Прокис Дж.

ЦИФРОВАЯ СВЯЗЬ



РАДИО И СВЯЗЬ

Proakis J.G.

DIGITAL COMMUNICATIONS



Mc Graw-Hill

Прокис Дж.

ЦИФРОВАЯ СВЯЗЬ

Перевод с английского под редакцией
Заслуженного деятеля наук Российской Федерации,
Действительного члена Академии инженерных наук
Российской Федерации, доктора технических наук,
профессора Кловского Д.Д.

МОСКВА
РАДИО И СВЯЗЬ
2000

УДК 621.396.2
ББК 32.884.1
П78

Федеральная программа поддержки книгоиздания России

Переводчики: Кловский Д.Д., Николаев Б.И.

Прокис Джон

П78 Цифровая связь. Пер. с англ. / Под ред. Д.Д. Кловского. - М.: Радио и связь. 2000. - 800 с.: ил.

ISBN 5-256-01434-X.

В систематизированном виде излагаются основные сведения по цифровым системам связи: элементам теории стохастических процессов и кодирования, вопросам определения и оценки параметров систем связи и их сигналов, блочным и сверточным канальным кодам, каналам с ограниченной полосой пропускания и линейными фильтрами, адаптивной коррекции, системам с мультинесущей, многоканальным пользовательским системам и т.п.

Для специалистов в области связи. Может быть полезна аспирантам и студентам вузов, обучающимся по направлению "Телекоммуникации".

ББК 32.884.1

ISBN 5-256-01434-X (рус.)
ISBN 007-051726-6 (англ.).

© McGraw-Hill, 1995
© Перевод на русский язык
Кловский Д.Д., Николаев Б.И., 2000
© Предисловие к русскому изданию,
примечания Кловский Д.Д., 2000

ПРЕДИСЛОВИЕ АВТОРА К РУССКОМУ ИЗДАНИЮ

Я признателен профессору Даниилу Кловскому за предложение написать предисловие к русскому переводу моей книги «Цифровая связь». С момента первого издания книги в 1983 г., книга была переведена и переиздана в 1989 и 1995 гг. с тем чтобы отразить новые результаты в синтезе и анализе характеристик качества цифровых систем связи.

Третье издание книги содержит много разделов, которые не были включены в первые два издания, особо это касается систем связи со многими пользователями. Однако область цифровой связи продолжает быстро развиваться. Значительными являются исследования в области алгоритмов кодирования и декодирования, особенно работы в области каскадных кодов с перемежением, называемых также турбокодами, и итеративных алгоритмов декодирования, которые предназначены для внедрения в следующем поколении цифровых сотовых систем связи и систем связи в пространственно-временных каналах. Эти вопросы не включены в третье издание. Тем не менее настоящее издание «Цифровой связи» знакомит читателя с базовыми элементами в синтезе и анализе качества цифровых систем связи.

Книга спланирована так, чтобы дать учебный материал для первого года специальной подготовки студентов по электрической связи. Она также рассчитана для использования в самостоятельной подготовке и как справочник для практикующего инженера, занятого разработкой цифровых систем связи. Я предполагаю, что читатель имеет базовые знания по теории линейных систем и некоторые предварительные знания по теории вероятностей и теории случайных процессов.

Глава 1 является введением в предмет и включает историческую справку, а также описание характеристик и моделей каналов.

Глава 2 содержит обзор основных понятий теории вероятностей и теории случайных процессов. В ней описываются некоторые функции распределения вероятностей и моменты, которые применяются на протяжении всей книги, а также граница Чернова, которая используется для получения границ качества цифровых систем связи.

Глава 3 рассматривает кодирование дискретных и аналоговых источников. Акцент делается на технику скалярного и векторного квантования. Сравнение различных решений выполняется на основе базовых результатов теории функции скорость-искажение.

В главе 4 читатель знакомится с сигналами цифровой модуляции и с характеристиками узкополосных сигналов и систем. В этой главе также обсуждаются спектральные характеристики сигналов цифровой модуляции.

В главе 5 исследуются вопросы синтеза устройств модуляции и методов оптимальной демодуляции и детектирования для цифровой передачи по каналу с аддитивным белым гауссовским шумом. Акцент делается на расчёте вероятности ошибок, как характеристике качества для различной техники цифровой передачи, и на обосновании требований к полосе частот канала для соответствующих сигналов.

Глава 6 посвящена оцениванию фазы несущей и методам тактовой синхронизации, основанным на алгоритмах максимального правдоподобия. Рассматриваются методы оценивания, как управляемые, так и не управляемые решениями (о переданных информационных символах).

И., 2000
здаию,
2000

В главе 7 обсуждается тема пропускной способности канала для некоторых моделей канала при случайном кодировании.

В главе 8 рассматриваются линейные блоковые и свёрточные коды. Обсуждаются методы синтеза кодов с использованием регистров сдвига и логических элементов, а также

алгоритмы декодирования, включающие декодирование мягких и жёстких решений. Анализируется качество этих алгоритмов в канале с аддитивным белым гауссовским шумом.

Глава 9 сконцентрирована на синтезе сигналов для ограниченных по полосе каналов. Эта глава включает материал по сигналам с парциальным откликом и кодам с ограниченным разбегом для формирования необходимого спектра сигнала.

В главе 10 обсуждаются вопросы демодуляции и детектирования сигналов, искажённых межсимвольной интерференцией. Особое внимание уделяется оптимальным и субоптимальным методам выравнивания и оценке их качества.

В главе 11 рассматривается адаптивное канальное выравнивание. Приведены алгоритмы минимума среднеквадратической ошибки и рекуррентные алгоритмы наименьших квадратов, а также характеристики их качества. Обсуждаются алгоритмы слепого выравнивания.

Глава 12 трактует вопросы модуляции в многоканальных системах и в системах со многими несущими. Последняя тема особенно актуальна с точки зрения некоторых важных приложений, которые появились в последние два десятилетия.

Глава 13 рассматривает широкополосные сигналы (с рассеянным спектром) и системы, использующие эти сигналы. Акцент делается на вопросы кодирования при синтезе широкополосных сигналов.

В главе 14 рассматриваются вопросы связи по каналам с замираниями. Для этих каналов рассматривается различные статистические модели замираний, с акцентом на замирания по Релею и Накагами. В эту главу также включён материал по использованию решетчатых кодов в каналах с замираниями.

В главе 15 обсуждаются проблемы синтеза систем связи со многими пользователями. Акцент делается на системах многостанционного доступа с кодовым разделением каналов (CDMA), детектировании сигналов и методах случайного доступа в таких системах, как ALOHA и CSMA.

Располагая 15 главами и варьируя темы, преподаватель имеет возможность построить одно- или двухсеместровый курс. Главы 3–6 образуют базу для рассмотрения методов цифровой модуляции-демодуляции и детектирования. Материал по канальному кодированию, рассмотренному в главах 7 и 8, может быть включён совместно с материалом по модуляции и демодуляции в односеместровый курс. Темы по канальному выравниванию, каналам с замираниями, широкополосным сигналам и по системам связи со многими пользователями могут быть включены в следующий семестр.

Я хочу выразить мою признательность профессору Кловскому за перевод книги на русский язык. Профессор Кловский является основоположником в области цифровой связи и особенно в области синтеза и анализа качества цифровой связи в каналах с рассеянием во времени и со случайно меняющимися параметрами. На протяжении всего периода перевода книги мы часто переписывались, и он сделал целый ряд критических замечаний и предложений, направленных на достижение ясности в изложении ряда вопросов в книге. Я благодарен ему за этот вклад.

Я имею существенную оговорку при публикации моей книги в России. Это касается ссылок лишь на несколько публикаций русских авторов в обсуждаемой области. Большинство ссылок, содержащихся в моей библиографии (имеющиеся в журнале IEEE, публикациях научных конференций, так же как и в книгах), написаны на английском языке. Я надеюсь, что русские читатели примут мои извинения за эти очевидные упущения, которые в большой степени объясняются моим недостаточным знанием русского языка.

ПРЕДИСЛОВИЕ РЕДАКТОРА РУССКОГО ПЕРЕВОДА

Книга известного американского учёного и специалиста в области цифровой связи профессора Джона Дж. Прокиса охватывает базовые вопросы синтеза и анализа систем цифровой связи в каналах различной физической природы. Это прежде всего: проводные каналы телефонии для передачи дискретных сообщений (данных), широкополосные проводные каналы передачи данных, радиоканалы различных диапазонов волн для передачи дискретных сообщений, каналы, образуемые в различных системах записи, хранения и считывания информации.

Книга не перегружена математическими выкладками и рассчитана на широкий круг читателей, имеющих базовые знания по теории линейных систем и некоторые начальные знания по основам теории вероятностей и теории случайных процессов.

Книга содержит 15 глав, соединённых единой методологией подачи материала, которая отличается простотой и ясностью изложения. Недаром книга Джона Дж. Прокиса «Цифровая связь» выдержала три переиздания и готовится 4-е издание.

Привлекательной особенностью книги является большое число задач, замыкающих все её главы (кроме первой). В книге приведена обширная и полезная библиография из 445 названий.

Особый интерес привлекают главы, посвящённые передаче данных по каналам с межсимвольной интерференцией (МСИ) и аддитивным белым гауссовским шумом (АБГШ), в частности, различным алгоритмам выравнивания характеристики канала и их реализации, как при использовании тестовых сигналов, так и при их отсутствии (слепые эквалайзеры). В этой области широко известны оригинальные исследования самого автора. Достаточно широко в книге представлен материал по использованию широкополосных (шумоподобных) сигналов в системах связи, в частности, в системах со многими пользователями с кодовым разделением сигналов пользователей (СДМА), а также элементы теории линейных систем телекоммуникаций со многими пользователями.

В книге имеется интересный материал по эффективному синтезу сигналов (и, в частности, выбору сигнальных созвездий) для частотно-ограниченных каналов.

Книга Дж. Прокиса содержит объёмный и интересный материал по кодированию источников и по канальному кодированию (с использованием линейных и нелинейных блоковых, свёрточных и более общих решётчатых кодов), по кодовой модуляции. Приведён обширный материал по декодированию мягких и жёстких решений демодулятора при различных моделях линейных каналов (с памятью и без памяти) с АБГШ. Особый интерес представляют эти результаты для выбора кодов в многопутевых радиоканалах с замираниями. Модели некоторых из них весьма характерны и для каналов акустической подводной связи.

При подготовке перевода нам пришлось в какой-то мере исправить недостаток книги Джона Прокиса, который автор и сам признаёт в своём предисловии к настоящему изданию, а именно почти полное отсутствие ссылок и освещения базовых результатов известных российских учёных и специалистов в области телекоммуникаций. Мы это сделали путём примечаний в тексте с указанием дополнительной литературы, помещённой после библиографии автора.

Для сокращения объёма книги было решено исключить некоторые повторы, а также не приводить предметный указатель в конце книги на 25 страницах. Вместе с тем сочтено целесообразным добавить список используемых в тексте сокращений (аббревиатур).

В ходе многочисленных контактов с Джоном Прокисом при работе над переводом книги и редактированием перевода мы лучше узнали друг друга. Он оказался весьма корректным автором, мгновенно реагирующим на критику и конструктивные замечания.

Моя личная встреча с Джоном Прокисом в 2000 году в США, которая прошла весьма интересно и продуктивно, позволила мне увидеть исключительно симпатичного, интеллигентного, высокоэрудированного и доброжелательного человека.

Надеюсь, что публикация в России перевода 3-го издания книги Джона Прокиса будет встречена с пониманием в широких кругах специалистов, занятых развитием и эксплуатацией цифровых систем связи, которые по достоинству оценят её как энциклопедию по вопросам общей теории передачи дискретных сообщений по каналам связи, а также приложений этой теории для синтеза систем цифровой связи в каналах различной физической природы. Книга будет полезна в качестве учебного пособия при изучении основ цифровой связи на различном уровне и как справочный материал для широкого круга специалистов.

Я хотел бы выразить большую признательность Б.И. Николаеву, который являлся моим незаменимым творческим помощником при переводе этой большой книги.

Д. Кловский



Дж. Дж. Прокис (слева) и Д.Д. Кловский в г. Лексингтоне, штат Массачусетс 2 августа 2000 г.

СПИСОК СОКРАЩЕНИЙ

Сокращение на русском	Сокращение на английском	Расшифровка
АБГШ	AWGN	Аддитивный белый гауссовский шум
АВ	VA	Алгоритм Витерби
АДИКМ	ADPCM	Адаптивная ДИКМ
АКН	KNA	Алгоритм Кловского-Николаева
АЛОХА	ALOHA	Система связи со случайным доступом пользователей в канал и узкополосными сигналами
АМ	AM	Амплитудная модуляция
АМ	ASK, (OOK)	Амплитудная (двоичная) манипуляция
АП	AJ	Антипомеховая (система)
АПМ	CMA	Алгоритм с постоянным модулем
АРУ	AGC	Автоматическая регулировка уровня
	AMI	Чередование инверсий единиц
АСКИО	ASCII	Американский стандарт кодирования информационного обмена
АСН	ZFA	Алгоритм сведения к нулю
БПФ	FFT	Быстрое преобразование Фурье
БЧХ	BCH	(Коды) Боуза–Чоудхури–Хоквингема
	GF(·)	Поле Галуа
ГПН	CWJ	Гармоническая помеха на несущей
	GTE	General Telephone and Electronics (Company)
ГУН	VCO	Генератор, управляемый напряжением
ДБН	NRZ	Двоичный (двухполярный) сигнал без возврата к нулю
ДБНП	NRZI	ДБН с памятью
ДБП	DSB	Двухполосная (амплитудная) модуляция
ДИБП	DMS	Дискретный источник без памяти
ДИКМ	DPCM	Дифференциальная ИКМ
ДКБП	DMC	Дискретный канал без памяти
ДМ	DM	Дельта-модуляция
ДМНМН	CVSD	ДМ с непрерывно меняющимся наклоном
ДПФ	DFT	Дискретное преобразование Фурье
ДСК	BSC	Двоичный симметричный канал
ДФМ	DPSK	Дифференциальная ФМ
ДЭ	FSE	Дробный эквалайзер
ЗПВ	LOS	За пределами прямой видимости
ИКМ	PCM	Импульсно-кодовая модуляция
ИФР	cdf	Интегральная функция распределения
КАМ	QAM	Квадратурная амплитудная модуляция
КИХ	FIR	Конечная импульсная характеристика
КЗФ	QMF	Квадратурные зеркальные фильтры
КПП	SBC	Кодирование подполосок
КСПО	QPRS	Квадратурный сигнал с парциальным откликом
КФМС	SQPSK	Квадратурная ФМ со сдвигом
ЛКВВС	VSELP	Линейный кодер (вокодер) с возбуждением векторной суммой
ЛККВ	CELP	Линейный кодер (вокодер) с кодовым возбуждением
ЛКП	LPC	Линейное кодирование с предсказанием
МАВ	MAP	Максимум апостериорной вероятности
МАС	ASK	Модуляция амплитудным сдвигом (манипуляция)
МДВР	TDMA	Множественный доступ с временным разделением
МДРС	SSMA	Многостанционный доступ с рассеянным спектром

МДЧР	FDMA	Множественный доступ с частотным разделением
МДКР	CDMA	Множественный доступ с кодовым разделением
ММС	MSK	Модуляция с минимальным сдвигом
МНФ	CPM	Модуляция с непрерывной фазой
МП	ML	Максимум правдоподобия
МППО	MLSE	Максимально правдоподобное последовательное оценивание
МСИ	ISI	Межсимвольная интерференция
МСКО	MMSE	Минимум СКО
МТИ	MIT	Массачусетский технологический институт
Н	CW	Несущая
НВП	LPI	Низкая вероятность перехвата
НК	LMS	(Критерий, алгоритм) наименьших квадратов
ОАВ	GVA	Обобщённый алгоритм Витерби
ОБП, ОМ	SSB	Однополосная модуляция
ОДПФ	IDFT	Обратное дискретное преобразование Фурье
ОКФМ	OQPSK	Офсетная квадратурная ФМ
ОНСД/ДС	CSMA/CD	Обнаружение несущей с детектированием столкновений
ОСР	DF	Обратная связь по решению
ОСШ	SNR	Отношение сигнал-шум
$ОСШ_0$	SNR_0	ОСШ на выходе согласованного фильтра
ПВВ	TDL	Петля с временной «вилкой»
ПВМС	PBI	Парциально-временной мешающий сигнал
ПЗ	DLL	Петля с задержкой
ПК	RC	Приподнятый косинус
ПКЛ	KLT	Преобразование Карунена–Лоэва
ПП	DS	Прямая (широкополосная) последовательность
прп	r r t	Примечание редактора перевода
ПЦПР	RWSSD	Приём в целом с поэлементным решением. Вместе с ОСР – синоним АКН
ПШ	PN	Псевдошумовой
Рейк	RAKE	Широкополосная система цифровой радиосвязи, впервые предложенная для КВ канала
РНК	RLS	Рекуррентный НК
СВ	HT	(Сигналы со) скачками времени (задержки)
СИД	LED	Светоизлучающий диод
СИИП	SIIP	Система с испытательным импульсом и предсказанием
СКО	MSE	Среднеквадратическая ошибка
СЧ	HF	Скачки частоты
СФПВ	jpdf	Совместная ФПВ
ТСП	TATS	Тактическая система передачи
ТУН	VCC	Таймер (задающий генератор), управляемый напряжением
ФАП	PLL	Фазовая автоподстройка
ФЗП	PLL	Фазо-замкнутая петля
ФПВ	pdf	Функция плотности вероятности
ФМ-2	PM-2	Двоичная фазовая модуляция
ФМ,МФС	PSK	Фазовая модуляция (манипуляция)
ЧМ, МЧС	FSK	Частотная модуляция (модуляция частотным сдвигом)
ЧМНФ	CPFSK	Частотная модуляция с непрерывной фазой
ЭМИ	ERP,EIRP	Эффективная мощность излучения
ЭНВП	ZFE	Эквалайзер с нулевыми взаимными помехами
ЭОСР	DFE	Эквалайзер с обратной связью по решению

Посвящается Фелии, Джорджу и Елене

1

ВВЕДЕНИЕ

В этой книге мы представляем основные принципы, которые лежат в основе анализа и синтеза систем цифровой связи. Предмет цифровой связи включает в себя передачу информации в цифровой форме от источника, который создаёт информацию для одного или многих мест назначения. Особенно важным для анализа и синтеза систем связи являются характеристики физических каналов, через которые передаётся информация. Характеристики канала обычно влияют на синтез базовых составных блоков системы связи. Ниже мы опишем элементы системы связи и их функции.

1.1. ЭЛЕМЕНТЫ СИСТЕМ ЦИФРОВОЙ СВЯЗИ

Функциональную схему и основные элементы цифровой системы связи поясняет рис. 1.1.1. Выход источника может быть либо аналоговым сигналом, как звуковой или видеосигнал, либо цифровым сигналом, как выход печатающей машины, – он дискретен во времени и имеет конечное число выходных значений. В системе цифровой связи сообщения, выданные источником, преобразуются в последовательность двоичных символов. В идеале мы можем представить выход источника сообщения небольшим числом двоичных символов (насколько это возможно). Другими словами, мы ищем эффективное представление выхода источника, которое приводит к источнику с наименьшей избыточностью или с полным её отсутствием. Процесс эффективного преобразования выхода источника – как аналогового, так и цифрового – в последовательность двоичных символов называют *кодированием источника или сжатием данных.*

Последовательность двоичных символов от кодера источника, который мы назовём *источником информации*, поступает на *кодер канала.* Цель кодера канала состоит в том, чтобы ввести управляемым способом некоторую избыточность в информационную двоичную последовательность, которая может использоваться в приёмнике, чтобы преодолеть влияние шума и интерференции, с которой сталкиваются при передачи сигнала через канал. Таким образом, добавленная избыточность служит для увеличения надёжности принятых данных и улучшает верность воспроизведения принятого сигнала. Фактически избыточность в информационной последовательности помогает приёмнику в декодировании переданной информационной последовательности. Например, тривиальной формой кодирования исходной двоичной последовательности является простое повторение каждого двоичного символа m раз, где m – некоторое целое положительное число. Более сложное (нетривиальное) кодирование сводится к преобразованию блока из k информационных символ в уникальную последовательность из n символов, называемую *кодовым словом.* Значение избыточности, вводимой при кодировании данных таким способом, измеряется отношением n/k. Обратная величина этого отношения, а именно k/n, названа *скоростью* кода.

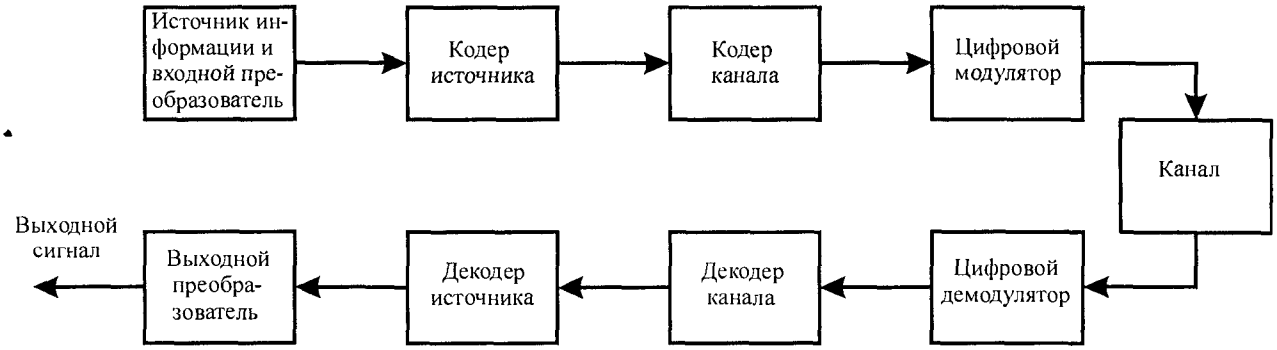


Рис. 1.1.1 Основные элементы цифровой системы связи

Двоичная последовательность на выходе кодера канала поступает на *цифровой модулятор*, который служит интерфейсом к каналу связи. Так как почти все каналы связи, с которыми сталкиваются на практике, способны к передаче электрических сигналов (волновых процессов), основная цель цифрового модулятора сводится к отображению информационной двоичной последовательности в соответствующий сигнал. Чтобы разобраться с этим вопросом, предположим, что кодированная информационная последовательность должна передать один бит за определённое время с постоянной скоростью R бит/с. Цифровой модулятор может просто отображать двоичный символ 0 в сигнал $s_0(t)$, а двоичный символ 1 – в сигнал $s_1(t)$. Таким способом каждый бит кодера передаётся отдельно. Мы называем это *двоичной модуляцией*. В качестве альтернативы модулятор может передавать b кодированных информационных битов одновременно, используя различные сигналы $s_i(t)$, $i=0, \ldots, M-1$, один сигнал для каждого из $M=2^b$ возможных b-битовых последовательностей. Мы назовём это *M–позиционной модуляцией* ($M>2$). Заметим, что информационная последовательность с b битами поступает на вход модулятора каждые b/R секунд. Следовательно, когда канальная скорость передачи данных R фиксирована, для передачи одного из M сигналов, соответствующих информационной последовательности из b бит, отведён в b раз больший интервал времени, чем при двоичной модуляции.

Канал связи – это физическая среда, которая используется для передачи сигнала от передатчика к приёмнику. При беспроволочной связи каналом может быть атмосфера (свободное пространство). С другой стороны, телефонные каналы обычно используют ряд физических сред, включая линии проводной связи, волоконно-оптические кабели и беспроволочные линии (например, микроволновую радиолинию). Для любой физической среды, используемой для передачи информации, существенно, что передаваемый сигнал подвержен случайным искажениям через такие механизмы, как воздействие аддитивного *теплового шума*, генерируемого электронными устройствами, воздействие промышленных помех (например, автомобильные помехи от системы зажигания), воздействие атмосферных помех (электрические разряды молнии во время грозы) и т.п.

На приёмной стороне системы цифровой связи *цифровой демодулятор* обрабатывает искажённый каналом передаваемый сигнал и преобразует его в последовательность чисел, которые представляют оценки переданных данных (двоичных или M-позиционных). Эта последовательность чисел поступает на канальный декодер, который пытается восстановить первоначальную информационную последовательность, используя знание канального кода и избыточности, содержащейся в принятых данных.

Мера качества работы демодулятора и декодера – это частота, с которой возникают ошибки декодируемой последовательности. Более точно, средняя вероятность ошибки на бит для выходных символов декодера является удобной характеристикой качества

демодулятора-декодера. Вообще говоря, вероятность ошибки является функцией от характеристик кода, форм сигналов, используемых для передачи информации по каналу, мощности передатчика, характеристик канала, а именно уровня шума, природы интерференции и т.д., и методов демодуляции и декодирования. Эти обстоятельства и их влияние на характеристики качества системы связи будут обсуждаться подробно в последующих главах.

На заключительной стадии, когда рассматривается аналоговый выход, декодер источника принимает выходную последовательность от декодера канала и, используя знание метода кодирования источника, применённого на передаче, пытается восстановить исходную форму сигнала источника. Ошибки декодирования и возможные искажения в кодере и декодере источника приводят к тому, что сигнал на выходе декодера источника является аппроксимацией исходного сигнала источника. Разность или некоторая функция разности между исходным и восстановленным сигналом является мерой искажения, внесённого цифровой системой связи.

1.2. КАНАЛЫ СВЯЗИ И ИХ ХАРАКТЕРИСТИКИ

Как было указано в предшествующем обсуждении, канал связи обеспечивает соединение передатчика и приёмника. Физический канал может быть двухпроводной линией, которая пропускает электрический сигнал, или стекловолокном, которое переносит информацию посредством модулированного светового луча, или подводным каналом океана, в котором информация передаётся акустически, или свободным пространством, по которому несущий информационный сигнал излучается при помощи антенны. Другие среды, которые могут характеризоваться как каналы связи – средства хранения данных, такие как магнитная лента, магнитные и оптические диски.

Одна общая проблема при передаче сигнала через любой канал – аддитивный шум. Вообще говоря, аддитивный шум создаётся часто внутри различных электронных компонентов, таких как резисторы и твёрдотельные устройства, используемых в системах связи. Эти шумы часто называют *тепловым шумом*. Другие источники шума и интерференции (наложения) могут возникать вне системы, например переходные помехи от других пользователей канала. Когда такой шум и переходные помехи занимают тот же самый диапазон частот, что и полезный сигнал, их влияние может быть минимизировано путем соответствующего выбора передаваемого сигнала и демодулятора в приемнике. Другие виды сигнальных искажений, которые могут встречаться при передаче сигнала по каналу, – это затухание сигнала, амплитудные и фазовые искажения сигнала и искажения сигнала, обусловленные многопутевым распространением волн.

Влияние шума может быть уменьшено увеличением мощности передаваемого сигнала. Однако конструктивные и другие практические соображения ограничивают уровень мощности передаваемого сигнала. Другое базовое ограничение – доступная ширина полосы частот канала. Ограничение ширины полосы обычно обусловлено физическими ограничениями среды и электрических компонентов, используемых в передатчике и приемнике. Эти два обстоятельства приводят к ограничению количества данных, которые могут быть переданы надёжно по любому каналу связи, как мы увидим в последующих главах книги. Ниже мы опишем некоторые из важных характеристик отдельных каналов связи.

Проводные каналы. Телефонная сеть экстенсивно использует проводные линии для передачи звукового сигнала, а также данных и видеосигналов. Витые проводные пары и коаксиальный кабель в основном дают электромагнитный канал, который обеспечивает

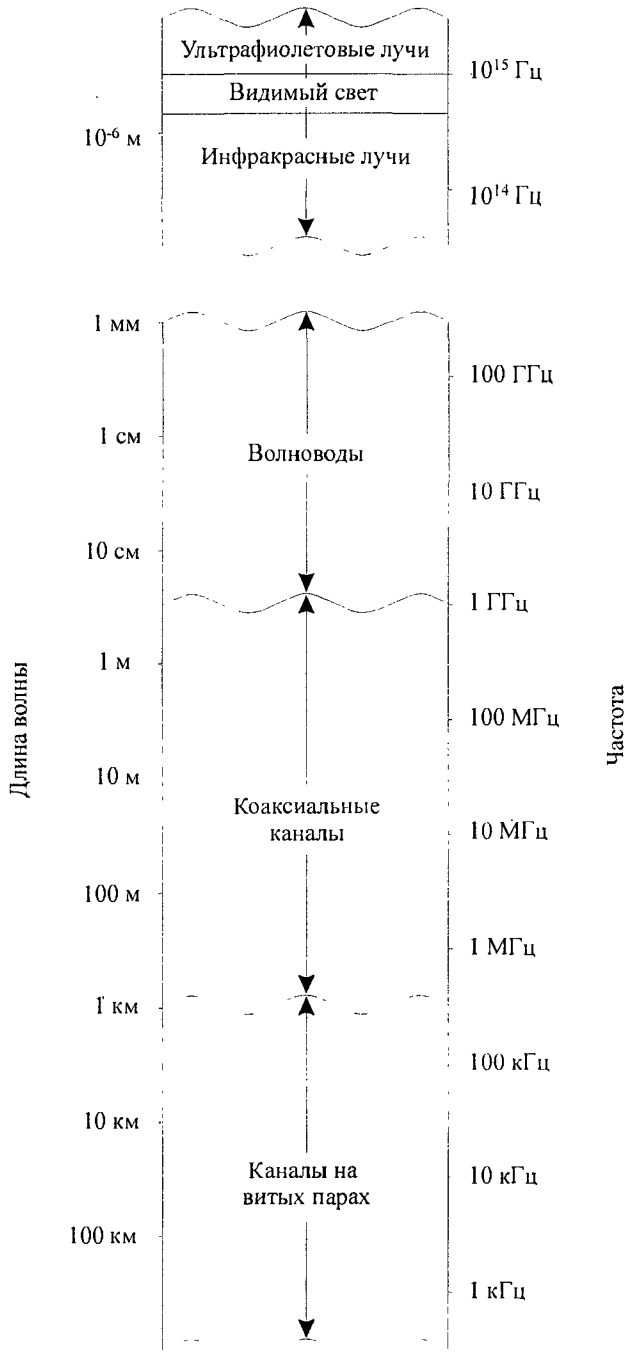


Рис. 1.2.1. Частотные диапазоны для каналов связи с направляющими системами

прохождение относительно умеренной ширины полосы частот. Телефонный провод, обычно используемый, чтобы соединить клиента с центральной станции, имеет ширину полосы несколько сотен килогерц. С другой стороны, коаксиальный кабель имеет обычно используемую ширину полосы частот несколько мегагерц. Рисунок 1.2.1 поясняет частотный диапазон используемых электромагнитных каналов, которые включают волноводы и оптический кабель.

Сигналы, передаваемые через такие каналы, искажаются по амплитуде и фазе, и, кроме того, на них накладывается аддитивный шум. Проводная линия связи в виде витой пары также склонна к интерференции переходных помех от рядом расположенных пар. Поскольку проводные каналы составляют большой процент каналов связи по всей стране и миру, широкие исследования были направлены на определение их свойств передачи и на уменьшение амплитудных и фазовых искажений в канале. В гл. 9 мы опишем методы синтеза оптимальных передаваемых сигналов и демодуляторов; в гл. 10 и 11 рассмотрим синтез канальных эквалайзеров (выравнивателей), которые компенсируют амплитудные и фазовые искажения в канале.

Волоконно-оптические каналы. Стекловолокно предоставляет проектировщику системы связи ширину полосы частот, которая на несколько порядков больше, чем у каналов с коаксиальным кабелем. В течение прошедшего десятилетия были разработаны оптические кабели, которые имеют относительно низкое затухание для сигнала, и высоконадёжные оптические устройства для генерирования и детектирования сигнала. Эти технологические достижения привели к быстрому освоению таких каналов как для внутренних систем электросвязи, так и для трансатлантических и мировых систем связи. С учётом большой ширины полосы частот, доступной на волоконно-оптических каналах,

стало возможно для телефонных компаний предложить абонентам широкий диапазон услуг электросвязи, включая передачу речи, данных, факсимильных и видеосигналов.

Передатчик или модулятор в волоконно-оптической системе связи – источник света, светоизлучающий диод (СИД) или лазер. Информация передается путем изменения (модуляции) интенсивности источника света посредством сигнала сообщения. Свет распространяется через волокно как световая волна, и она периодически усиливается (в случае цифровой передачи детектируется и восстанавливается ретрансляторами) вдоль тракта передачи, чтобы компенсировать затухания сигнала.

В приемнике интенсивность света детектируется фотодиодом, чей выход является электрическим сигналом, который изменяется пропорционально мощности света на входе фотодиода. Источники шума в волоконно-оптических каналах – это фотодиоды и электронные усилители.

Предполагается, что волоконно-оптические каналы заменят почти все каналы проводной линии связи в телефонной сети на рубеже столетия.

Беспроводные (радио) каналы. В системах беспроводной связи (радиосвязи) электромагнитная энергия передается в среду распространения антенной, которая служит излучателем. Физические размеры и структура антенны зависят прежде всего от рабочей частоты. Чтобы получить эффективное излучение электромагнитной энергии, размеры антенны должны быть больше чем 1/10 длины волны. Следовательно, передача радиостанции с АМ на несущей, допустим, f_c = 1 МГц, соответствующей длине волны $\lambda = c/f_c$ = 300 м, требует антенны с диаметром по крайней мере 30 м. Другие важные характеристики и свойства антенн для беспроволочной передачи описаны в гл. 5.

Рисунок 1.2.2 поясняет различные диапазоны частот для радиосвязи. Способы распространения электромагнитных волн в атмосфере и в свободном пространстве можно разделить на три категории, а именно: распространение поверхностной волной, распространение пространственной волной, распространение прямой волной. В диапазоне очень низких частот (ОНЧ) и звуковом диапазоне, в которых длины волн превышают 10 км, земля и ионосфера образуют волновод для распространения электромагнитных волн. В этих частотных диапазонах сигналы связи фактически распространяются вокруг всего земного шара. По этой причине эти диапазоны частот прежде всего используются во всём мире для решения навигационных задач с берега до кораблей.

Ширина полосы частот канала, доступной в этих диапазонах, относительно мала (обычно составляет 1...10 % центральной частоты), и, следовательно, информация, которая передаётся через эти каналы, имеет относительно низкую скорость передачи и обычно неприемлема для цифровой передачи.

Доминирующий тип шума на этих частотах обусловлен грозовой деятельностью вокруг земного шара, особенно в тропических областях. Интерференция возникает из-за большого числа станций в этих диапазонах частот.

Распространение земной волной, как иллюстрируется на рис. 1.2.3, является основным видом распространения для сигналов в полосе средних частот (0,3...3 МГц). Это – диапазон частот, используемый для радиовещания с АМ и морского радиовещания. При АМ радиовещании и распространении земной волной дальность связи, даже при использовании мощных радиостанций, ограничена 150 км. Атмосферные шумы, промышленные шумы и тепловые шумы от электронных компонентов приёмника являются основными причинами искажений сигналов, передаваемых в диапазоне средних частот.

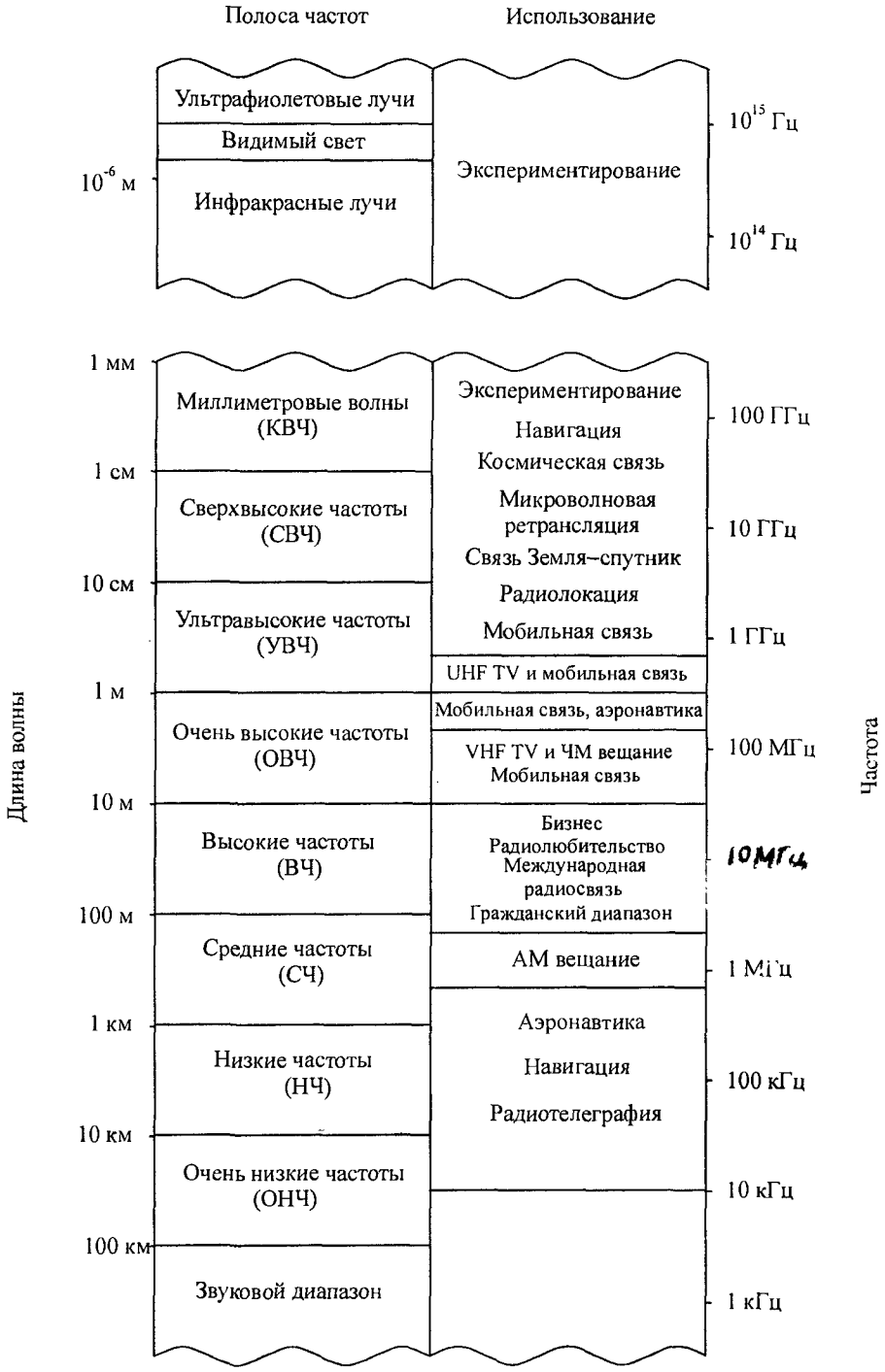


Рис. 1.2.2. Частотные диапазоны для беспроводных каналов связи
[*Carlson* (1975), 2-е изд., © *McGraw-Hill Book Co.*]

Частным случаем распространения пространственной волны является ионосферное распространение, иллюстрируемое рис. 1.2.4. Оно сводится к отражению (отклонение или рефракция волны) передаваемого сигнала от ионосферы, которая состоит из нескольких слоёв заряженных частиц, расположенных на высоте 50...400 км от поверхности земли. В дневное время суток разогрев нижних слоёв атмосферы солнцем обусловливает появление нижнего слоя на высоте ниже 120 км. Эти нижние слои, особенно *D*-слой, вызывают

поглощение частот ниже 2 МГц, таким образом ограничивая распространение ионосферной волной радиопередач АМ радиовещания. Однако в течение ночных часов электронная концентрация частиц в нижних слоях ионосферы резко падает, и частотное поглощение, которое встречается в дневное время, значительно сокращается. Как следствие, мощные радиовещательные сигналы с АМ могут распространяться на большие расстояния посредством отражения от ионосферных слоев (которые располагаются на высоте от 140 до 400 км над поверхностью земли), и земной поверхности.

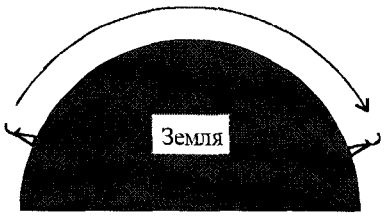


Рис. 1.2.3. Иллюстрация распространения поверхностной волной

Часто возникающая проблема при ионосферном распространении электромагнитной волны в частотном диапазоне ВЧ – это *многопутёвость*. Многопутёвость образуется потому, что передаваемый сигнал достигает приёмника по многим путям с различными задержками. Это обычно приводит к межсимвольной интерференции в системе цифровой связи. Более того, сигнальные компоненты, прибывающие по различным путям распространения, могут суммироваться таким образом, что это приводит к явлению, названному *замираниями*. Это большинство людей испытало при слушании отдалённой радиостанции ночью, когда ионосферная волна является доминирующим способом распространения. Аддитивный шум в ВЧ диапазоне – это комбинация атмосферных помех и теплового шума. Распространение ионосферной волны прекращается на частотах выше ≈30 МГц, что является границей диапазона ВЧ. Однако возможно ионосферно-тропосферное распространение на частотах в диапазоне от 30 до 60 МГц, обусловленное рассеянием сигналов от нижних слоев ионосферы. Также можно связаться на расстоянии нескольких сотен миль при помощи тропосферного рассеяния в диапазоне от 40 до 300 МГц. Тропосферное рассеяние обуславливается рассеянием сигнала благодаря частицам в атмосфере на высотах порядка 10 км. Обычно ионосферное и тропосферное рассеяние вызывает большие сигнальные потери и требует большой мощности передатчика и относительно больших размеров антенн.

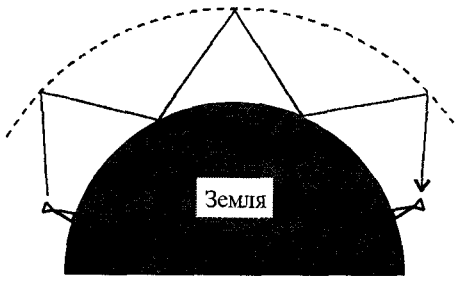


Рис. 1.2.4. Иллюстрация распространения пространственной волной

Частоты выше 30 МГц проходят через ионосферу с относительно малыми потерями и делают возможным спутниковую и внеземную связь. Следовательно, на частотах УВЧ диапазона и выше основным способом электромагнитного распространения волн является

распространение в пределах прямой видимости (ППВ). Для земных систем связи это означает, что передающая и приемная антенны должны быть в прямой видимости с относительно малой преградой (или ее отсутствием). По этой причине передача телевизионных станций в УВЧ и СВЧ диапазонах частот для достижения широкой зоны охвата осуществляется антеннами на высоких опорах.

Вообще, зона охвата для ППВ распространения ограничена кривизной поверхности земли. Если передающая антенна установлена на высоте h м над поверхностью земли, расстояние до радиогоризонта, не принимая во внимание физические преграды, такие как горы, приблизительно $d = \sqrt{15h}$ км. Например, антенна телевидения, установленная на высоте 300 м, обеспечивает покрытие территории приблизительно 67 км. Другой пример – релейные системы микроволновой радиосвязи, экстенсивно используемые для передачи телефонных и видеосигналов на частотах выше чем 1 МГц, имеют антенны, установленные на высоких опорах или сверху на высоких зданиях.

Доминирующий шум, ограничивающий качество системы связи в ВЧ и УВЧ диапазонах, – тепловой шум, создаваемый во входных цепях приемника, и космические шумы, уловленные антенной. На частотах в диапазоне СВЧ выше чем 10 ГГц при распространения сигнала главную роль играют атмосферные условия. Например, на частоте 10 ГГц затухание меняется приблизительно от 0,003 дБ/км при лёгком дожде до 0,3 дБ/км при тяжёлом дожде. На частоте 100 ГГц затухание меняется приблизительно от 0,1 дБ/км при легком дожде до 6 дБ/км при тяжёлом дожде. Следовательно, в этом частотном диапазоне тяжелый дождь вызывает чрезвычайно высокие потери при распространении, которые могут приводить к отказу системы обслуживания (полный обрыв в системе связи).

На частотах выше КВЧ (крайне высокие частоты) полосы мы имеем диапазон инфракрасного и видимого излучений – области электромагнитного спектра, который может использоваться для применения ППВ оптической связи в свободном пространстве. До настоящего времени эти диапазоны частот использовались в экспериментальных системах связи типа связи между спутниками.

Подводные акустические каналы. За последние 40 лет исследования океанской деятельности непрерывно расширялись. Это связано с усилением потребности передать данные, собранные датчиками, размещенными под водой и на поверхности океана. Оттуда данные передаются к центру сбора информации.

Электромагнитные волны не распространяются на большие расстояния под водой, за исключением крайне низких частот. Однако передача сигналов таких низких частот предельно дорога из-за чрезвычайно больших и мощных передатчиков. Затухание электромагнитных волн в воде может быть выражено *глубиной поверхностного слоя*, которая является расстоянием, на котором сигнал ослабляется в **e** раз. Для морской воды глубина поверхностного слоя $\delta = 250/\sqrt{f}$, где f выражена в герцах, а δ – в метрах. Например, для частоты 10 кГц глубина поверхностного слоя 2,5 м. Напротив, акустические сигналы распространяются на расстояния порядка десятков и даже сотен километров.

Подводный акустический канал ведет себя как многопутевой канал благодаря сигнальным отражениям от поверхности и дна моря. Из-за случайного движения волны сигнальные продукты многопутевого (многолучевого) распространения приводят к случайным во времени задержкам распространения и в итоге к замираниям сигнала. Кроме того, имеется частотно-зависимое затухание, которое приблизительно пропорционально квадрату частоты сигнала. Глубинная скорость номинально равна приблизительно 1500 м/с, но реальное значение выше или ниже номинального значения в зависимости от глубины, на которой сигнал распространяется.

Окружающий океанский акустический шум вызван креветкой, рыбой и различными млекопитающими. Ближние гавани добавляют к окружающему шуму промышленный шум. Несмотря на эту помеховую окружающую среду, возможно проектировать и выполнять эффективные и безопасные подводные акустические системы связи для передачи цифровых сигналов на большие расстояния.

Системы хранения информации и системы поиска информации составляют значительную часть систем повседневной обработки данных. Это магнитная лента, включая цифровую наклонно-строчную звукозапись, и видеолента, магнитные диски, используемые для хранения больших количеств данных компьютера, оптические диски, используемые для хранения данных компьютера. Компакт-диски – также пример систем хранения информации, которые могут рассматриваться как каналы связи. Процесс запоминания данных на магнитной ленте или магнитном или оптическом диске эквивалентен передаче сигнала по телефону или радиоканалу. Процесс считывания и сигнальные процессы, используемые в системах хранения, чтобы восстанавливать запасенную информацию, эквивалентен функциям, выполняемым приемником в системе связи для восстановления передаваемой информации.

Аддитивный шум, издаваемый электронными контактами, и интерференция от смежных дорожек обычно представлены в сигнале считывания записанной информации точно так, как это имеет место в системе проводной телефонии или системе радиосвязи. Количество данных, которые можно хранить, ограничено размером диска или ленты и плотностью записи (числом битов, хранящихся на единице площади), которая может быть достигнута электронными системами и головками записи-считывания. Например, плотность упаковки 10^8 бит на квадратный сантиметр демонстрировалась в экспериментальной системе хранения на магнитном диске. (Текущие коммерческие магнитные изделия хранения достигают значительно меньшей плотности.) Скорость, с которой данные могут быть записаны на диске или ленте, и скорость, с которой информация может считываться, также ограничены механическими и электрическими подсистемами, входящими в систему хранения информации.

Кодирование канала и модуляция – существенные компоненты хорошо разработанной цифровой магнитной или оптической системы хранения. В процессе считывания сигнал демодулируется и его избыточность, введённая кодером канала, используется для исправления ошибок считывания.

1.3. МАТЕМАТИЧЕСКИЕ МОДЕЛИ КАНАЛОВ СВЯЗИ

При синтезе систем связи для передачи информации через физические каналы мы используем математические модели, которые отображают наиболее важные характеристики среды передачи. Затем математическая модель канала используется для синтеза кодера и модулятора в передатчике и демодулятора и декодера в приёмнике. Ниже мы приводим краткое описание моделей каналов, которые часто используются для отображения многих физических каналов, с которыми мы сталкиваемся на практике.

Канал с аддитивным шумом. Самая простая математическая модель для канала связи – это канал с аддитивным шумом, иллюстрируемый на рис. 1.3.1. В этой модели передаваемый сигнал $s(t)$ подвержен воздействию лишь аддитивного шумового процесса $n(t)$. Физически аддитивный шум возникает от посторонних электрических помех, электронных компонентов и усилителей в приёмнике систем связи, а также из-за интерференции сигналов.

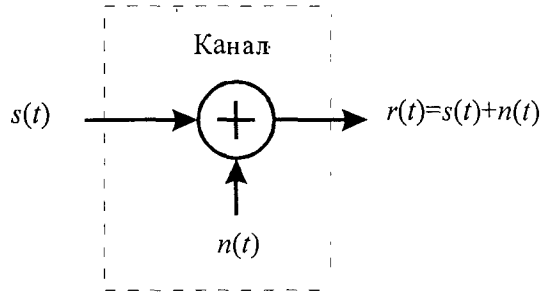


Рис. 1.3.1. Канал с аддитивным шумом

Если шум обусловлен в основном электронными компонентами и усилителями в приёмнике, его можно описать как тепловой шум. Этот тип шума характеризуется статистически как *гауссовский шумовой процесс*. Как следствие, результирующую математическую модель обычно называют *каналом с аддитивным гауссовским шумом*. Поскольку эта модель применима к широкому классу физических каналов связи и имеет простую математическую интерпретацию, она является преобладающей моделью канала при анализе и синтезе систем связи. Затухание каналов легко включается в модель. Если при прохождении через канал сигнал подвергается ослаблению, то принимаемый сигнал

$$r(t) = \alpha s(t) + n(t), \qquad (1.3.1)$$

где α – коэффициент затухания линейного канального фильтра.

Линейный фильтровой канал. В некоторых физических каналах, таких как проводные телефонные каналы, фильтры используются для того, чтобы гарантировать, что передаваемые сигналы не превышают точно установленные ограничения на ширину полосы и, таким образом, не интерферируют друг с другом. Такие каналы обычно характеризуются математически как линейные фильтровые каналы с аддитивным шумом, что иллюстрируется на рис. 1.3.2. Следовательно, если на вход канала поступает сигнал $s(t)$, на выходе канала имеем сигнал

$$r(t) = s(t) * c(t) + n(t) = \int_{-\infty}^{\infty} c(\tau) \cdot s(t-\tau) d\tau + n(t), \qquad (1.3.2)$$

где $c(t)$ – импульсная характеристика линейного фильтра, а $*$ обозначает свертку.

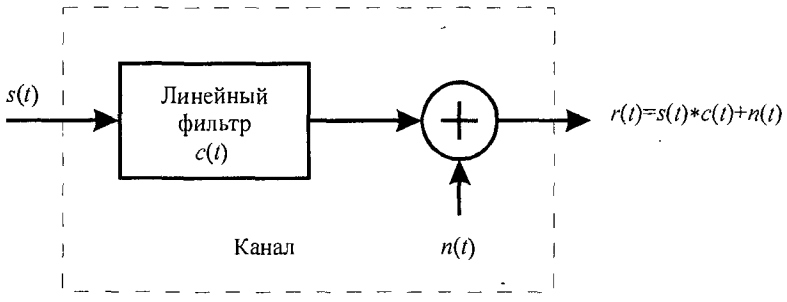


Рис. 1.3.2. Линейный фильтровой канал с аддитивным шумом

Линейный фильтровой канал с переменными параметрами. Физические каналы, такие как подводные акустические каналы и ионосферные радиоканалы, которые возникают в условиях меняющегося во времени многопутевого распространения передаваемого сигнала, могут быть описаны математически как линейные фильтры с переменными параметрами. Такие линейные фильтры характеризуются меняющимися во времени импульсной характеристикой канала $c(\tau,t)$, где $c(\tau,t)$ – отклик канала в момент времени t на δ-импульс, поданный ко входу в момент $t-\tau$. Таким образом, τ представляет

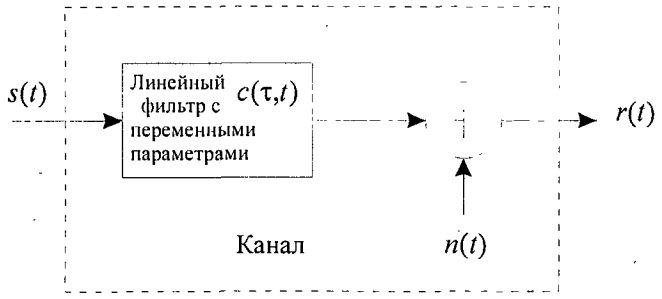


Рис. 1.3.3. Линейный фильтровой канал с переменными параметрами и аддитивным шумом

«ретроспективную» переменную. Линейный фильтровой канал с переменными параметрами и аддитивным шумом иллюстрируется на рис. 1.3.3.

Для входного сигнала $s(t)$ выходной сигнал канала

$$r(t) = s(t) * c(\tau; t) + n(t) = \int_{-\infty}^{\infty} c(\tau; t) \cdot s(t-\tau) d\tau + n(t). \qquad (1.3.3)$$

Хорошей моделью для многопутевого распространения волн через физические каналы типа ионосферы (на частотах ниже 30 МГц) и каналы подвижной сотовой радиосвязи является частный случай (1.3.3), когда переменная во времени импульсная характеристика канала имеет вид

$$c(\tau; t) = \sum_{k=1}^{L} a_k(t) \delta(\tau - \tau_k), \qquad (1.3.4)$$

где $\{a_k(t)\}$ определяет возможные меняющиеся во времени коэффициенты затухания для L путей распространения, $\{(\tau_k)\}$ – соответствующие им времена задержки. Если (1.3.4) подставить в (1.3.3), то принимаемый сигнал

$$r(t) = \sum_{k=1}^{L} a_k(t) s(\tau - \tau_k) + n(t). \qquad (1.3.5)$$

Следовательно, полученный сигнал состоит из L компонентов распространения, где каждый компонент умножается на $a_k(t)$ и запаздывает на τ_k.

Три математические модели, описанные выше, адекватно характеризуют большинство физических каналов, с которыми сталкиваются на практике. Эти три модели канала используются в книге для анализа и синтеза систем связи.

1.4. ИСТОРИЧЕСКИЙ ОБЗОР РАЗВИТИЯ ЦИФРОВОЙ СВЯЗИ

Следует отметить, что самая ранняя форма электрической связи, а именно *телеграфная связь*, была системой цифровой связи. Электрический телеграф был разработан Сэмюэлем Морзе и демонстрировался в 1837 г. Морзе изобрел двоичный код переменой длины, в котором буквы английского алфавита представлены последовательностью точек и тире (кодовые слова). В этом коде часто встречающиеся буквы представлены короткими кодовыми словами, в то время как буквы, встречающиеся менее часто, – более короткими кодовыми словами. Таким образом, *код Морзе* был предшественником методов кодирования источников кодом переменной длины, описанных в гл. 3.

Почти 40 годами позже, в 1875 г., Эмиль Бодо изобрёл код для телеграфной связи, в котором каждая буква кодировалась двоичным кодом фиксированной длины 5. В коде Бодо элементы двоичного кода имеют равную длину и именуются посылкой и паузой.

Хотя Морзе принадлежит первая электрическая система цифровой связи (телеграфная связь), начало того, что мы теперь считаем современной теорией цифровой связи, следует из работ Найквиста (1924), исследовавшего проблему определения максимальной скорости передачи, которую можно обеспечить по телеграфному каналу данной ширины полосы частот без межсимвольной интерференции (МСИ). Он сформулировал модель телеграфной системы, в которой передаваемый сигнал имеет общую форму

$$s(t) = \sum_n a_n g(t - nT), \qquad (1.4.1)$$

где $g(t)$ – базовая форма импульса (несущей); $\{a_n\}$ – последовательность данных в двоичном коде (± 1), передаваемых со скоростью $1/T$ бит/с.

Найквист пытался определить оптимальную форму импульса $g(t)$ с ограниченной полосой W Гц и максимизировать скорость передачи данных в предположении, что импульс не вызывает МСИ в точках отсчёта kT, $k = 0, \pm 1, \pm 2, \ldots$ Эти исследования привели его к заключению, что максимальная скорость передачи равна $2W$ отсч./с. Эту скорость теперь называют скоростью Найквиста. Более того, эту скорость передачи можно достичь при использовании импульса $g(t) = \sin 2\pi Wt/(2\pi Wt)$. Эта форма импульса допускает восстановление данных без межсимвольных помех в выборочные моменты времени. Результат Найквиста эквивалентен версии теоремы отсчётов[1] для сигналов с ограниченной полосой, который был позже точно сформулирован Шенноном (1948)[2]. Теорема отсчётов гласит, что сигнал с шириной полосы частот W может быть восстановлен по его отсчётам, взятым со скоростью Найквиста $2W$, путем использования интерполяционной формулы

$$s(t) = \sum_n s\left(\frac{n}{2W}\right) \frac{\sin\left[2\pi W\left(t - n/(2W)\right)\right]}{2\pi W\left(t - n/(2W)\right)}. \qquad (1.4.2)$$

В продолжение работы Найквиста Хартли (1928) рассмотрел вопрос о количестве данных, которые могут быть переданы надежно по каналу с ограниченной полосой частот, когда для последовательной передачи данных используются импульсы со многими амплитудными уровнями. С учетом шума и другой интерференции Хартли показал, что приемник может надежно оценивать амплитуду принятого сигнала с некоторой точностью A_δ. Это исследование привело Хартли к заключению, что имеется максимальная скорость передачи данных по каналу с ограниченной полосой частот, зависящая от максимальной амплитуды сигнала $A_{\max}$ (фиксированной максимальной мощности) и величины A_δ.

---

[1] Теорема отсчётов (Котельникова) на самом деле дуальна теореме Найквиста: в первой речь идет о передаче непрерывного сигнала с помощью его отсчётов (по каналу с дискретным временем), а во второй – о передаче дискретного сигнала (последовательности отсчётов) по непрерывному каналу. В первой теореме $\Delta t \le 1/(2W)$, а во второй – $\Delta t \ge 1/(2W)$. На практике никогда не достигается равенство, поэтому в первой теореме $\Delta t < 1/(2W)$, а во второй – $\Delta t > 1/(2W)$ (прп – так будем сокращённо обозначать примечания редактора перевода).

[2] Теорема отсчётов в теории связи была впервые сформулирована и доказана В.А. Котельниковым (1933) [1] (дополнительные ссылки на литературные источники, введенные редактором перевода, даны отдельным списком), причём в более общем виде, чем (1.4.2). Общий вид теоремы отсчётов следует из (1.4.2), если в этой формуле заменить $(t - n/(2W))$ на $(t - n\Delta t)$, где $\Delta t \le 1/(2W)$ (прп).

Другим значительным вкладом в развитие теории связи была работа Винера (1942), который рассмотрел проблему оценивания полезного сигнала $s(t)$ на фоне аддитивного шума $n(t)$, исходя из наблюдения принимаемого сигнала $r(t) = s(t) + n(t)$. Эта проблема возникает при демодуляции сигналов. Винер определил линейный фильтр, выход которого является лучшей среднеквадратической аппроксимацией полезного сигнала $s(t)$. Полученный фильтр назван *оптимальным линейным* (*винеровским*) *фильтром* [1].

Результаты Хартли и Найквиста по максимальной скорости передачи цифровой информации были предшественниками работ Шеннона (1948), который установил математические основы передачи информации по каналам связи и нашел фундаментальные ограничения для систем цифровой связи. В своей пионерской работе Шеннон сформулировал основную проблему надежной передачи информации в терминах статистической теории связи, используя вероятностные модели для информационных источников и каналов связи. Применяя вероятностный подход, он нашёл универсальную логарифмическую меру для количества информации источника. Он также показал, что существует некоторый предельный показатель, характеризующий скорость передачи информации по каналу связи, зависящий от величины мощности передатчика, ширины полосы и интенсивности аддитивного шума, названный им пропускной способностью канала. Например, в случае аддитивного белого (с равномерным спектром) гауссовского шума идеальный частотно-ограниченный канал с шириной полосы W имеет пропускную способность C', бит/с, которая определяется формулой

$$C' = W \log_2\left(1 + \frac{P}{WN_0}\right), \qquad (1.4.3)$$

где P – средняя мощность сигнала, а N_0 – спектральная плотность мощности аддитивного шума. Значение параметра пропускной способности канала C' состоит в том, что если информационная скорость (производительность) источника R меньше, чем $C'(R < C')$, то теоретически возможно обеспечить надёжную (свободную от ошибок) передачу через канал соответствующим кодированием. С другой стороны, если $R > C'$, то надежная передача невозможна, независимо от способов обработки сигнала на передаче и приеме. Таким образом, Шеннон установил основные ограничения передачи информации и породил новое направление, которое теперь называется *теорией информации*[2]

Другой важный вклад в области цифровой связи – это работа Котельникова (1947), который провел тщательный анализ различных систем цифровой связи, основанный на

---

[1] Первые работы по корреляционной теории случайных процессов выполнены в 1934 г. А.Я. Хинчиным [2]. Первая работа по фильтрации сигналов на фоне помех по среднеквадратичному критерию качества выполнена в 1939 г. А.Н. Колмогоровым [3]. Поэтому оптимальный фильтр, работающий по среднеквадратическому критерию качества, с большим основанием следует называть и действительно называют фильтром Колмогорова–Винера. По исследованию уравнений Винера–Хопфа, определяющих оптимальную фильтрацию, выдающиеся результаты получены М.Г. Крейном в 1954 г. [4]. Рекуррентные алгоритмы оптимальной линейной фильтрации нестационарных марковских процессов независимо друг от друга найдены в 1960 г. Р.Л. Стратоновичем [5] и Р. Калманом. Основы теории нелинейной фильтрации марковских случайных процессов заложены работами Р.Л. Стратоновича в 1959–1961 гг. [6, 7] (прп).

[2] Первоначальные результаты Шеннона по теории информации были расширены как самим автором, так и многими математиками, в первую очередь в России: А.Н. Колмогоровым и А.Я. Хинчиным в 1956 г. [8, 9] и Р.Л. Добрушиным в 1959 г. [10].

геометрическом представлении.[1] Исследование Котельникова было позже развито Возенкрафтом и Джекобсом (1965).

Вслед публикациям Шеннона появилась классическая работа Хемминга (1950) по кодам с обнаружением и с исправлением ошибок, которые противодействуют вредному влиянию канального шума. Работа Хемминга стимулировала многих исследователей, которые в последующие годы открыли ряд новых и мощных кодов, многие из которых сегодня внедрены в современные системы связи.

Увеличение спроса на передачу данных в течение последних 3–4 десятков лет и развитие более сложных интегральных схем вело к созданию эффективных и надежных систем цифровой связи. В свете этих достижений оригинальные результаты Шеннона и обобщение его результатов по максимальным ограничениям на передачу информации по каналу и по достижимым характеристикам качества служили маяком при разработке любых проектов систем связи. Теоретические пределы, полученные Шенноном и другими исследователями, способствовали развитию теории информации и служат конечной целью в продолжающихся усилиях по разработке и развитию более эффективных систем цифровой связи.

За ранними работами Шеннона, Котельникова и Хемминга появилось много новых достижений в области цифровой связи. Некоторые из наиболее заметных достижений следующие:

- разработка новых блоковых кодов Маллером (1954), Ридом (1954) Ридом и Соломоном (1960), Боузом и Рой-Чоудхури (1960) и Гоппом (1970–1971);

---

[1] Вклад В.А. Котельникова в теорию связи более существенен. В его докторской диссертации «Теория потенциальной помехоустойчивости» (защищённой в 1947 г. на заседании Учёного совета Московского энергетического института) он впервые сформулировал задачу оптимального статистического синтеза приёмных устройств в неискажающем (однопутевом) линейном канале с аддитивным белым гауссовским шумом (АБГШ) в её современном виде как задачу различения гипотез и проанализировал с новых позиций различные системы связи, установив потенциальные ограничения на возможные виды модуляции [11]. Большую роль в распространении идей и методов статистической теории связи сыграли несколько книг А.А. Харкевича, появившиеся в 1955–1963 гг. Ими зачитывались студенты, преподаватели и специалисты самых различных направлений [12, 13, 14]. Первые работы по исследованию помехоустойчивости систем связи, в том числе при замираниях сигналов, выполнены в 1946 г. А.Н. Щукиным [15], В.И. Сифоровым [16, 17], в 1951 г. В.С. Мельниковым [18, 19] и В.И. Бунимовичем [20]. Вслед за монографией В.А. Котельникова появились первые монографии по теории оптимальной (когерентной и некогерентной) обработки сигналов в однопутевых каналах с аддитивным гауссовским шумом, в том числе при замираниях: в 1960 г. Л.А. Вайнштейна и В.Д. Зубакова [21], в 1961 г. Л.С. Гуткина [22], в 1963 г. Л.М. Финка [23], А.А. Фельдбаума. [24]. В 1959 г. Д.Д. Кловский впервые получил [25] оптимальный (по правилу максимального правдоподобия) алгоритм демодуляции с обратной связью по решению для каналов с МСИ и переменными параметрами (для многопутевых каналов) с АБГШ при анализе на сигнальном интервале (тактовом интервале передачи). В 1970 г. Д.Д. Кловский и Б.И. Николаев обобщили этот алгоритм на случай анализа на интервале произвольной длительности, появился алгоритм приёма в целом с поэлементным решением [26] ПЦПР или АКН. Этот алгоритм обеспечивает примерно ту же помехоустойчивость, что и алгоритм Витерби (АВ), предложенный в 1972 Форни для демодуляции в каналах с МСИ, но требует меньших вычислительных затрат. Основы методов преодоления априорной неопределённости при обработке сигналов и получения систем, близких к оптимальным, которые сохраняют желаемые свойства при изменении параметров сигналов и помех и, кроме того, являются практически реализуемыми, заложены в 1963 г. работами А.А. Фельдбаума и Б.Р. Левина [24,27] и продолжены в работах В.Г. Репина и Г.П. Тартаковского [28]. Совместно оптимальные алгоритмы обнаружения, различения и оценивания параметров при обработке сигналов были начаты работами Б.Р. Левина и Ю.С. Шинакова в 1977 г. [29] и продолжены в работах А.П. Трифонова и Ю.С. Шинакова [30]. Оригинальные результаты в этом направлении (оптимальные оценочно-корреляционные алгоритмы обработки сигналов) получены в 1978 г. Ю.Г. Сосулиным [31] (прп).

- разработка каскадных кодов Форни (1966) [1];
- разработка эффективных в вычислительном отношении БЧХ кодов, например, алгоритма Берлекампа-Месси (см. Чейн, 1964; Берлекамп, 1968) [2];
- разработка сверточных кодов и алгоритмов декодирования Возенкрафтом и Рейффеном (1961), Фано (1963), Зигангировым (1966), Елинеком (1969), Форни (1970, 1972) и Витерби (1967, 1971);
- разработка решетчато-кодированной модуляции Унгербоеком (1982), Форни и др. (1984), Ваем (1987) и др.[3];
- разработка эффективных алгоритмов кодирования источника для сжатия данных, таких как алгоритм Зива и Лемпела (1977, 1978) и Линда и др. (1980).

1.5. БИБЛИОГРАФИЧЕСКИЕ ЗАМЕЧАНИЯ И ССЫЛКИ

Имеются некоторые исторические обзоры, посвящённые развитию радиотехники и систем связи в течение последнего столетия.[4] Их можно найти в книгах МакМагона (1984), Мильмана (1984) и Ридера и Финка (1984). Мы уже процитировали классические работы Найквиста (1924), Хартли (1928), Котельникова (1947), Шеннона (1948) и Хемминга (1950 и некоторые другие важные работы, опубликованные после 1950 г. Сборник работ Шеннона был издан книгой IEEE Press под редакцией Слоэна и Вайнера (1993)[5]. Другие сборники работ, опубликованные IEEE Press, которые могут представить интерес для читателя: Key Papers in the Development of Coding Theory, под ред. Берлекампа (1974), и Key Papers in the Development of Information Theory, под ред. Слепяна (1974).

---

[1] Идеи каскадного кодирования Форни обобщены в 1972 г. Э.Л. Блохом и В.В. Зябловым [32, 33]; ими же предложены обобщённые каскадные коды, нашедшие применение в каналах с ограниченной полосой (прп).

[2] Идеи порогового (мажоритарного) декодирования Месси были существенно развиты в 1968 г. В.О. Колесником и Е.Г. Мирончиковым [34] (прп).

[3] В.И. Коржик, С.А. Осмоловский и Л.М. Финк предложили в 1972 г. новый реализуемый практически стохастический подход к кодированию в системах с обратным каналом, посредством которого в любых двоичных каналах вероятность необнаруженной ошибки не превосходит заданную величину, определяемую параметрами кода, а не свойствами канала [35, 36, 37] (прп).

[4] Интересный обзор по развитию статистической теории связи до конца 70-х годов XX столетия и её приложениям можно найти в монографии [38], выпущенной издательством «Связь» в 1979 г. под редакцией Б.Р. Левина (прп).

[5] Первый сборник работ К. Шеннона был издан значительно раньше в Москве в 1963 г. [39] (прп).

2

ВЕРОЯТНОСТЬ И СЛУЧАЙНЫЕ ПРОЦЕССЫ

Теория вероятностей и случайных процессов – это существенный математический инструмент при проектировании систем цифровой связи. Этот инструмент важен при статистическом моделировании источников, которые выдают аналоговый сигнал, преобразуемый затем в цифровую форму, при определении характеристик канала, через который передаётся цифровая информация, при создании приёмника, который обрабатывает сигнал, несущий информацию из канала, и при оценке характеристик качества систем связи. Мы коснёмся лишь ограниченной части теории вероятностей и теории случайных процессов. Приведём ряд определений и основных понятий из теории вероятностей и теории случайных процессов, и несколько результатов, которые являются особенно важными при проектировании эффективных систем цифровой связи и оценке их характеристик.

Мы ожидаем, что большинство наших читателей имеют некоторое априорное представление о теории вероятностей и теории случайных процессов, так что наше изложение они воспримут, прежде всего, как обзор. Эти читатели извлекут с выгодой для себя дополнительную информацию из чтения интересного материала по этим вопросам, имеющего инженерную направленность и содержащегося в учебниках Давенпорта и Рута (1958 г.), Давенпорта (1970 г.), Папулиса (1984 г.), Хелстрома (1991 г.), и Леона–Гарсиа (1994 г.).

2.1. ВЕРОЯТНОСТЬ

Рассмотрим, например, такой эксперимент, как бросание игральной кости с рядом возможных исходов. Выборочное пространство S эксперимента состоит из набора всех возможных его исходов. В случае игральной кости

$$S = \{1, 2, 3, 4, 5, 6\}, \tag{2.1.1}$$

где целые числа 1...6 представляют числа, указанные на шести сторонах игральной кости. Эти шесть возможных исходов – выборочные (характерные) точки эксперимента. Событием является некоторая часть от S, которая может состоять из любого числа характерных точек. Например, событие A, определённое как

$$A = \{2, 4\}, \tag{2.1.2}$$

состоит из результатов 2 и 4. Дополнение к событию A, обозначаемое $\overline{A}$, состоит из всех характерных точек в S, которых нет в A, следовательно,

$$\overline{A} = \{1, 3, 5, 6\}. \tag{2.1.3}$$

Два события считают взаимоисключающими (несовместными), если они не имеют никаких общих характерных точек – т.е. если появление одного результата исключает появление другого. Например, если A определено как в (2.1.2), а событие B определим как

$$B = \{1, 3, 6\}, \tag{2.1.4}$$

тогда A и B – несовместные события. Точно так же A и $\overline{A}$ – несовместны.

Объединение (сумма) двух событий – это событие, которое состоит из всех

характерных точек двух событий. Например, если B определено, как в (2.1.4), а событие C – как

$$C = \{1, 2, 3\}, \tag{2.1.5}$$

тогда объединение событий B и C, обозначаемое $B \cup C$, является событием

$$D = B \cup C = \{1, 2, 3, 6\}. \tag{2.1.6}$$

Точно так же $A \cup \overline{A} = S$, где S – всё выборочное пространство, определяющее достоверное событие.

Пересечение двух событий – событие, которое состоит из характерных точек, общих для обоих событий. Таким образом, если $E = B \cap C$ представляет пересечение событий B и C, определяемых (2.1.4) и (2.1.5) соответственно, то

$$E = \{1, 3\}.$$

Если события несовместны, их пересечение – событие с нулевой вероятностью, обозначаемое как $\varnothing$. Например, $A \cap B = \varnothing$ и $A \cap \overline{A} = \varnothing$.

Определения для объединения и пересечения событий можно непосредственно расширить на более чем два события. Каждому событию A из пространства S приписывается его вероятность $P(A)$. При назначении вероятностной меры для событий мы принимаем аксиоматическую точку зрения. Это означает, что мы полагаем, что вероятность событий А удовлетворяет условию $P(A) \geq 0$. Мы также полагаем, что вероятность всего выборочного пространства S (достоверного события) $P(S)$=1. Третья аксиома касается вероятности взаимоисключающих (несовместных) событий. Предположим, что A_i, i=1,2,..., являются рядом (возможно, бесконечным) несовместных событий в выборочном пространстве S, так что

$$A_i \cap A_j = \varnothing, \quad i \neq j = 1, 2, \ldots$$

Тогда вероятность объединения (суммы) этих несовместных событий удовлетворяет условию

$$P\left(\bigcup_i A_i\right) = \sum_i P(A_i). \tag{2.1.7}$$

Например, в случае бросания игральной кости каждый возможный исход (событие) имеет вероятность 1/6. Событие, определённое (2.1.2), состоит из двух несовместных подсобытий или исходов, следовательно, $P(A)$=2/6=1/3. Аналогично вероятность события $A \cup B$, где A и B – несовместные события, определённые соответственно (2.1.2) и (2.1.4), равна $P(A)+P(B)$=1/3+1/2=5/6.

Совместные события и совместные вероятности. Предположим, что мы имеем дело не с одним, а с двумя экспериментами и рассматриваем их исходы. В качестве примера двух экспериментов можно рассматривать два отдельных бросания одной игральной кости или одно бросание двух игральных костей. В любом случае выборочное пространство S состоит из 36 дублетов (i, j), где i, j = 1,2,...,6. Если бросание производится чисто, то каждой точке выборочного пространства назначаем вероятность 1/36. Мы теперь можем рассматривать, например, объединённые события вида $\{i$ – чётное, j=3$\}$ и определять соответствующие вероятности таких событий, зная вероятности всех возможных характерных точек.

Вообще, если один эксперимент имеет возможные исходы A_i, i=1, 2,..., n, а второй эксперимент – B_j, j=1, 2,..., m, тогда объединённый эксперимент имеет возможные совместные исходы (A_i, B_j), i=1, 2,..., n, j=1, 2,..., m. Каждому объединённому исходу (A_i, B_j) присваивается вероятность $P(A_i, B_j)$, которая удовлетворяет условиям

$$0 \le P(A_i, B_j) \le 1.$$

В предположении, что исходы B_j, j=1, 2,..., m, являются несовместными, получаем

$$\sum_{j=1}^{m} P(A_i, B_j) = P(A_i). \qquad (2.1.8)$$

Точно так же, если исходы A_i, i=1, 2,..., n, являются несовместными, то

$$\sum_{i=1}^{n} P(A_i, B_j) = P(B_j). \qquad (2.1.9)$$

Далее, если все результаты из двух экспериментов несовместны, то

$$\sum_{i=1}^{n} \sum_{j=1}^{m} P(A_i, B_j) = 1. \qquad (2.1.10)$$

Обобщение вышеупомянутого положения на более чем два эксперимента очевидно.

Условные вероятности. Рассмотрим комбинированный эксперимент, в котором исход встречается с вероятностью $P(A, B)$. Предположим, что событие B произошло, и мы желаем определить вероятность того, что при этом произошло событие A. Эта вероятность называется *условной вероятностью* события A при условии, что событие B имеет место, и определяется как

$$P(A|B) = \frac{P(A,B)}{P(B)} \qquad (2.1.11)$$

в предположении, что $P(B)>0$. Подобным же образом вероятность события B при условии, что событие A имело место, определяется как

$$P(B|A) = \frac{P(A,B)}{P(A)} \qquad (2.1.12)$$

в предположении, что $P(A)>0$. Формулы (2.1.11) и (2.1.12) могут быть переписаны в виде

$$P(A,B) = P(A|B)P(B) = P(B|A)P(A). \qquad (2.1.13)$$

Соотношения в (2.1.11)–(2.1.13) применимы также к единственному эксперименту, в котором A и B являются двумя событиями, определёнными на выборочном пространстве S, а $P(A,B)$ интерпретируется как вероятность $A \cap B$. Т.е. $P(A,B)$ определяет вероятность одновременного наступления (пересечения) событий A и B. Например, рассмотрим события B и C, определённые (2.1.4) и (2.1.5) соответственно, для единственного бросания кости. Совместное событие состоит из выборочных точек {1,3}. Условная вероятность события C при условии, что B произошло, равна

$$P(C|B) = \frac{2/6}{3/6} = \frac{2}{3}.$$

В единственном эксперименте мы наблюдаем, что, когда два события A и B несовместны, $A \cap B = \varnothing$ и, следовательно, $P(A|B) = 0$. Так же, если A входит в B, тогда $A \cap B = A$ и, следовательно,

$$P(A|B) = \frac{P(A)}{P(B)}.$$

С другой стороны, если B входит в A, мы имеем $A \cap B = B$ и, следовательно,

$$P(A|B) = \frac{P(B)}{P(B)} = 1.$$

Чрезвычайно полезные соотношения для условных вероятностей выражаются теоремой Байеса, которая гласит, что если A_i, i=1,2,...,n, являются несовместными событиями, так что

$$\bigcup_{i=1}^{n} A_i = S,$$

и B – произвольное событие с отличной от нуля вероятностью, тогда

$$P(A_i|B) = \frac{P(A_i,B)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{n} P(B|A_j)P(A_j)}. \qquad (2.1.14)$$

Мы используем эту формулу в гл. 5 для нахождения структуры оптимального приёмника для системы цифровой связи, в которой события A_i, i=1, 2,..., n, представляют в нашем случае возможные передаваемые сообщения на данном временном интервале, а $P(A_i)$ представляют их *априорные вероятности*, B – принятый сигнал, подверженный действию шума, который содержит передаваемое сообщение (одно из A_i), а $P(A_i|B)$ является *апостериорной вероятностью* A_i при условии, что наблюдается принятый сигнал B.

Статистическая независимость. Статистическая независимость двух или большего числа событий – другое важное понятие теории вероятности. Она обычно возникает, когда мы рассматриваем два или больше экспериментов или результатов повторений одного эксперимента. Чтобы пояснить это понятие, мы рассматриваем события A и B и их условную вероятность $P(A|B)$, которая является вероятностью события A при условии, что событие B произошло. Предположим, что появление события A не зависит от появления события B. Это значит, что

$$P(A|B) = P(A). \qquad (2.1.15)$$

Подставив (2.1.15) в (2.1.13), получаем результат

$$P(A,B) = P(A)P(B). \qquad (2.1.16)$$

Это означает, что совместная вероятность событий A и B определяется произведением элементарных или собственных вероятностей событий $P(A)$ и $P(B)$. Когда события A и B удовлетворяют соотношению (2.1.16), их называют *статистически независимыми.*

Например, рассмотрим два последовательных эксперимента бросания кости. Пусть A представляет выборочные точки с чётными номерами {2,4,6} в первом бросании, а B представляет чётно нумерованную выборку {2,4,6} во втором бросании. В случае правильной кости мы считаем что вероятность $P(A)$= 3/6=1/2 и $P(B)$=3/6=1/2. Теперь вероятность совместного исхода – чётно нумерованный результат при первом бросании и чётно нумерованный результат при втором бросании – является вероятностью результата для девяти возможных пар (i,j), i = 2,4,6, j = 2,4,6, которая равна 9/36 = 1/4. Но мы имеем также

$$P(A,B) = P(A)P(B) = 1/4.$$

Таким образом, результаты A и B статистически независимы. Точно так же мы можем говорить, что исходы двух экспериментов статистически независимы.

Понятие статистической независимости может быть расширено на три и большее число событий. Три статистически независимых события A_1, A_2 и A_3 должны удовлетворять следующим условиям:

$$\begin{aligned} &P(A_1,A_2) = P(A_1)P(A_2); \qquad P(A_1,A_3) = P(A_1)P(A_3); \\ &P(A_2,A_3) = P(A_2)P(A_3); \qquad P(A_1,A_2,A_3) = P(A_1)P(A_2)P(A_3). \end{aligned} \qquad (2.1.17)$$

В общем случае события A_i, i=1, 2,..., n, являются статистически независимыми при условии, что вероятности совместного наступления 2, 3,... n событий в любой комбинации определяются произведением вероятностей индивидуальных событий.

2.1.1. Случайные величины, распределение вероятностей и плотности вероятностей

Для данного эксперимента, имеющего выборочное пространство S с элементами $s \in S$, мы определяем функцию $X(s)$, область определения которой S, а областью значений является набор чисел на вещественной оси. Функцию $X(s)$ называют *случайной величиной*. Например, если мы бросаем монету, возможными результатами являются орёл (Н) и решка (Т), так что пространство S содержит 2 точки, маркированные как Н и Т. Предположим, что мы определяем функцию $X(s)$ так, что

$$X(s)=\begin{cases} 1 & (s=H), \\ -1 & (s=T). \end{cases} \qquad (2.1.18)$$

Таким образом, мы отображаем два возможных результата бросания монеты в виде двух точек (± 1) на вещественной оси. Другой эксперимент – бросание игральной кости с возможными исходами $S=\{1, 2, 3, 4, 5, 6\}$. Случайная переменная, определённая на этом выборочном пространстве, может быть $X(s)=s$. В этом случае результаты эксперимента отображаются целыми числами {1, 2, 3, 4, 5, 6}. Можно положить $X(s)=s^2$, тогда возможные результаты отображаются целыми числами {1, 4, 9, 16, 25, 36}. Это примеры *дискретных случайных величин*.

Хотя мы использовали в качестве примеров эксперименты, которые имеют конечное множество возможных исходов, имеется много физических систем, эксперименты в которых дают непрерывные выходные результаты. Например, шумовое напряжение, создаваемое электронным усилителем, имеет непрерывную амплитуду. Как следствие, выборочное пространство S амплитуд напряжения $\upsilon \in S$ непрерывно и таким же является отображение $X(\upsilon)=\upsilon$. В таком случае случайную величину[1] X называют *непрерывной случайной величиной*. Для случайной величины X рассмотрим событие $\{X \le x\}$, где x – любое вещественное число в интервале $(-\infty;+\infty)$. Определим вероятность этого события как $P(X \le x)$ и обозначим её как $F(x)$, т.е.

$$F(x)=P(X \le x) \quad (-\infty < x < \infty). \qquad (2.1.19)$$

Функция $F(x)$ названа *функцией распределения вероятности* случайной величины X. Её также называют *интегральной* (*кумулятивной*) *функцией распределения* (ИФР). Так как $F(x)$ – это вероятность, то её значения ограничены интервалом $0 \le F(x) \le 1$. Фактически $F(-\infty)=0$ и $F(\infty)=1$. Например, дискретная случайная величина, полученная при бросании монеты и определённая (2.1.18), имеет ИФР, показанную на рис. 2.1.1(*a*). Здесь имеются два скачка $F(x)$: один при $x=-1$ и другой при $x=1$. Точно так же случайная величина $X(s)=s$, полученная при бросании игральной кости, имеет ИФР, показанную на рис. 2.1.1 (*b*). В этом случае $F(x)$ имеет шесть скачков, в каждом из x =1,..., 6.

[1] Случайную величину $X(s)$ обычно обозначают просто X.

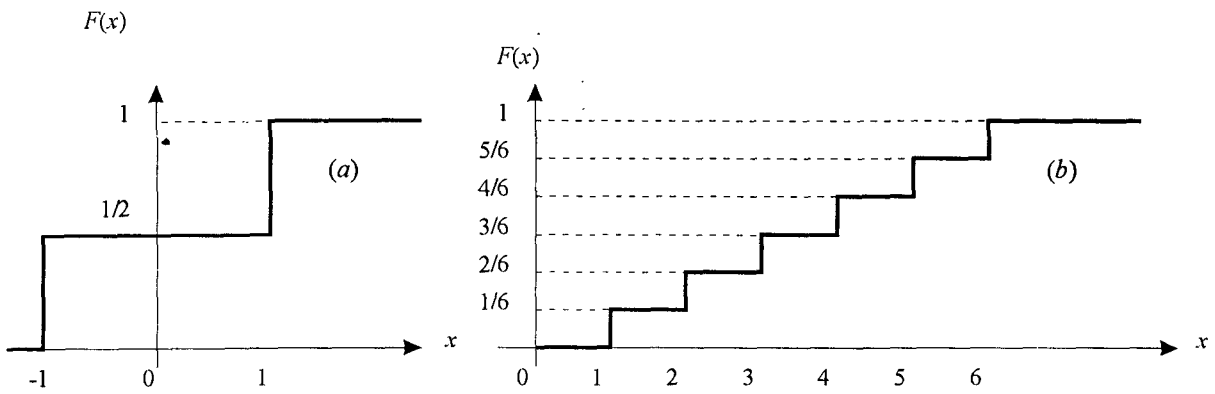


Рис. 2.1.1 Примеры интегральных функций распределения двух дискретных случайных величин

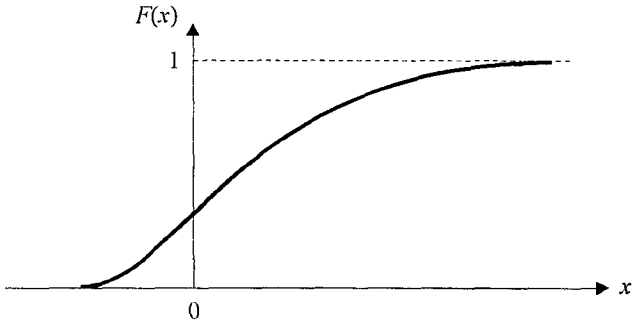


Рис. 2.1.2 Пример интегральной функции распределения непрерывной случайной переменной

ИФР непрерывной случайной величины обычно изменяется так, как показано на ис. 2.1.2. Это гладкая, неубывающая функция. В некоторых практических задачах мы южем также сталкиваться со случайной величиной смешанного типа. ИФР такой лучайной величины является гладкой неубывающей функцией в отдельных частях ещественной оси и содержащей скачки в ряде дискретных значений x. Пример такой ИФР ллюстрируется рис. 2.1.3.

Производная от ИФР $F(x)$, обозначаемая как $p(x)$, называется *функцией плотности вероятности* (ФПВ) случайной величины X. Таким образом, имеем

$$p(x)=\frac{dF(x)}{dx} \qquad (-\infty < x < \infty), \tag{2.1.20}$$

ли, что эквивалентно,

$$F(x)=\int_{-\infty}^{x} p(u)du \qquad (-\infty < x < \infty). \tag{2.1.21}$$

Так как $F(x)$ – неубывающая функция, то $p(x) \geq 0$. Когда случайная величина искретная или смешанного типа, ФПВ содержит δ-импульсы в точках нарушения епрерывности $F(x)$. В таких случаях дискретная часть $p(x)$ может быть выражена как

$$p(x)=\sum_{i=1}^{n} P(X=x_i)\delta(x-x_i), \tag{2.1.22}$$

ıе x_i, i=1, 2,..., n являются возможными дискретными значениями случайной величины; $(X=x_i)$, $i=1,2,...n$, являются вероятностями, а $\delta(x)$ обозначает δ-функцию.

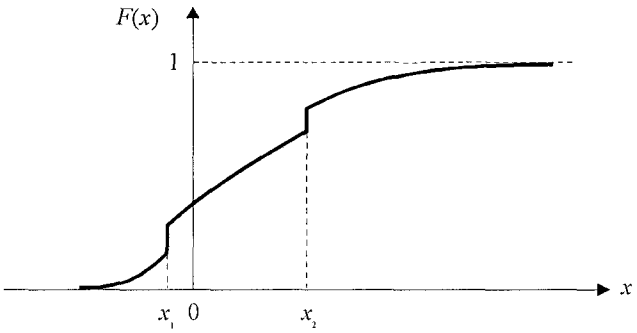


Рис. 2.1.3 Пример интегральной функции распределения случайной переменной смешанного типа

Часто мы сталкиваемся с проблемой определения вероятности того, что случайная величина X находится в интервале (x_1, x_2), где $x_2 > x_1$. Чтобы определить вероятность этого события, начнём с события $\{X \le x_2\}$. Это событие всегда можно выразить как объединение двух несовместных событий $\{X \le x_1\}$ и $\{x_1 < X \le x_2\}$. Следовательно, вероятность события $\{X \le x_2\}$ можно выразить как сумму вероятностей несовместных событий. Таким образом, мы имеем

$$P(X \le x_2) = P(X \le x_1) + P(x_1 < X \le x_2),$$
$$F(x_2) = F(x_1) + P(x_1 < X \le x_2)$$

или эквивалентное соотношение

$$P(x_1 < X \le x_2) = F(x_2) - F(x_1) = \int_{x_1}^{x_2} p(x)dx. \qquad (2.1.23)$$

Другими словами, вероятность события $\{x_1 < X \le x_2\}$ – это площадь под ФПВ в пределах $x_1 < X \le x_2$.

Многомерные случайные величины, совместные распределения вероятностей и совместные плотности вероятностей. Когда имеем дело с комбинированными экспериментами или повторениями одного эксперимента, мы сталкиваемся с многомерными случайными величинами и их ИФР и ФПВ. Многомерные случайные величины – в основном многомерные функции – определены на выборочном пространстве при комбинированном эксперименте. Начнём с двух случайных величин X_1 и X_2, каждая из которых может быть непрерывной, дискретной или смешанной. Совместная интегральная функция распределения (СИФР) для двух случайных величин определяется так:

$$F(x_1, x_2) = P(X_1 \le x_1, X_2 \le x_2) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2} p(u_1, u_2)du_1\, du_2, \qquad (2.1.24)$$

где $p(x_1, x_2)$ – совместная функция плотности вероятности (СФПВ). Последнюю можно также выразить в виде

$$p(x_1, x_2) = \frac{\partial^2}{\partial x_1 \partial x_2} F(x_1, x_2). \qquad (2.1.25)$$

Когда СФПВ $p(x_1, x_2)$ интегрируется по одной из переменных, мы получаем ФПВ по другой переменной, т.е.

$$\int_{-\infty}^{\infty} p(x_1, x_2)dx_1 = p(x_2), \qquad \int_{-\infty}^{\infty} p(x_1, x_2)dx_2 = p(x_1). \qquad (2.1.26)$$

ФПВ $p(x_1)$ и $p(x_2)$, полученные путём интегрирования СФПВ по одной из

переменных, называют *собственными* (маргинальными) ФПВ. Далее, если $p(x_1, x_2)$ интегрировать по двум переменным, получим

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x_1, x_2)dx_1\, dx_2 = F(\infty,\infty) = 1. \qquad (2.1.27)$$

Заметим также, что $F(-\infty,-\infty) = F(-\infty,x_2) = F(x_1,-\infty) = 0$.

Обобщение вышеуказанных соотношений на многомерные случайные величины очевидно. Предположим, что X_i, i=1, 2,..., n, являются случайными величинами с СИФР

$$F(x_1, x_2, ..., x_n) = P(X_1 \le x_1, X_2 \le x_2, ..., X_n \le x_n) = \\ = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\cdots\int_{-\infty}^{x_n} p(u_1, u_2, ..., u_n)\, du_1\, du_2 \ldots du_n, \qquad (2.1.28)$$

где $p(x_1, x_2, ..., x_n)$ – совместная ФПВ. Беря частные производные от $F(x_1, x_2, ..., x_n)$, заданной (2.1.28), получаем

$$p(x_1, x_2, ..., x_n) = \frac{\partial^n}{\partial x_1 \partial x_2 \ldots \partial x_n} F(x_1, x_2, ..., x_n). \qquad (2.1.29)$$

Любое число переменных в $p(x_1, x_2, ..., x_n)$ можно исключить путём интегрирования по этим переменным. Например, интегрируя по x_2 и x_3, получаем

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x_1, x_2, x_2, ..., x_n)dx_2\, dx_3 = p(x_1, x_4, .., x_n). \qquad (2.1.30)$$

Следует также, что $F(x_1,\infty,\infty,x_4,...,x_n) = F(x_1,x_4,x_5,...,x_n)$, а $F(x_1,-\infty,-\infty,x_4,..,x_n) = 0$.

Условные функции распределения вероятности. Рассмотрим две случайные величины X_1 и X_2 с СФПВ $p(x_1, x_2)$. Предположим, что мы желаем определить вероятность того, что случайная величина $X_1 < x_1$ при условии, что

$$x_2 - \Delta x_2 < X_2 \le x_2,$$

где Δx_2 – некоторое положительное приращение. Таким образом, мы желаем определить вероятность события $(X_1 \le x_1 | x_2 - \Delta x_2 < X_2 \le x_2)$. Используя соотношения, приведённые ранее для условной вероятности события, вероятность события $(X_1 \le x_1 | x_2 - \Delta x_2 < X_2 \le x_2)$ можно определить как вероятность совместного события $(X_1 \le x_1, x_2 - \Delta x_2 < X_2 \le x_2)$, делённую на вероятность события $(X_2 - \Delta X_2 < X_2 \le X_2)$. Таким образом,

$$P(X_1 \le x_1 | x_2 - \Delta x_2 < X_2 \le x_2) = \frac{\int_{-\infty}^{x_1}\int_{x_2-\Delta x_2}^{x_2} p(u_1, u_2)du_1\, du_2}{\int_{x_2-\Delta x_2}^{x_2} p(u_2)du_2} = \\ = \frac{F(x_1, x_2) - F(x_1, x_2 - \Delta x_2)}{F(x_2) - F(x_2 - \Delta x_2)}. \qquad (2.1.31)$$

Предполагая, что ФПВ $p(x_1, x_2)$ и $p(x_2)$ являются непрерывными функциями на интервале $(x_2 - \Delta x_2, x_2)$, мы можем делить числитель и знаменатель (2.1.31) на Δx_2 и взять предел при $\Delta x_2 \to 0$. Таким образом, мы получим

$$P(X \le x_1 | X_2 = x_2) \equiv F(x_1|x_2) \frac{\partial F(x_1, x_2)/\partial x_2}{\partial F(x_2)/\partial x_2} = \\ = \frac{\partial\left[\int_{-\infty}^{x_1}\int_{-\infty}^{x_2} p(u_1, u_2)du_1\, du_2\right]\Big/\partial x_2}{\partial\left[\int_{-\infty}^{x_2} p(u_2)du_2\right]\Big/\partial x_2} = \frac{\int_{-\infty}^{x_1} p(u_1, x_2)du_1}{p(x_2)}, \qquad (2.1.32)$$

что является условной ИФР случайной величины X_1 при заданной величине X_2. Заметим, что $F(-\infty|x_2)=0$ и $F(\infty|x_2)=1$. Путём дифференцирования правой части (2.1.32) по x_1 мы получаем условную ФПВ $p(x_1|x_2)$ в форме

$$p(x_1|x_2)=\frac{p(x_1,x_2)}{p(x_2)}. \qquad (2.1.33)$$

В качестве альтернативы мы можем выразить совместную ФПВ $p(x_1,x_2)$ через условную ФПВ $p(x_1|x_2)$ или $p(x_2|x_1)$ как

$$p(x_1,x_2)=p(x_1|x_2)p(x_2)=p(x_2|x_1)p(x_1). \qquad (2.1.34)$$

Обобщение соотношений, данных выше, на многомерные случайные величины не вызывает затруднений. Начиная с совместной ФПВ случайных величин X_i, i=1, 2,..., n, можно написать

$$p(x_1,x_2,...,x_n)=p(x_1,x_2,...,x_k|x_{k+1},...,x_n)p(x_{k+1},...,x_n), \qquad (2.1.35)$$

где k – любое целое число в пределах $1<k<n$. Совместная условная ИФР, соответствующая СФПВ $p(x_1,x_2,...,x_k|x_{k+1},...,x_n)$, равна

$$F(x_1,x_2,...,x_k|x_{k+1},...,x_n)=\frac{\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_k}p(u_1,u_2,...,u_k,x_{k+1},...x_n)du_1\,du_2\,...du_k}{p(x_{k+1},...x_n)}. \qquad (2.1.36)$$

Условные ИФР удовлетворяют соотношениям, ранее установленным для таких функций, таким как

$$F(\infty,x_2,...,x_k|x_{k+1},...,x_n)=F(x_2,x_3,...,x_k|x_{k+1},...,x_n),$$
$$F(-\infty,x_2,...,x_k|x_{k+1},...,x_n)=0.$$

Статистически независимые случайные величины. Мы уже определили статистическую независимость двух или больше событий из выборочного пространства S. Понятие статистической независимости может быть распространено на случайные величины, определённые на выборочном пространстве и полученные при комбинированном эксперименте или при повторении единственного эксперимента. Если эксперименты приводят к несовместным исходам, вероятность результата в одном эксперименте не зависит от результата в любом другом эксперименте[1]. Т.е. совместная вероятность результатов определяется произведением вероятностей, соответствующих каждому результату. Следовательно, случайные величины, соответствующие результатам в экспериментах, независимы в том смысле, что их СФПВ (или СИФР) определяется произведением соответствующих ФПВ (или ИФР). Следовательно, многомерные случайные величины статистически независимы, если, и только если

$$F(x_1,x_2,...,x_n)=F(x_1)F(x_2)\cdots F(x_n), \qquad (2.1.37)$$

или в качестве альтернативы

$$p(x_1,x_2,...,x_n)=p(x_1)p(x_2)\cdots p(x_n). \qquad (2.1.38)$$

[1] Правильнее было бы говорить о зависимых и независимых событиях безотносительно к способу проведения эксперимента (прп)

2.1.2. Функции от случайных величин

Проблему, которая часто возникает в практических приложениях теории вероятности, можно сформулировать так. Дана случайная величина X, которая характеризуется своей ФПВ $p(x)$, и надо найти ФПВ случайной величины $Y = g(X)$, где $g(X)$ – некоторая заданная функция от X. Если преобразование g от X к Y взаимно однозначное, определить $p(y)$ относительно просто. Однако, если преобразование не является взаимно однозначным, как в случае, например, когда $Y = X^2$, мы должны быть более внимательны в определении $p(y)$..

Пример 2.1.1. Рассмотрим случайную величину Y, определённую как

$$Y = aX + b, \tag{2.1.39}$$

где a и b – константы. Мы предположим, что $a>0$. Если $a<0$, подход тот же (см. задачу 2.3). Заметим, что это преобразование, иллюстрируемое рис. 2.1.4, (a), является линейным и монотонным.

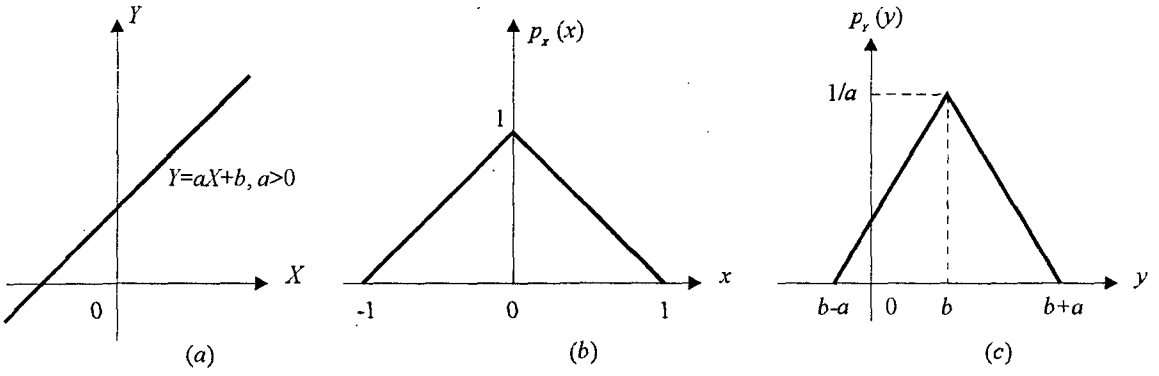


Рис. 2.1.4 Линейное преобразование случайной переменной и пример соответствующих ФПВдля X и Y

Пусть $F_x(x)$ и $F_y(y)$ определяют ИФР для X и Y соответственно[1]. Тогда

$$F_Y(y) = P(Y \le y) = P(aX + b \le y) = P\left(X \le \frac{y-b}{a}\right) = \int_{-\infty}^{(y-b)/a} p_X(x)dx = F_X\left(\frac{y-b}{a}\right) \tag{2.1.40}$$

Дифференцируя (2.1.40) по y, получаем зависимость между соответствующими ФПВ

$$p_Y(y) = \frac{1}{a} p_X\left(\frac{y-b}{a}\right). \tag{2.1.41}$$

Таким образом, (2.1.40) и (2.1.41) определяют ИФР и ФПВ случайной величины Y через ИФР и ФПВ случайной величины X для линейного преобразования (2.1.39). Чтобы проиллюстрировать это преобразование для определённой ФПВ $p(x)$, рассмотрим пример распределения на рис. 2.1.4,(b). Полученная ФПВ для преобразования (2.1.39) показана на рис. 2.1.4,(c).

Пример 2.1.2. Рассмотрим случайную величину Y, определённую как

$$Y = aX^3 + b, a > 0. \tag{2.1.42}$$

Как в примере (2.1.1), преобразование X в Y взаимно однозначное, следовательно,

[1] Чтобы избежать ошибки прн замене переменных, использованы индексы для соответствующих ФПВ и ИФР.

$$F_Y(y)=P(Y\le y)=P(aX^3+b\le y)=P\left[X\le\left(\frac{y-b}{a}\right)^{1/3}\right]=F_X\left[\left(\frac{y-b}{a}\right)^{1/3}\right]. \qquad (2.1.43)$$

Дифференцирование (2.1.43) по y даёт соотношение между двумя ФПВ

$$p_Y(y)=\frac{1}{3a[(y-b)/a]^{2/3}}p_X\left[\left(\frac{y-b}{a}\right)^{1/3}\right]. \qquad (2.1.44)$$

Пример 2.1.3. Случайная величина Y определена как

$$Y=aX^2+b,\ a>0. \qquad (2.1.45)$$

В отличие от примеров (2.1.1) и (2.1.2), связь между X и Y, иллюстрируемая рис. 2.1.5, теперь не взаимно однозначная. Чтобы найти ИФР для Y, заметим, что

$$F_Y(y)=P(Y\le y)=P(aX^2+b\le y)=P\left(|X|\le\sqrt{\frac{y-b}{a}}\right).$$

Следовательно,

$$F_Y(y)=F_X\left(\sqrt{\frac{y-b}{a}}\right)-F_X\left(-\sqrt{\frac{y-b}{a}}\right). \qquad (2.1.46)$$

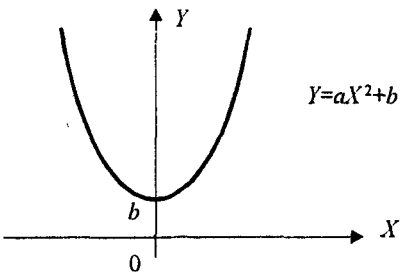


Рис. 2.1.5. Квадратичное преобразоваиие случайиой переменной X

Дифференцируя (2.1.46) по y, мы получим ФВП Y через ФВП X в виде

$$p_Y(y)=\frac{p_X\left[\sqrt{(y-b)/a}\right]}{2a\sqrt{[(y-b)/a]}}+\frac{p_X\left[-\sqrt{(y-b)/a}\right]}{2a\sqrt{[(y-b)/a]}} \qquad (2.1.47)$$

Для примера (2.1.3) мы замечаем, что уравнение $g(x)=ax^2+b=y$ имеет два вещественных решения:

$$x_1=\sqrt{\frac{y-b}{a}},\qquad x_2=-\sqrt{\frac{y-b}{a}},$$

и что $p_Y(y)$ содержит два слагаемых, соответствующих этим двум решениям:

$$p_Y(y)=\frac{p_X\left[x_1=\sqrt{(y-b)/a}\right]}{\left|g'\left[x_1=\sqrt{(y-b)/a}\right]\right|}+\frac{p_X\left[x_2=-\sqrt{(y-b)/a}\right]}{\left|g'\left[x_2=-\sqrt{(y-b)/a}\right]\right|}, \qquad (2.1.48)$$

где $g'(x)$ означает первую производную от $g(x)$ по x.

В общем случае предположим, что $x_1, x_2, \ldots, x_n$ являются вещественными корнями уравнения $g(x)=y$. Тогда ФПВ для случайной величины $Y=g(X)$ можно выразить так

$$p_Y(y)=\sum_{i=1}^{n}\frac{p_X(x_i)}{|g'(x_i)|}, \qquad (2.1.49)$$

где корни x_i, i=1, 2, …, n являются функциями от y.

Теперь рассмотрим функции от многомерных случайных величин. Предположим, что X_i, i=1, 2, …, n, являются случайными величинами с СФПВ $p_X(x_1, x_2, \ldots, x_n)$ и что Y_i, i=1, 2, …, n – другой ряд случайных величин, связанных с X_i функциями

$$Y_i = g_i(X_1, X_2, \ldots, X_n), \qquad i = 1, 2, \ldots, n. \tag{2.1.50}$$

Считаем, что $g_i(X_1, X_2, \ldots, X_n)$, i=1, 2, …, n, являются однозначными обратимыми функциями с непрерывными частными производными. Под «обратимыми» мы понимаем то, что X_i, i=1, 2, …, n, можно выразить как функции от Y_i, i=1, 2, …, n, в форме

$$X_i = g_i^{-1}(Y_1, Y_2, \ldots, Y_n), \quad i = 1, 2, \ldots, n, \tag{2.1.51}$$

причём обратные функции также считаются однозначными с непрерывными частными производными. Задача сводится к определению СФПВ Y_i, i=1, 2, …, n, т.е. $p_Y(y_1, y_2, \ldots, y_n)$, через заданную СФПВ $p_X(x_1, x_2, \ldots, x_n)$.

Чтобы найти нужное соотношение, положим, что R_X означает область в n–мерном пространстве случайных переменных X_i, i=1, 2, …, n, и что R_Y является областью взаимно-однозначного отображения в R_X, определенной функциями $Y_i = g_i(X_1, X_2, \ldots, X_n)$.

Очевидно, что

$$\iint_{R_Y} \ldots \int p_Y(y_1, y_2, \ldots, y_n) dy_1 dy_2 \ldots dy_n = \iint_{R_X} \ldots \int p_X(x_1, x_2, \ldots, x_n) dx_1 dx_2 \ldots dx_n. \tag{2.1.52}$$

Путем замены переменных в многомерном интеграле в правой части (2.1.52) по формулам

$$x_i = g_i^{-1}(y_1, y_2, \ldots, y_n) \equiv g_i^{-1}, \; i = 1, 2, \ldots, n.$$

получаем

$$\begin{aligned} &\iint_{R_Y} \ldots \int p_Y(y_1, y_2, \ldots, y_n) dy_1 dy_2 \ldots dy_n = \\ &\iint_{R_X} \ldots \int p_X(x_1 = g_1^{-1}, x_2 = g_2^{-1}, \ldots, x_n = g_n^{-1}) |J| dy_1 dy_2 \ldots dy_n, \end{aligned} \tag{2.1.53}$$

где J – якобиан преобразования, равный определителю

$$J = \begin{vmatrix} \dfrac{\partial g_1^{-1}}{\partial y_1} & \dfrac{\partial g_2^{-1}}{\partial y_1} & \cdots & \dfrac{\partial g_n^{-1}}{\partial y_1} \\ \vdots & \vdots & & \vdots \\ \dfrac{\partial g_1^{-1}}{\partial y_n} & \dfrac{\partial g_2^{-1}}{\partial y_n} & \cdots & \dfrac{\partial g_n^{-1}}{\partial y_n} \end{vmatrix}. \tag{2.1.54}$$

Следовательно, искомое соотношение для СФПВ всех Y_i, i=1, 2, …, n,

$$p_Y(y_1, y_2, \ldots, y_n) = p_X(x_1 = g_1^{-1}, x_2 = g_2^{-1}, \ldots, x_n = g_n^{-1}) |J|. \tag{2.1.55}$$

Пример 2.1.4. Важное функциональное соотношение между двумя рядами n-мерных случайных величин, которое часто встречается на практике, – линейное преобразование

$$Y_i = \sum_{j=1}^{n} a_{ij} X_j, \; i = 1, 2, \ldots, n, \tag{2.1.56}$$

где $\{a_{ij}\}$ – постоянные. Можно воспользоваться матричной формой преобразования

$$\mathbf{Y} = \mathbf{AX}, \tag{2.1.57}$$

где $\mathbf{X}$ и $\mathbf{Y}$ являются n-мерными векторами, а $\mathbf{A}$ – матрица размером $n \times n$. Предположим, что матрица $\mathbf{A}$ – невырожденная. Тогда матрица $\mathbf{A}$ обратима, и

$$\mathbf{X} = \mathbf{A}^{-1}\mathbf{Y}. \tag{2.1.58}$$

Эквивалентная скалярная запись

$$X_i = \sum_{j=1}^{n} b_{ij} Y_j, i = 1,2,\ldots,n, \qquad (2.1.59)$$

где $\{b_{ij}\}$ – элементы обратной матрицы $\mathbf{A}^{-1}$. Якобиан этого преобразования $J = 1/\det \mathbf{A}$. Следовательно,

$$P_Y(y_1, y_2, \ldots, y_n) = p_X\left(x_1 = \sum_{j=1}^{n} b_{1j} y_j, x_2 = \sum_{j=1}^{n} b_{2j} y_j, \ldots, x_n = \sum_{j=1}^{n} b_n y_j \right) \frac{1}{|\det A|}. \qquad (2.1.60)$$

2.1.3. Статистическое усреднение случайных величин

Усреднение играет важную роль для характеристики результатов эксперимента и случайных величин, определенных на выборочном пространстве эксперимента. В частности, представляют интерес первый и второй моменты одной случайной величины и совместные моменты, такие как корреляция и ковариация между парой случайных величин в многомерном ряде случайных величин. Также большой интерес представляет характеристическая функция случайной величины и совместные характеристические функции для многомерного ряда случайных величин. Этот раздел посвящается определению этих важных статистических средних.

Сначала мы рассмотрим случайную величину X, характеризуемую ФПВ $p(x)$. *Математическое ожидание* от X определяется как

$$E(X) \equiv m_x = \int_{-\infty}^{\infty} x p(x) dx, \qquad (2.1.61)$$

где $E(\cdot)$ означает математическое ожидание (статистическое усреднение). Это первый момент случайной величины X. В общем, n-й момент определяется как

$$E(X^n) = \int_{-\infty}^{\infty} x^n p(x) dx. \qquad (2.1.62)$$

Теперь предположим, что мы определяем случайную величину $Y = g(X)$, где $g(X)$ – некоторая произвольная функция от случайной величины X. Математическое ожидание Y определяется как

$$E(Y) = E[g(X)] = \int_{-\infty}^{\infty} g(x) p(x) dx. \qquad (2.1.63)$$

В частности, если $Y = (X - m_x)^n$, где m_X – математическое ожидание X, то

$$E(Y) = E\left[(X - m_x)^n\right] = \int_{-\infty}^{\infty} (x - m_x)^n p(x) dx. \qquad (2.1.64)$$

Это математическое ожидание названо n-м *центральным моментом* случайной величины X, так как это момент, взятый относительно среднего. Если $n = 2$, центральный момент называется *дисперсией* случайной величины и обозначается σ_X^2. Таким образом,

$$\sigma_X^2 = \int_{-\infty}^{\infty} (x - m_X)^2 p(x) dx. \qquad (2.1.65)$$

Этот параметр является мерой рассеяния случайной величины X. Раскрывая выражение $(x - m_X)^2$ в интеграле (2.1.65) и учитывая, что математическое ожидание от константы равно константе, получим выражение, которое определяет дисперсию через первый и второй моменты:

$$\sigma_X^2 = E(X^2) - [E(X)]^2 = E(X^2) - m_X^2. \qquad (2.1.66)$$

Для случая двух случайных величин X_1 и X_2 с СФПВ $p(X_1, X_2)$ мы определяем *совместный момент* как

$$E(X_1^k X_2^n) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1^k x_2^n p(x_1, x_2) dx_1 dx_2. \qquad (2.1.67)$$

и *совместный центральный момент* как

$$E\left[(X_1-m_1)^k(X_2-m_2)^n\right]= \\ =\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x_1-m_1)^k(x_2-m_2)^n p(x_1,x_2)dx_1\,dx_2, \tag{2.1.68}$$

где $m_i=E(X_i)$. С точки зрения приложений важное значение имеет совместный момент и совместный центральный момент, когда $k=n=1$. Эти совместные моменты называют *корреляцией* и *ковариацией* случайных величин X_1 и X_2.

При рассмотрении многомерных случайных величин мы можем определять совместные моменты произвольного порядка. Однако наиболее полезные для практических приложений моменты – это корреляция и ковариации между парами случайных величин. Для детализации предположим, что X_i, i=1, 2, ..., n, являются случайными величинами с СФПВ $p(x_1, x_2, \ldots, x_n)$. Пусть $p(x_i,x_j)$ – СФПВ случайных величин X_i и X_j. Тогда корреляция между X_i и X_j определяется совместным моментом

$$E(X_iX_j)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x_ix_j\,p(x_i,x_j)dx_idx_j\,, \tag{2.1.69}$$

а ковариация между X_i и X_j равна

$$\mu_{ij}\equiv E\left[(X_i-m_i)(X_j-m_j)\right]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x_i-m_i)(x_j-m_j)p(x_i,x_j)dx_idx_j= \\ =\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x_ix_jp(x_i,x_j)dx_idx_j-m_im_j=E(X_iX_j)-m_im_j. \tag{2.1.70}$$

Матрица размера $n\times n$ с элементами μ_{ij} называется *ковариационной матрицей* случайных величин X_i, i=1, 2, ..., n. Мы встретимся с ковариационной матрицей при обсуждении совместных гауссовских случайных величин в разделе 2.1.4.

Две случайные величины называют *некоррелированными*, если $E(X_iX_j)= =E(X_i)E(X_j)=m_im_j$. В этом случае их ковариация μ_{ij}=0. Заметим, что если X_i и X_j статистически независимы, они также не коррелированы. Однако, если X_i и X_j некоррелированы, они не обязательно статистически независимы.

Говорят, что две случайные величины *ортогональны*, если $E(X_iX_j)=0$. Заметим, что это условие имеет место, когда X_i и X_j не коррелированы и либо одна, либо обе случайные величины имеют нулевое среднее.

Характеристические функции. *Характеристическая функция* случайной величины X определяется как статистическое среднее

$$E(e^{j\nu X})\equiv\psi(j\nu)=\int_{-\infty}^{\infty}e^{j\nu x}p(x)dx, \tag{2.1.71}$$

где переменная ν вещественная, $j=\sqrt{-1}$. Заметим, что $\psi(j\nu)$ можно определить как преобразование Фурье[1)] от ФПВ $p(x)$. Тогда обратное преобразование Фурье дает

$$p(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\psi(j\nu)e^{-j\nu x}d\nu\,: \tag{2.1.72}$$

Очень полезное свойство характеристической функции – ее связь с моментами случайной величины. Заметим, что первая производная от (2.1.71) по ν

$$\frac{d\psi(j\nu)}{d\nu}=j\int_{-\infty}^{\infty}xe^{j\nu x}p(x)dx.$$

Вычисляя производную при ν=0, получаем для первого момента (среднего)

[1)] Обычно преобразованне Фурье от функции $g(u)$ определяется как $G(\nu)=\int_{-\infty}^{\infty}g(u)e^{-j\nu u}du$, которое отличается от (2.1.71) отрицательным знаком в экспоненте. Но это тривиальное отличие, и мы называем интеграл в (2.1.71) преобразованием Фурье.

$$E(X)=m_X=-j\frac{d\psi(jv)}{dv}\bigg|_{v=0}. \qquad (2.1.73)$$

Дифференцирование можно продолжить, и n-я производная от $\psi(jv)$ при v=0 определяет n-й момент:

$$E(X^n)=(-j)^n\frac{d^n\psi(jv)}{dv^n}\bigg|_{v=0}. \qquad (2.1.74)$$

Таким образом, моменты случайных величин можно определять через характеристические функции. С другой стороны, предположим, что характеристическую функцию можно представить рядом Тейлора относительно точки v=0, т.е.

$$\psi(jv)=\sum_{n=0}^{\infty}\left[\frac{d^n\psi(jv)}{dv^n}\right]_{v=0}\frac{v^n}{n!}. \qquad (2.1.75)$$

Используя соотношение (2.1.74) в (2.1.75), мы получаем выражение для характеристической функции через моменты в виде

$$\psi(jv)=\sum_{n=0}^{\infty}E(X^n)\frac{(jv)^n}{n!}. \qquad (2.1.76)$$

Характеристическая функция дает простой метод для определения ФПВ суммы независимых случайных величин. Чтобы это проиллюстрировать, предположим, что X_i, i=1, 2, ... n, – ряд статистически независимых случайных величин, и пусть

$$Y=\sum_{i=1}^{n}X_i. \qquad (2.1.77)$$

Задача сводится к нахождению ФПВ от Y. Мы определим ФПВ от Y, найдя сначала её характеристическую функцию, а затем вычислив обратное преобразование Фурье. Итак,

$$\begin{aligned}\psi_Y(jv)&=E\left(e^{jvY}\right)=\\&=E\left[\exp\left(jv\sum_{i=1}^{n}X_i\right)\right]=E\left[\prod_{i=1}^{n}\left(e^{jvX_i}\right)\right]=\\&=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left(\prod_{i=1}^{n}e^{jvx_i}\right)p(x_1,x_2,\ldots x_n)dx_1\,dx_2\ldots dx_n.\end{aligned} \qquad (2.1.78)$$

Так как случайные величины статистически независимы, $p(x_1,x_2,\ldots x_n)=$ $=p(x_1)p(x_2)\ldots p(x_n)$ и n-мерный интеграл в (2.1.78) сводится к произведению n простых интегралов, каждый из которых определяет характеристическую функцию одного X_i. Следовательно,

$$\psi_Y(jv)=\prod_{i=1}^{n}\psi_{X_i}(jv). \qquad (2.1.79)$$

Если помимо статистической независимости все X_i имеют одинаковое распределение, тогда все $\Psi_{Xi}(jv)$ идентичны.

Соответственно

$$\psi_Y(jv)=\left[\psi_{X_i}(jv)\right]^n. \qquad (2.1.80)$$

Окончательно ФПВ Y определяется обратным преобразованием Фурье, как дано в (2.1.72).

Поскольку характеристическая функция суммы n статистически независимых случайных величин равна произведению характеристических функций индивидуальных

случайных переменных X_i, i=1, 2, ... n, отсюда следует, что в области преобразования ФПВ Y является n-кратной сверткой ФПВ от X_i. Обычно n-кратную свёртку выполнить непосредственно более сложно, чем воспользоваться методом характеристической функции для нахождения распределения ФПВ для Y, как описано выше.

Если мы имеем дело с n-мерными случайными величинами, необходимо определить n-мерные преобразования Фурье от СФПВ. В частности, если X_i, i=1, 2... n, – случайные величины с ФПВ $p(x_1, x_2, \ldots x_n)$, *n-мерная характеристическая функция* определяется как

$$\psi_Y(jv_1, jv_2, \ldots jv_n) = E\left[\exp\left(j\sum_{i=1}^{n} v_i X_i\right)\right] = \\ = \int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty}\exp\left(j\sum_{i=1}^{n} v_i X_i\right)p(x_1, x_2, \ldots x_n)dx_1\, dx_2 \ldots dx_n. \quad (2.1.81)$$

Специальный интерес представляет двухмерная характеристическая функция

$$\psi(jv_1, jv_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{j(v_1x_1+v_2x_2)}p(x_1, x_2)\,dx_1 dx_2. \quad (2.1.82)$$

Заметим, что частные производные от $\psi(jv_1, jv_2)$ по v_1 и v_2 можно использовать для получения совместных моментов. Например, легко видеть, что

$$E(X_1X_2) = -\left.\frac{\partial^2\psi(jv_1, jv_2)}{\partial v_1\,\partial v_2}\right|_{v_1=v_2=0}. \quad (2.1.83)$$

Моменты более высоких порядков можно получить аналогичным образом.

2.1.4. Некоторые часто используемые распределения

В последующих главах мы встретим несколько различных типов случайных величин. В этом разделе мы перечислим эти новые часто встречающиеся случайные величины, их ФПВ, ПФР и моменты. Мы начнём с биномнального распределения, которое является распределением дискретной случайной величины, а затем представим распределение некоторых непрерывных случайных величин.

Биномиальное распределение. Пусть X – дискретная случайная величина, которая принимает два возможных значения, например X = 1 или X =0, с вероятностью p и $1-p$ соответственно. Соответствующая ФПВ для x показана на рис. 2.1.6.

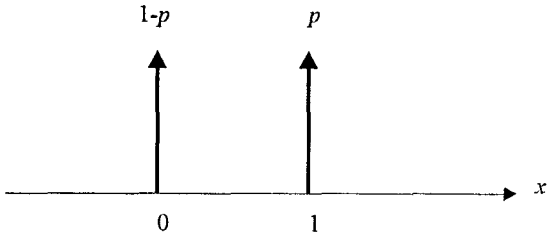


Рис. 2.1.6. Функция распределения вероятностей X

Теперь предположим, что

$$Y = \sum_{i=1}^{n} X_i,$$

где X_i, i=1,2...n, – статистически независимые и идентично распределенные случайные величины с ФПВ, показанной на рис. 2.1.6. Какова функция распределения Y?

Чтобы ответить на этот вопрос, заметим, что изначально Y – это ряд целых чисел от 0 до n. Вероятность того, что Y=0, просто равна вероятности того, что все X_i=0. Так как Y_i статистически независимы, то

$$P(Y = 0) = (1-p)^n.$$

Вероятность того, что Y=1, равна вероятности того, что одно слагаемое X_i=1, а остальные равны нулю. Так как это событие может возникнуть n различными путями,

$$P(Y=1)=np(1-p)^{n-1}.$$

Далее, вероятность того, что Y=k, равна вероятности того, что k значений X_i=1, а $n-k$ равна нулю. Так как теперь имеется

$$C_n^k=\binom{n}{k}\equiv\frac{n!}{k!(n-k)!} \tag{2.1.84}$$

различных комбинаций, которые приводят к результату $\{Y=k\}$, получаем

$$P(Y=k)=C_n^k p^k(1-p)^{n-k}, \tag{2.1.85}$$

где C_n^k – биномиальный коэффициент. Следовательно, ФПВ Y можно выразить как

$$p(y)=\sum_{k=0}^{n}P(Y=k)\delta(y-k)=\sum_{k=0}^{n}\binom{n}{k}p^k(1-p)^{n-k}\delta(y-k). \tag{2.1.86}$$

ИФР для Y

$$F(y)=P(Y\le y)=\sum_{k=0}^{[y]}\binom{n}{k}p^k(1-p)^{n-k}, \tag{2.1.87}$$

где $[y]$ означает наибольшее целое число m, такое, что $m\le y$.

ИФР (2.1.87) характеризует биномиальное распределение случайной величины.

Первые два момента Y равны

$$\begin{aligned}&E(Y)=np,\\&E(Y^2)=np(1-p)+n^2p^2,\\&\sigma^2=np(1-p),\end{aligned} \tag{2.1.88}$$

а характеристическая функция

$$\psi(jv)=(1-p+pe^{jv})^n. \tag{2.1.89}$$

Равномерное распределение. ФПВ и ИФР равномерно распределенной случайной величины X показан на рис. 2.1.7.

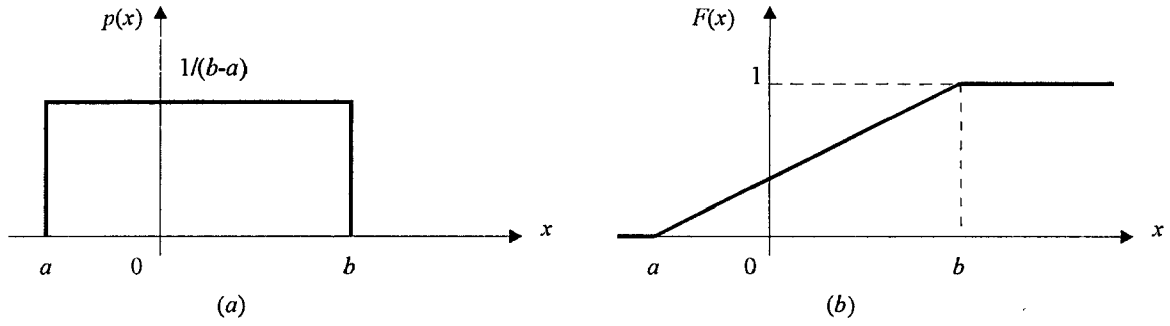


Рис. 2.1.7. Графики ФПВ и ИФР для равномерно распределенной случайной величины

Первые два момента X равны

$$\begin{aligned}&E(X)=\tfrac{1}{2}(a+b),\\&E(X^2)=\tfrac{1}{3}(a^2+b^2+ab),\\&\sigma^2=\tfrac{1}{12}(a-b)^2,\end{aligned} \tag{2.1.90}$$

а характеристическая функция равна

$$\psi(jv)=\frac{e^{jvb}-e^{jva}}{jv(b-a)} \tag{2.1.91}$$

Гауссовское распределение. ФПВ гауссовской или нормально распределенной случайной величины определяется формулой

$$p(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-(x-m_X)^2/2\sigma^2}, \tag{2.1.92}$$

где m_X – математическое ожидание, а σ^2 – дисперсия случайной величины. ИФР равна

$$F(x)=\int_{-\infty}^{x}p(u)du=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x}e^{-(u-m_x)^2/2\sigma^2}du=\frac{1}{2}\frac{2}{\sqrt{\pi}}\int_{-\infty}^{(x-m_x)/\sqrt{2}\sigma}e^{-t^2}dt=\frac{1}{2}+\frac{1}{2}\mathrm{erf}\left(\frac{x-m_x}{\sqrt{2}\sigma}\right), \tag{2.1.93}$$

где $\mathrm{erf}(x)$ – функция ошибок, которая определяется выражением

$$\mathrm{erf}(x)=\frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt. \tag{2.1.94}$$

ФПВ и ПФР иллюстрируются на рис. 2.1.8.

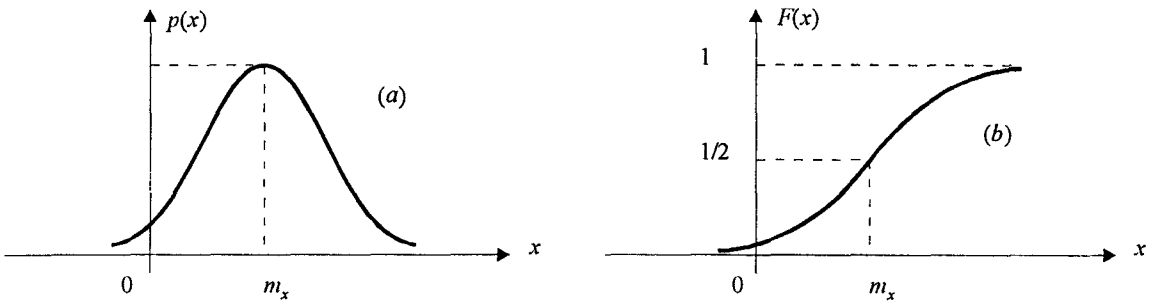


Рис. 2.1.8. Графики ФПВ (*a*) и ИФР (*b*) гауссовской случайной величины

ИФР $F(x)$ можно также выразить через дополнительную функцию ошибок, т.е.

$$F(x)=1-\frac{1}{2}\mathrm{erfc}\left(\frac{x-m_x}{\sqrt{2}\sigma}\right),$$

где

$$\mathrm{erfc}(x)=\frac{2}{\sqrt{\pi}}\int_x^\infty e^{-t^2}dt=1-\mathrm{erf}(x). \tag{2.1.95}$$

Заметим, что $\mathrm{erf}(-x)=-\mathrm{erf}(x)$, $\mathrm{erfc}(-x)=2-\mathrm{erfc}(x)$, $\mathrm{erf}(0)=\mathrm{erfc}(\infty)=0$ и $\mathrm{erf}(\infty)=\mathrm{erfc}(0)=1$. Для $x > m_x$ дополнительная функция ошибок пропорциональна площади под частью гауссовской ФПВ. Для больших значений X дополнительная функция ошибок $\mathrm{erf}(x)$ может быть аппроксимирована рядом

$$\mathrm{erfc}(x)=\frac{e^{-x^2}}{x\sqrt{\pi}}\left(1-\frac{1}{2x^2}+\frac{1\cdot3}{2^2x^4}+\frac{1\cdot3\cdot5}{2^3x^6}+\cdots\right), \tag{2.1.96}$$

причем ошибка аппроксимации меньше, чем последнее удерживаемое слагаемое.

Функция, которая обычно используется для площади под частью гауссовской ФПВ, обозначается через $Q(x)$ и определяется как

$$Q(x)=\frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-t^2/2}dt,\quad x\ge 0. \tag{2.1.97}$$

Сравнивая (2.1.95) и (2.1.97), находим

$$Q(x)=\frac{1}{2}\mathrm{erfc}\left(\frac{x}{\sqrt{2}}\right). \tag{2.1.98}$$

Характеристическая функция гауссовской случайной величины со средним m_X и дисперсией σ^2 равна

$$\psi(jv)=\int_{-\infty}^{\infty}e^{jvx}\left[\frac{1}{\sqrt{2\pi}\sigma}e^{-(x-m_x)^2/2\sigma^2}\right]dx=e^{jvm_x-(1/2)v^2\sigma^2}. \tag{2.1.99}$$

Центральные моменты гауссовской случайной величины равны

$$E\left[(X-m_x)^k\right]\equiv\mu_k=\begin{cases}1\cdot3\cdots(k-1)\sigma^k & (\text{четные } k)\\ 0 & (\text{нечетные } k),\end{cases} \tag{2.1.100}$$

а обычные моменты можно выразить через центральные моменты

$$E\left(X^k\right)=\sum_{i=0}^{k}\binom{k}{i}m_x^i\mu_{k-i}\,. \tag{2.1.101}$$

Сумма n статистически независимых гауссовских случайных величин также является гауссовской случайной величиной. Чтобы это продемонстрировать, предположим

$$Y=\sum_{i=1}^{n}X_i\,, \tag{2.1.102}$$

где X_i, i=1,2...n – независимые случайные величины со средними m_i и дисперсиями σ_i^2. Используя результат (2.1.79), мы находим, что характеристическая функция Y равна

$$\psi_Y(jv)=\prod_{i=1}^{n}\psi_{X_i}(jv)=\prod_{i=1}^{n}e^{jvm_i-v^2\sigma_i^2/2}=e^{jvm_Y-v^2\sigma_Y^2/2}\,, \tag{2.1.103}$$

где

$$m_Y=\sum_{i=1}^{n}m_i;\qquad \sigma_Y^2=\sum_{i=1}^{n}\sigma_i^2\,. \tag{2.1.104}$$

Следовательно, Y является гауссовской случайной величиной со средним m_Y и дисперсией σ_Y^2.

Хи-квадрат-распределение. Случайная величина с хи-квадрат-распределением порождается гауссовской случайной величиной, в том смысле, что ее формирование можно рассматривать как преобразование последней. Для конкретности, пусть $Y=X^2$, где X – гауссовская случайная величина. Тогда Y имеет хи-квадрат-распределение. Мы различаем два вида хи–квадрат распределения. Первое называется *центральным хи-квадрат-распределением*, и получается, когда X имеет нулевое среднее значение. Второе называется *нецентральным хи-квадрат-распределением*, и получается, когда X имеет ненулевое среднее значение.

Сначала рассмотрим центральное хи-квадрат-распределение. Пусть X – гауссовская случайная величина с нулевым средним и дисперсией σ^2. Поскольку $Y=X^2$, результат даётся функцией (2.1.47) с параметрами a=1 и b=0. Таким образом, получаем ФПВ Y в виде

$$p_Y(y)=\frac{1}{\sqrt{2\pi y}\sigma}e^{-y/2\sigma^2}\,,\quad y\ge 0. \tag{2.1.105}$$

ИФР для Y

$$F_Y(y)=\int_0^y p_Y(u)du=\frac{1}{\sqrt{2\pi}\sigma}\int_0^y\frac{1}{\sqrt{u}}e^{-u/2\sigma^2}du\,, \tag{2.1.106}$$

которое не может быть выражено в замкнутом виде. Характеристическая функция, однако, может быть выражена в замкнутой форме:

$$\psi(jv)=\frac{1}{\left(1-j2v\sigma^2\right)^{1/2}}\,. \tag{2.1.107}$$

Теперь предположим, что случайная величина Y определяется как

$$Y=\sum_{i=1}^{n}X_i^2\,, \tag{2.1.108}$$

где X_i, i=1, 2, ..., n, – статистически независимые и одинаково распределенные гауссовские случайные величины с нулевыми средними и дисперсией σ^2. Вследствие статистической независимости X_i характеристическая функция Y

$$\psi_Y(jv)=\frac{1}{\left(1-j2v\sigma^2\right)^{n/2}}\,. \tag{2.1.109}$$

Обратное преобразование этой характеристической функции дает ФПВ

$$p_Y(y)=\frac{1}{\sigma^n 2^{n/2}\Gamma\left(\frac{1}{2}n\right)}y^{n/2-1}e^{-y/2\sigma^2}\,,\ y\ge 0, \tag{2.1.110}$$

где $\Gamma(p)$ – гамма-функция, определённая как

$$\Gamma(p)=\int_0^\infty t^{p-1}e^{-t}dt\,,\ p>0,$$

$$\Gamma(p)=(p-1)!\,,\ p\text{ – целое число},\ p>0, \tag{2.1.111}$$

$$\Gamma(\tfrac{1}{2}) = \sqrt{\pi}, \quad \Gamma(\tfrac{3}{2}) = \tfrac{1}{2}\sqrt{\pi}.$$

Эта ФПВ является обобщением (2.1.105) и названа *хи-квадрат- (или гамма-) ФПВ с n степенями свободы*. Она иллюстрируется рис. 2.1.9.

Случай, когда n=2, определяет экспоненциальное распределение.

Первые два момента Y равны

$$\begin{aligned} E(Y) &= n\sigma^2, \\ E(Y^2) &= 2n\sigma^4 + n^2\sigma^4, \\ \sigma_y^2 &= 2n\sigma^4. \end{aligned} \tag{2.1.112}$$

ИФР Y равна

$$F_Y(y) = \int_0^y \frac{1}{\sigma^n 2^{n/2}\Gamma(\frac{1}{2}n)} u^{n/2-1} e^{-u/2\sigma^2} du, \quad y \geq 0. \tag{2.1.113}$$

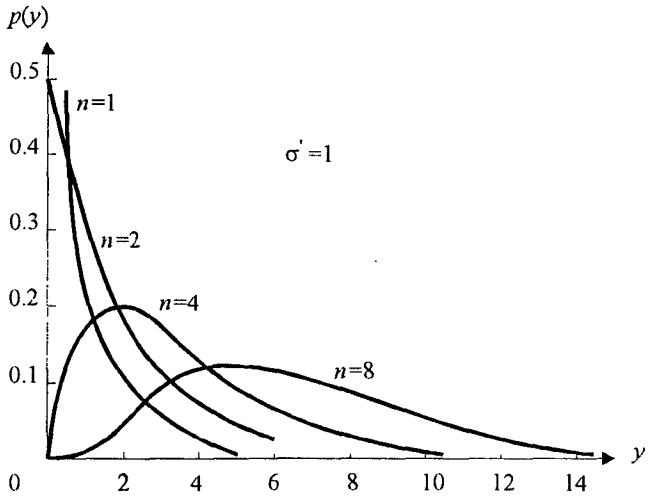


Рис. 2.1.9 Графики ФПВ для случайной величины с хи-квадрат-распределением для нескольких значений степеней свободы

Этот интеграл преобразуется к неполной гамма-функции, которая была табулирована Пирсоном (1965).

Если n четно, интеграл (2.11.113) можно выразить в замкнутом виде.

В частности, пусть $m = \frac{1}{2}n$, где m – целое. Тогда, используя повторно интегрирование по частям, получаем

$$F_Y(y) = 1 - e^{-y/2\sigma^2} \sum_{k=0}^{m-1} \frac{1}{k!}\left(\frac{y}{2\sigma^2}\right)^k, y = 0. \tag{2.1.114}$$

Теперь рассмотрим нецентральное хи-квадрат-распределение, которое является результатом возведения в квадрат гауссовской случайной величины с ненулевым средним. Если X – гауссовская случайная величина со средним m_X и дисперсией σ^2, случайная величина $Y=X^2$ имеет ФПВ

$$p_Y(y) = \frac{1}{\sqrt{2\pi y}\sigma} e^{-(y+m_x^2)/2\sigma^2} \operatorname{ch}\left(\frac{\sqrt{y}m_x}{\sigma^2}\right), \quad y \geq 0. \tag{2.1.115}$$

Этот результат получается при использовании (2.1.47) для гауссовской ФПВ с распределением (2.1.92). Характеристическая функция для ФПВ

$$\psi_Y(jv) = \frac{1}{(1 - j2v\sigma^2)^{1/2}} e^{jm_X^2 v/(1-j2v\sigma^2)}. \tag{2.1.116}$$

Для обобщения результатов предположим, что Y является суммой квадратов гауссовских случайных величин, определенных (2.1.108). Все X_i, $i = 1,2,...,n$, предполагаются статистически независимыми со средними m_i, $i = 1,2,...,n$, и одинаковыми дисперсиями σ^2. Тогда характеристическая функция, получаемая из (2.1.116), при использовании соотношения (2.1.79) равна

$$\psi_Y(jv)=\frac{1}{\left(1-j2v\sigma^2\right)^{n/2}}\exp\left(\frac{jv\sum_{i=1}^{n}m_i^2}{1-j2v\sigma^2}\right). \qquad (2.1.117)$$

Обратное преобразование Фурье от этой характеристической функции даёт ФПВ

$$p_Y(y)=\frac{1}{2\sigma^2}\left(\frac{y}{s^2}\right)^{(n-2)/4}e^{-(s^2+y)/2\sigma^2}I_{n/2-1}\left(\sqrt{y}\frac{s}{\sigma^2}\right),\ y\ge 0, \qquad (2.1.118)$$

где введено обозначение

$$s^2=\sum_{i=1}^{n}m_i^2, \qquad (2.1.119)$$

а $I_\alpha(x)$ –модифицированная функция Бесселя первого рода порядка α, которую можно представить бесконечным рядом

$$I_\alpha(x)=\sum_{k=0}^{\infty}\frac{(x/2)^{\alpha+2k}}{k!\Gamma(\alpha+k+1)},\ x\ge 0. \qquad (2.1.120)$$

ФПВ, определяемая (2.1.118), называется *нецентральным хи-квадрат-распределением с n степенями свободы*. Параметр s^2 назван *параметром нецентральности распределения*. ИФР для нецентрального хи-квадрат-распределения с n степенями свободы

$$F_Y(y)=\int_0^y\frac{1}{2\sigma^2}\left(\frac{u}{s^2}\right)^{(n-2)/4}e^{-(s^2+u)/2\sigma^2}I_{n/2-1}\left(\sqrt{u}\frac{s}{\sigma^2}\right)du. \qquad (2.1.121)$$

Этот интеграл не выражается в замкнутой форме. Однако, если $m=\frac{1}{2}n$ – целое число, ИФР можно выразить через обобщённую Q-функцию Маркума, которая определяется как

$$\begin{aligned}Q_m(a,b)&=\int_b^x x\left(\frac{x}{a}\right)^{m-1}e^{-(x^2+a^2)/2}I_{m-1}(ax)dx\\&=Q_1(a,b)+e^{(a^2+b^2)/2}\sum_{k=1}^{m-1}\left(\frac{b}{a}\right)^k I_k(ab),\end{aligned} \qquad (2.1.122)$$

где

$$Q_1(a,b)=e^{-(a^2+b^2)/2}\sum_{k=0}^{x}\left(\frac{a}{b}\right)^k I_k(ab),\ b>a>0 \qquad (2.1.123)$$

Если заменить переменную интегрирования u в (1.2.121) на x, причём $x^2=u/\sigma^2$, и положить, что $a^2=s^2/\sigma^2$, тогда можно легко найти

$$F_Y(y)=1-Q_m\left(\frac{s}{\sigma},\frac{\sqrt{y}}{\sigma}\right). \qquad (2.1.124)$$

В заключение заметим, что первые два момента для центрального хи-квадрат-распределения случайных величин равны

$$\begin{aligned}E(Y)&=n\sigma^2+s^2,\\E(Y^2)&=2n\sigma^4+4\sigma^2s^2+(n\sigma^2+s^2)^2,\\\sigma_y^2&=2n\sigma^4+4\sigma^2s^2.\end{aligned} \qquad (2.1.125)$$

Релеевское распределение. Релеевское распределение часто используется как модель для статистики сигналов, переданных через радиоканалы, таких как, например, в сотовой радиосвязи. Это распределение тесно связано с центральным хи-квадрат-распределением. Чтобы это проиллюстрировать, положим, что $Y=X_1^2+X_2^2$, где X_1 и X_2 – статистически независимые гауссовские случайные величины с нулевыми средними и одинаковой дисперсией σ^2. Из изложенного выше следует, что Y имеет хи-квадрат-распределение с двумя степенями свободы. Следовательно, ФПВ для Y

$$p_Y(y)=\frac{1}{2\sigma^2}e^{-y/2\sigma^2},\ y\ge 0. \qquad (2.1.126)$$

Теперь предположим, что мы определяем новую случайную величину

$$R=\sqrt{X_1^2+X_2^2}=\sqrt{Y}. \qquad (2.1.127)$$

Выполнив простые преобразования в (2.1.126), получим для ФПВ R

$$p_R(r) = \frac{r}{\sigma^2} e^{-r^2/2\sigma^2}, \; r = 0. \tag{2.1.128}$$

Это ФПВ для релеевской случайной величины. Соответствующая ИФР равна

$$F_R(r) = \int_0^r \frac{u}{\sigma^2} e^{-u^2/2\sigma^2} du = 1 - e^{-r^2/2\sigma^2}, \; r \geq 0. \tag{2.1.129}$$

Моменты от R равны

$$E(R^k) = \left(2\sigma^2\right)^{k/2} \Gamma\left(1 + \tfrac{1}{2}k\right), \tag{2.1.130}$$

а дисперсия

$$\sigma_r^2 = \left(2 - \tfrac{1}{2}\pi\right)\sigma^2. \tag{2.1.131}$$

Характеристическая функция для распределённой по Релею случайной величины

$$\Psi_R(jv) = \int_0^\infty \frac{r}{\sigma^2} e^{-r^2/2\sigma^2} e^{jvr} dr. \tag{2.1.132}$$

Этот интеграл можно выразить так:

$$\begin{aligned} \psi_R(jv) &= \int_0^\infty \frac{r}{\sigma^2} e^{-r^2/2\sigma^2} \cos vr\, dr + j \int_0^\infty \frac{r}{\sigma^2} e^{-r^2/2\sigma^2} \sin vr\, dr = \\ &= {}_1F_1\left(1, \frac{1}{2}, -\frac{1}{2}v^2\sigma^2\right) + j\sqrt{\frac{1}{2}}\pi v\sigma^2 e^{-v^2\sigma^2/2}, \end{aligned} \tag{2.1.133}$$

где ${}_1F_1(1,1/2,-a)$ – это вырожденная гипергеометрическая функция, определяемая как

$${}_1F_1(\alpha, \beta; x) = \sum_{k=0}^{\infty} \frac{\Gamma(\alpha + k)\Gamma(\beta) x^k}{\Gamma(\alpha)\Gamma(\beta + k) k!}, \quad \beta \neq 0, -1, -2, \ldots \tag{2.1.134}$$

Боули (1990) показал, что ${}_1F_1(1,1/2,-a)$ можно выразить как

$${}_1F_1\left(1, \frac{1}{2}; -a\right) = -e^{-a} \sum_{k=0}^{\infty} \frac{a^k}{(2k-1)k!}. \tag{2.1.135}$$

Как обобщение полученных выше выражений рассмотрим случайную величину

$$R = \sqrt{\sum_{i=1}^{n} X_i^2}, \tag{2.1.136}$$

где X_i, i=1, 2, ..., n, статистически независимые одинаково распределенные гауссовские случайные величины с нулевым средним. Ясно, что $Y=R^2$ имеет хи-квадрат-распределение с n степенями свободы. Его ФПВ задаётся формулой (2.1.100). Простые преобразования переменной в (2.1.110) приводят к ФПВ для R в виде

$$p_R(r) = \frac{r^{n-1}}{2^{(n-2)/2}\sigma^n \Gamma\left(\frac{1}{2}n\right)} e^{-r^2/2\sigma^2}, \quad r \geq 0. \tag{2.1.137}$$

Как следствие фундаментальной зависимости между центральным хи-квадрат-распределением и релеевским распределением, соответствующая ИФР достаточно простая. Так, для любого n ИФР для R можно представить в форме неполной гамма-функции. В специальном случае, когда n чётно, т.е. когда n=2m, ИФР для R может быть представлено в замкнутой форме

$$F_R(r) = 1 - e^{-r^2/2\sigma^2} \sum_{k=0}^{m-1} \frac{1}{k!} \left(\frac{r^2}{2\sigma^2}\right)^k, \quad r \geq 0. \tag{2.1.138}$$

В заключение приведём формулу для k-го момента R

$$E(R^k) = \left(2\sigma^2\right)^{k/2} \frac{\Gamma\left(\frac{1}{2}(n+k)\right)}{\Gamma\left(\frac{1}{2}n\right)}, \quad k \geq 0, \tag{2.1.139}$$

справедливую для любого n.

Распределение Райса. В то время как распределение Релея связано с центральным хи-квадрат-распределением, распределение Райса связано с нецентральным хи-квадрат-распределением. Чтобы проиллюстрировать эту связь, положим $Y=X_1^2+X_2^2$, где X_1 и X_2 – статистически независимые гауссовские случайные величины со средним m_i, i=1, 2 и одинаковой дисперсией σ^2. Из предыдущего рассмотрения мы знаем, что Y имеет нецентральное хи-квадрат-распределение с параметром отклонения $s^2=m_1^2+m_2^2$. ФПВ для Y получаем из (2.1.118), а при n=2 находим

$$p_Y(y)=\frac{1}{2\sigma^2}e^{-(s^2+y)/2\sigma^2}I_0\left(\sqrt{y}\frac{s}{\sigma^2}\right),\quad y\ge 0. \tag{2.1.140}$$

Теперь введём новую переменную $R=Y^{1/2}$.

ФПВ для R получается из (2.1.140) путём замены переменной

$$p_R(r)=\frac{r}{\sigma^2}e^{-(r^2+s^2)/2\sigma^2}I_0\left(\frac{rs}{\sigma^2}\right),\quad r\ge 0. \tag{2.1.141}$$

Функция (2.1.141) называется распределением Райса.

Как будет показано в гл. 5, эта ФПВ характеризует статистику огибающей гармонического сигнала, подверженному воздействию узкополосного гауссовского шума. Она также используется для статистики сигнала, переданного через некоторые радиоканалы. ИФР для R легко найти из (2.1.124) для случая, когда m=1. Это даёт

$$F_R(r)=1-Q_1\left(\frac{s}{\sigma},\frac{r}{\sigma}\right),\quad r\ge 0, \tag{2.1.142}$$

где $Q_1(a,b)$ определяется (2.1.123).

Для обобщения приведённого выше результата пусть R определяется (2.1.136), где X_i, i=1, 2, ... n - статистически независимые случайные величины со средними m_i, i=1, 2, ... n и одинаковыми дисперсиями σ^2. Случайная величина $R^2=Y$ имеет нецентральное хи-квадрат-распределение с n-степенями свободы и нецентральным параметром s^2, определяемое (2.1.119). Её ФПВ определятся (2.1.118), следовательно, ФПВ для R равна

$$p_R(r)=\frac{r^{n/2}}{\sigma^2 s^{(n-2)/2}}e^{-(r^2+s^2)/2\sigma^2}I_{n/2-1}\left(\frac{rs}{\sigma^2}\right),\quad r\ge 0, \tag{2.1.143}$$

а соответствующая ИФР

$$F_R(r)=P(R\le r)=P(\sqrt{Y}\le r)=P(Y\le r^2)=F_Y(r^2), \tag{2.1.144}$$

где $F_Y(r^2)$ определяется (2.1.121). В частном случае, когда $m=n/2$ – целое число, имеем

$$F_R(r)=1-Q_m\left(\frac{s}{\sigma},\frac{r}{\sigma}\right),\quad r\ge 0, \tag{2.1.145}$$

которое следует из (2.1.124). В заключение отметим, что k-й момент от R

$$E(R^k)=(2\sigma^2)^{k/2}e^{-s^2/2\sigma^2}\frac{\Gamma\left(\frac{1}{2}(n+k)\right)}{\Gamma\left(\frac{1}{2}n\right)}{}_1F_1\left(\frac{n+k}{2},\frac{n}{2};\frac{s^2}{2\sigma^2}\right),\quad k\ge 0, \tag{2.1.146}$$

где ${}_1F_1(\alpha,\beta;x)$ – вырожденная гипергеометрическая функция.

***m*-распределение Накагами.** И распределение Релея, и распределение Райса часто используется для описания статистики флуктуаций сигнала на выходе многопутевого канала с замираииями. Эта модель канала рассматривается в гл. 14. Другое распределение, часто используемое для характеристики статистики сигналов, передаваемых через многопутевые каналы с замираниями – это m-распределение Накагами. ФПВ для этого распределения дано Накагами (1960)

$$p_R(r)=\frac{2}{\Gamma(m)}\left(\frac{m}{\Omega}\right)^m r^{2m-1}e^{-mr^2/\Omega},\quad r\ge 0, \tag{2.1.147}$$

где Ω определяется как

$$\Omega=E(R^2), \tag{2.1.148}$$

а параметр m определяется как отношение моментов и назван параметром замираний:

$$m=\frac{\Omega^2}{E\left[\left(R^2-\Omega\right)^2\right]},\quad m\ge\frac{1}{2}. \tag{2.1.149}$$

Нормализованную версию для (2.1.147) можно получить путём введения другой случайной величины $X=R/\sqrt{\Omega}$ (см. задачу 2.15). n-й момент от R равен

$$E(R^n)=\frac{\Gamma\left(m+\frac{1}{2}n\right)}{\Gamma(m)}\left(\frac{\Omega}{m}\right)^{n/2}.$$

При m=1 можно видеть, что (2.1.147) приводит к распределению Релея. При значениях m удовлетворяющих условию $0{,}5\le m\le 1$, получаем ФПВ, которая имеет более протяжённые хвосты, чем при распределении Релея. При значениях $m>1$ хвосты ФПВ распределения Накагами убывают быстрее, чем для распределения Релея. Рисунок 2.1.10 иллюстрирует ФПВ для различных значений m.

Многомерное гауссовское распределение. Из многих многопараметрических или многомерных распределений, которые могут быть определены, многопараметрическое распределение Гаусса наиболее важное и наиболее часто используется на практике. Введем это распределение и рассмотрим его основные свойства.

Предположим, что X_i, i=1, 2, ... n являются гауссовскими случайными величинами со средними m_i, i=1, 2...n, дисперсиями σ_i^2, i=1, 2, ... n и ковариациями μ_{ij}, i= i=1, 2, ... n. Ясно, что $\mu_{ii}=\sigma_i^2$, i=1, 2, ... n. Пусть **M** – это матрица ковариаций размерности $n\times n$ с элементами $\{\mu_{ij}\}$. Пусть **X** определяет $n\times 1$ вектор-столбец случайных величин и пусть m_x означает $n\times 1$ вектор-столбец средних значений m_i, i=1, 2, ... n. Совместная ФПВ гауссовских случайных величин X_i, i=1, 2 ... n, определяется так

$$P(X_1,X_2,\ldots,X_n)=\frac{1}{(2\pi)^{n/2}(\det \mathbf{M})^{1/2}}\exp[-\frac{1}{2}(\mathbf{x}-\mathbf{m}_x)^{\mathrm{T}}\mathbf{M}^{-1}(\mathbf{x}-\mathbf{m}_x)] \qquad (2.1.150)$$

где $\mathbf{M}^{-1}$ – матрица, обратная **M**, и $\mathbf{x}^{\mathrm{T}}$ означает транспонирование **x**.

Характеристическая функция, соответствующая этой n-мерной совместной ФПВ

$$\psi(j\mathbf{v})=E(\mathrm{e}^{j\mathbf{v}^{\mathrm{T}}\mathbf{x}}),$$

где **V** – n-мерный вектор с элементами ν_i, i=1, 2, ... n.

Вычисление этого n-мерного преобразования Фурье даёт результат

$$\psi(j\mathbf{v})=\exp(j\mathbf{m}_x^{\mathrm{T}}\mathbf{v}-\tfrac{1}{2}\mathbf{v}^{\mathrm{T}}\mathbf{M}\mathbf{v}). \qquad (2.1.151)$$

Важнейший частный случай (2.1.150) – это бипараметрическая или двухмерная гауссовская ФПВ. Вектор средних $\mathbf{m}_x$ и ковариационная матрица **M** для этого случая

$$\mathbf{m}_x=\begin{bmatrix}m_1\\ m_2\end{bmatrix},\quad \mathbf{M}=\begin{bmatrix}\sigma_1^2 & \mu_{12}\\ \mu_{12} & \sigma_2^2\end{bmatrix}, \qquad (2.1.152)$$

где совместный центральный момент μ_{12} определяется так:

$$\mu_{12}=E[(X_1-m_1)(X_2-m_2)].$$

Удобно ввести нормированный коэффициент ковариации

$$\rho_{ij}=\frac{\mu_{ij}}{\sigma_i\sigma_j},\quad i\neq j, \qquad (2.1.153)$$

где ρ_{ij} удовлетворяет условию $0\le|\rho_{ij}|\le 1$. В двухмерном случае обычно опускают индексы в μ_{12} и ρ_{12}, тогда ковариациониая матрица выражается в виде

$$\mathbf{M}=\begin{bmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{bmatrix}. \qquad (2.1.154)$$

Обратная матрица

$$\mathbf{M}^{-1}=\frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)}\begin{bmatrix}\sigma_2^2 & -\rho\sigma_1\sigma_2\\ -\rho\sigma_1\sigma_2 & \sigma_1^2\end{bmatrix}, \qquad (2.1.155)$$

и $\det\mathbf{M}=\sigma_1^2\sigma_2^2(1-\rho^2)$.

Подставляя выражение $\mathbf{M}^{-1}$ в (2.1.150), получаем для двухмерной ФПВ гауссовских случайных величин

$$p(x_1,x_2)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left[-\frac{\sigma_2^2(x_1-m_1)^2-2\rho\sigma_1\sigma_2(x_1-m_1)(x_2-m_2)+\sigma_1^2(x_2-m_2)^2}{2\sigma_1^2\sigma_2^2(1-\rho^2)}\right]. \qquad (2.1.156)$$

Заметим, что если ρ=0, СФПВ $p(x_1,x_2)$ в (2.1.156) превращается в произведение $p(x_1)p(x_2)$, где $p(x_i)$, i=1, 2, – собственные ФПВ. Поскольку ρ является мерой корреляции между X_1 и X_2, то видим, что если гауссовские случайные величины не коррелированы, они также статистически независимы.

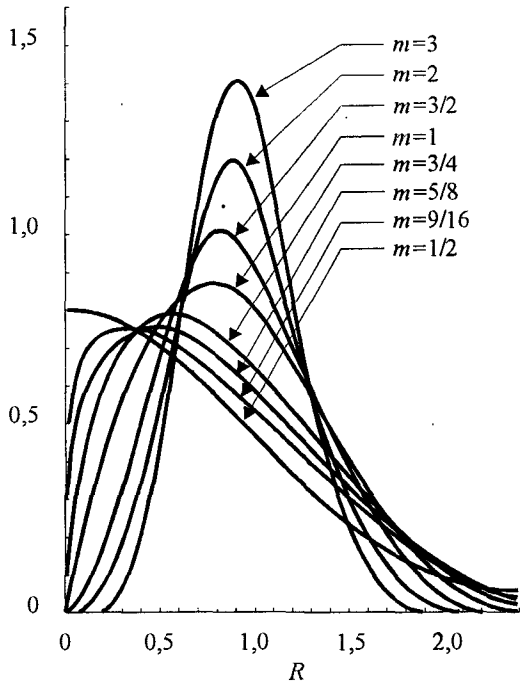


Рис. 2.1.10. Графики ФПВ для m-распределения при Ω=1. m - параметр замираний. (*Mijagaki* и др., *1978*)

Это важное свойство гауссовских случайных величин, которое, вообще говоря, не выполняется для других распределений. Оно распространяется на n-мерные гауссовские случайные величины непосредственно. Это означает, что если $\rho_{ij}=0$ при $i\neq j$, то случайные величины X_i, $i=$ $i=1, 2, \ldots n$ являются некоррелированными и, следовательно, статистически независимыми.

Теперь рассмотрим линейные преобразования n гауссовых случайных величин X_i, $i=$ $i=1, 2, \ldots n$, с вектором средних $\mathbf{m}_x$ и ковариационной матрицей **M**.

Пусть

$$\mathbf{Y}=\mathbf{AX}, \qquad (2.1.157)$$

где **A** – невырожденная матрица. Как показано раньше, якобиан этого преобразования $J=1/\det \mathbf{A}$. Подставляя $\mathbf{X}=\mathbf{A}^{-1}\mathbf{Y}$ в (2.1.150), получим СФПВ для **Y** в виде

$$p(\mathbf{y})=\frac{1}{(2\pi)^{n/2}(\det \mathbf{M})^{1/2}\det \mathbf{A}}\exp[-\frac{1}{2}(\mathbf{A}^{-1}\mathbf{y}-\mathbf{m}_x)^{\mathrm{T}}\mathbf{M}^{-1}(\mathbf{A}^{-1}\mathbf{y}-\mathbf{m}_x)]=$$
$$=\frac{1}{(2\pi)^{n/2}(\det \mathbf{Q})^{1/2}}\exp[-\frac{1}{2}(\mathbf{y}-\mathbf{m}_y)^{\mathrm{T}}\mathbf{Q}^{-1}(\mathbf{y}-\mathbf{m}_y)], \qquad (2.1.158)$$

где вектор $\mathbf{m}_y$ и матрица **Q** определяются так

$$\mathbf{m}_y=\mathbf{Am}_x$$
$$\mathbf{Q}=\mathbf{AMA}^{\mathrm{T}}. \qquad (2.1.159)$$

Таким образом, мы показали, что линейное преобразование ряда совместно гауссовских случайных величин приведёт к другому ряду также совместно гауссовских величин.

Предположим, что мы хотим с помощью линейных преобразований перейти к n статистически независимым случайным величинам. Как выбрать в этом случае матрицу **A**? Из предыдущего обсуждения мы знаем, что гауссовские случайные величины статистически независимы, если они попарно не коррелированы, т.е. если ковариационная матрица **Q** является диагональной. Следовательно, мы должны потребовать

$$\mathbf{AMA}^{\mathrm{T}}=\mathbf{D}, \qquad (2.1.160)$$

где **D** – диагональная матрица. Матрица **M** – это ковариационная матрица, следовательно, она положительно определённая. Одно решение (2.1.160) сводится к выбору ортогональной матрицы **A** ($\mathbf{A}^{\mathrm{T}}=\mathbf{A}^{-1}$), состоящей из столбцов, которые являются собственными векторами ковариационной матрицы **M**. Тогда **D** является диагональной матрицей с диагональными элементами, равными собственным векторам ковариационной матрицы **M**.

Пример 2.1.5. Рассмотрим двухмерную гауссовскую ФПВ с ковариационной матрицей

$$\mathbf{M}=\begin{vmatrix}1 & 1/2\\ 1/2 & 1\end{vmatrix}.$$

Определим преобразование **A**, которое приводит к некоррелированным случайным величинам. Сначала решим задачу о собственных значениях **M**. Характеристическое уравнение, которое их определяет,

$$\det(\mathbf{M}-\lambda\mathbf{I})=0,$$
$$(1-\lambda)^2-1/4=0,$$
$$\lambda=3/2,\ 1/2.$$

Далее мы определим два собственных вектора. Если **a** означает собственный вектор, имеем уравнение

$$(\mathbf{M}-\lambda\mathbf{I})\mathbf{a}=0.$$

При $\lambda_1=3/2$ и $\lambda_2=1/2$ мы получаем собственные векторы

$$\mathbf{a}_1=\begin{bmatrix}\sqrt{1/2}\\ \sqrt{1/2}\end{bmatrix},\quad \mathbf{a}_2=\begin{bmatrix}\sqrt{1/2}\\ -\sqrt{1/2}\end{bmatrix}.$$

Следовательно,

$$A=\sqrt{1/2}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}.$$

Легко показать, что $\mathbf{A}^{-1}=\mathbf{A}^{\mathrm{T}}$ и $\mathbf{AMA}^{\mathrm{T}}=\mathbf{D}$, где диагональные элементы **D** равны 3/2 и 1/2.

2.1.5. Верхняя граница для вероятностей «хвостов»

При определении характеристик систем цифровой связи часто необходимо определить площадь, ограниченную хвостами ФПВ. Мы назовём эту площадь вероятностью хвостов. В этом разделе мы представим две верхние границы для вероятности хвостов.

Первая, полученная из неравенства Чебышева, до некоторой степени грубая. Вторая, называемая границей Чернова, более плотная.

Неравенство Чебышева. Допустим, что X – произвольная случайная величина с ограниченным средним значением m_x и ограниченной дисперсией σ_x^2. Для произвольного положительного числа δ

$$P(|X-m_x|\ge\delta)\le\frac{\sigma_x^2}{\delta^2}. \qquad (2.1.161)$$

Это соотношение называется *неравенством Чебышева*. Доказательство этой границы относительно простое. Имеем

$$\sigma_x^2=\int_{-\infty}^{\infty}(x-m_x)^2p(x)dx\ge\int_{|x-m_x|\ge\delta}(x-m_x)^2p(x)dx\ge\delta^2\int_{|x-m_x|\ge\delta}p(x)dx=\delta^2P(|X-m_x|\ge\delta).$$

Таким образом, справедливость неравенства установлена.

Очевидно, что неравенство Чебышева непосредственно даёт верхнюю границу площади, ограниченной хвостами ФПВ $p(y)$, где $Y=X-m_x$, т.е. для площади под $p(y)$ в интервале $(-\infty,-\delta)$ и (δ,∞). Следовательно, неравенство Чебышева можно выразить в виде

$$1-[F_Y(\delta)-F_Y(-\delta)]\le\frac{\sigma_x^2}{\delta^2} \qquad (2.1.162)$$

или эквивалентным образом:

$$1-[F_X(m_x+\delta)-F_X(m_x-\delta)]\le\frac{\sigma_x^2}{\delta^2}. \qquad (2.1.163)$$

На границу Чебышева можно посмотреть с другой точки зрения. Используя случайную величину с нулевым средним $Y=X-m_x$, для удобства определим функцию $g(Y)$ в виде

$$g(Y)=\begin{cases}1 & (|Y|\ge\delta),\\ 0 & (|Y|<\delta).\end{cases} \qquad (2.1.163)$$

Поскольку функция $g(Y)$ равна 0 или 1 с вероятностью соответственно $P[|Y|<\delta]$ и $P[|Y|\ge\delta]$, её среднее значение

$$E[g(Y)]=P(|Y|\ge\delta) \qquad (2.1.165)$$

Теперь предположим, что мы используем для $g(Y)$ верхнюю квадратичную границу, т.е.

$$g(Y)\le\left(\frac{Y}{\delta}\right)^2. \qquad (2.1.166)$$

График для $g(Y)$ и верхняя граница показаны на рис. 2.1.11. Из графиков следует, что

$$E[g(Y)]\le E\left(\frac{Y^2}{\delta^2}\right)=\frac{E(Y^2)}{\delta^2}=\frac{\sigma_y^2}{\delta^2}=\frac{\sigma_x^2}{\delta^2}.$$

Так как $E[g(Y)]$ является вероятностью хвоста, как это следует из (2.1.165), мы получили границу Чебышева.

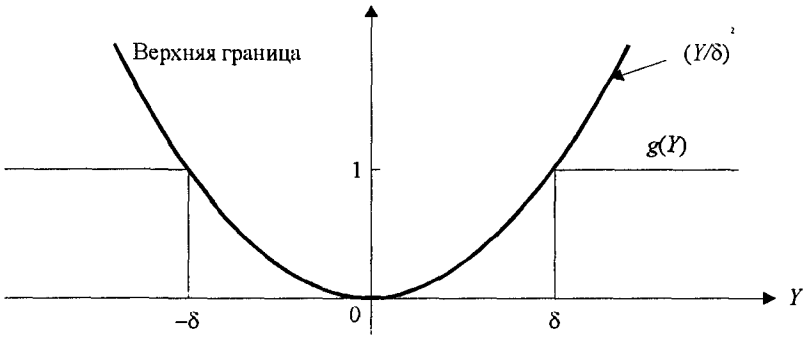


Рис. 2.1.11. Квадратичная верхняя граница для $g(Y)$, используемая для получения вероятности хвостов (граница Чебышева)

Для многих практических приложений эта чебышевская граница чрезмерно груба. Это можно объяснить неточностью квадратичной функции как верхней границы $g(Y)$. Имеется много других функций, которые можно использовать в качестве верхней границы $g(Y)$. В частности, граница Чернова часто оказывается более плотной.

Граница Чернова. Чебышевская граница, данная выше, включает площадь, ограниченную обоими хвостами ФПВ. В некоторых приложениях мы интересуемся лишь площадью, ограниченной одним хвостом: либо в интервале (δ,∞), либо в интервале $(-\infty,-\delta)$. В таком случае мы можем получить весьма плотную верхнюю границу путем огибания функции $g(Y)$ посредством экспоненты с параметром, который может оптимизировать верхнюю границу так плотно, насколько это возможно. Конкретно мы рассмотрим вероятность хвоста в интервале (δ,∞). Введем огибающую для $g(Y)$ из соотношения

$$g(Y) \le e^{\nu(Y-\delta)}, \qquad (2.1.167)$$

где $g(Y)$ теперь определена как

$$g(Y) = \begin{cases} 1 & (Y \ge \delta), \\ 0 & (Y < \delta), \end{cases} \qquad (2.1.168)$$

а $\nu \ge 0$ – параметр, который следует оптимизировать.

Графики для $g(Y)$ и экспоненциальной верхней границы даны на рис. 2.1.12.

Математическое ожидание $g(Y)$ равно

$$E[g(Y)] = P(Y \ge \delta) \le E\left[e^{\nu(Y-\delta)}\right]. \qquad (2.1.169)$$

Эта граница справедлива для любых $\nu \ge 0$.

Наиболее плотную верхнюю границу можно получить путем выбора значений, которые минимизируют $E\left[e^{\nu(Y-\delta)}\right]$.

Необходимое условие минимизации

$$\frac{d}{d\nu} E\left[e^{\nu(Y-\delta)}\right] = 0. \qquad (2.1.170)$$

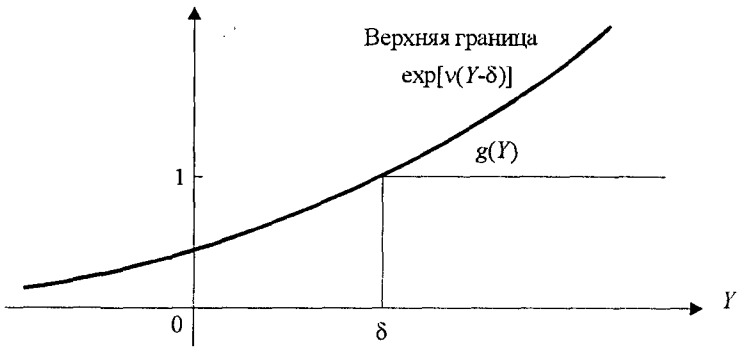


Рис. 2.1.12. Экспоненциальная верхняя граница для $g(Y)$, используемая для получения всроятности хвоста (граница Чернова)

Но можно изменить порядок дифференцирования и вычисление математического ожидания так, что

$$\frac{d}{d\nu}E\left[e^{\nu(Y-\delta)}\right]=E\left[\frac{d}{d\nu}e^{\nu(Y-\delta)}\right]=E\left[(Y-\delta)e^{\nu(Y-\delta)}\right]=$$
$$=e^{-\nu\delta}\left[E\left(Ye^{\nu Y}\right)-\delta E\left(e^{\nu Y}\right)\right]=0.$$

Следовательно, величина ν, которая обеспечивает плотную верхнюю границу определяется решением уравнения

$$E\left(Ye^{\nu Y}\right)-\delta E\left(e^{\nu Y}\right)=0. \qquad (2.1.171)$$

Пусть $\hat{\nu}$ является решением (2.1.171). Тогда из (2.1.169) следует, что верхняя граница для вероятности одного хвоста определяется так:

$$P(Y\geq\delta)\leq e^{-\hat{\nu}\delta}E\left[e^{\hat{\nu}\delta}\right]. \qquad (2.1.172)$$

Это – граница Чернова для вероятности верхнего хвоста дискретной или непрерывной случайной величины с нулевым средним[1]. Эту границу можно использовать, чтобы показать, что $Q(x)\leq e^{-x^2/2}$, где $Q(x)$ – площадь, определяющая вероятность хвоста гауссовской ФПВ (см. задачу 2.18).

Верхнюю границу для вероятности нижнего хвоста можно получить аналогичным путем:

$$P(Y\leq\delta)\leq e^{-\hat{\nu}\delta}E\left[e^{\hat{\nu}Y}\right], \qquad (2.1.173)$$

где $\hat{\nu}$ – решение (2.1.171) и $\delta<0$.

Пример 2.1.6. Рассмотрим ФПВ Лапласа

$$p(y)=\tfrac{1}{2}e^{-|y|}, \qquad (2.1.174)$$

которая проиллюстрирована на рис. 2.1.13.

Вычислим вероятность правого хвоста исходя из границы Чернова и сравним его с действительной вероятностью хвоста, которая равна

$$P(Y\geq\delta)=\int_{\delta}^{\infty}\tfrac{1}{2}e^{-y}\,dy=\tfrac{1}{2}e^{-\delta}. \qquad (2.1.175)$$

[1] Заметим, что $E(e^{\nu Y})$ для действительных ν не является характеристической функцией Y. Ее называют моментной производящей функцией Y.

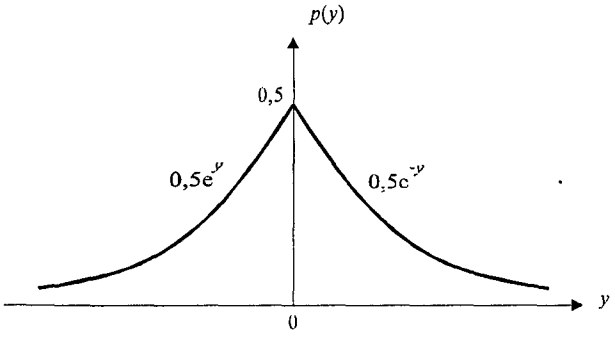


Рис. 2.1.13. График ФПВ для случайной величины, распределенной по Лапласу

Чтобы найти $\hat{\nu}$ из решения (2.1.171), мы должны определить моменты $E\left(Ye^{\nu Y}\right)$ и $E\left(e^{\nu Y}\right)$. Для ФПВ (2.1.174) находим

$$E\left(Ye^{\nu Y}\right)=\frac{2\nu}{(\nu+1)^2(\nu-1)^2},$$
$$E\left(e^{\nu Y}\right)=\frac{1}{(1+\nu)(1-\nu)}. \qquad (2.1.176)$$

Подставив эти моменты в (2.1.171), получим квадратное уравнение

$$\nu^2\delta+2\nu-\delta=0,$$

которое имеет решение

$$\hat{\nu}=\frac{-1\pm\sqrt{1+\delta^2}}{\delta}. \qquad (2.1.177)$$

Так как $\hat{\nu}$ должно быть положительной величиной, один из двух корней исключается. Таким образом,

$$\hat{\nu}=\frac{-1+\sqrt{1+\delta^2}}{\delta}. \qquad (2.1.178)$$

В заключение вычислим верхнюю границу в (2.1.172), ограничиваясь $E\left(e^{\hat{\nu}Y}\right)$, используя второе решение в (2.1.176) и подставляя для $\hat{\nu}$ решение (2.1.178). Результат равен

$$P(Y\ge\delta)=\frac{\delta^2}{2\left(-1+\sqrt{1+\delta^2}\right)}e^{1-\sqrt{1+\delta^2}}. \qquad (2.1.179)$$

Для $\delta>>1$ из (2.1.179) следует

$$P(Y\ge\delta)\le\tfrac{\delta}{2}e^{-\delta}. \qquad (2.1.180)$$

Заметим, что граница Чернова уменьшается экспоненциально с ростом δ. Следовательно, она тесно аппроксимирует действительную вероятность хвоста, определяемую (2.1.175). Напротив, чебышевская верхняя граница для вероятности верхнего хвоста, полученная как половина вероятности двух хвостов (вследствие симметрии ФПВ), равна

$$P(Y\ge\delta)\le\frac{1}{\delta^2}.$$

Следовательно, эта граница очень неточная.

Если случайная величина имеет ненулевое среднее, граница Чернова может быть обобщена, как мы сейчас покажем.

Если $Y = X - m_x$, имеем

$$P(Y \geq \delta) = P\left(X - m_x \geq \delta\right) = P\left(X \geq m_x + \delta\right) = P\left(X \geq \delta_m\right),$$

где $\delta_m = m_x + \delta$.

Так как $\delta > 0$, то $\delta_m > m_x$.

Пусть функция $g(X)$ определяется как

$$g(X) = \begin{cases} 1 & \left(X \geq \delta_m\right), \\ 0 & \left(X < \delta_m\right), \end{cases} \tag{2.1.181}$$

а верхняя граница – как

$$g(X) \leq e^{\nu\left(X - \delta_m\right)}. \tag{2.1.182}$$

Далее исследование идентично шагам, отражённым в (2.1.169)–(2.1.172). Окончательный результат таков:

$$P\left(X \geq \delta_m\right) \leq e^{-\hat{\nu}\delta_m} E\left[e^{\hat{\nu}X}\right], \tag{2.1.183}$$

где $\delta_m > m_x$ и $\hat{\nu}$ является решением уравнения

$$E\left[Xe^{\nu X}\right] - \delta_m E\left[e^{\nu X}\right] = 0. \tag{2.1.184}$$

Аналогичным путем можно найти границы Чебышева для вероятности нижнего хвоста. Для $\delta < 0$ имеем

$$P(X - m_x \leq \delta) = P(X \leq m_x + \delta) = P(X \leq \delta_m) \leq E\left(e^{\nu\left(X - \delta_m\right)}\right). \tag{2.1.185}$$

Из нашего предыдущего исследования очевидно, что (2.1.185) приводит к границе

$$P\left(X \leq \delta_m\right) \leq e^{-\hat{\nu}\delta_m} E\left[e^{\hat{\nu}X}\right], \tag{2.1.186}$$

где $\delta_m < m_x$ и $\hat{\nu}$ является решением (2.1.184).

2.1.6. Суммы случайных величин и центральная предельная теорема

Выше мы рассмотрели вопрос о нахождении ФПВ для суммы статистически независимых случайных величин. В этом разделе мы снова рассмотрим сумму статистически независимых случайных величин, но наш подход будет иным и не зависит от частных ФПВ случайных величин в сумме. В частности, предположим, что слагаемые суммы – статистически независимые и одинаково распределенные случайные величины, каждая из которых имеет ограниченные средние значения и ограниченную дисперсию.

Пусть Y определяется как нормированная сумма, называемая выборочным средним

$$Y = \frac{1}{n}\sum_{i=1}^{n} X_i. \tag{2.1.187}$$

Сначала определим верхние границы вероятности хвостов Y, а затем докажем очень важную теорему, определяющую ФПВ Y в пределе, когда n стремится к бесконечности.

Случайная величина Y, определенная (2.1.187), часто встречается при оценивании среднего случайной величины X по ряду наблюдений X_i, i=1, 2, ..., n. Другими словами, X_i могут рассматриваться как независимые выборочные реализации из распределения $F_X(x)$, а Y является оценкой среднего m_X.

Математическое ожидание Y равно

$$E(Y) = m_Y = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = m_X.$$

Дисперсия Y равна

$$\sigma_y^2 = E\left(Y^2\right) - m_y^2 = E\left(Y^2\right) - m_x^2 = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} E\left(X_i X_j\right) - m_x^2 = \frac{1}{n^2}\sum_{i=1}^{n} E\left(X_i^2\right) +$$

$$+\frac{1}{n^2}\sum_{\substack{i=1\\ i\neq j}}^{n}\sum_{j=1}^{n} E\left(X_i X_j\right) - m_x^2 = \frac{1}{n}\left(\sigma_x^2 + m_x^2\right) + \frac{1}{n^2} n(n-1) m_x^2 - m_x^2 = \frac{\sigma_x^2}{n}.$$

Если Y рассматривать как оценку среднего m, видим, что его математическое ожидание равно m, а его дисперсия уменьшается с ростом объема выборки X. Если X неограниченно возрастает, дисперсия стремится к нулю. Оценка параметра (в данном случае m), которая удовлетворяет условиям. что её математическое ожидание стремится к истинному значению параметра. а дисперсия строго к нулю, называется состоятельной оценкой.

Хвостовую вероятность случайной величины Y можно оценить сверху, используя границы, данные в разд. 2.1.5. Неравенство Чебышева применительно к Y имеет вид

$$P\left(\left|Y - m_y\right| \geq \delta\right) \leq \frac{\sigma_y^2}{\delta^2},$$

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - m_x\right| \geq \delta\right) \leq \frac{\sigma_x^2}{n\delta^2}. \qquad (2.1.188)$$

В пределе, когда $n\to\infty$, из (2.1.188) следует

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - m_x\right| \geq \delta\right) = 0. \qquad (2.1.189)$$

Следовательно, вероятность того, что оценка среднего отличается от истинного значения m_x больше, чем на δ ($\delta>0$), стремится к нулю, если n неограниченно растет. Это положение является формой закона больших чисел. Так как верхняя граница сходится к нулю относительно медленно, т.е. обратно пропорционально X, выражение (2.1.188) называют *слабым законом больших чисел*.

Если к случайной величине Y применить границу Чернова, содержащую экспоненциальную зависимость от n, тогда получим плотную верхнюю границу для вероятности одного хвоста. Следуя процедуре, изложенной в разд. 2.1.5, найдем, что вероятность хвоста для Y определяется выражением

$$P\left(Y - m_y \geq \delta\right) = P\left(\frac{1}{n}\sum_{i=1}^{n} X_i - m_x \geq \delta\right) =$$

$$= P\left(\sum_{i=1}^{n} X_i \geq n\delta_m\right) \leq E\left\{\exp\left[\nu\left(\sum_{i=1}^{n} X_i - n\delta_m\right)\right]\right\}, \qquad (2.1.190)$$

где $\delta_m = m_x + \delta$ и $\delta>0$. Но X_i, $i=1, 2, \ldots, n$ статистически независимы и одинаково распределены. Следовательно,

$$E\left\{\exp\left[\nu\left(\sum_{i=1}^{n} X_i - n\delta_m\right)\right]\right\} = e^{-\nu n\delta_m} E\left[\exp\left(\nu\sum_{i=1}^{n} X_i\right)\right] =$$

$$= e^{-\nu n\delta_m}\prod_{i=1}^{n} E\left(e^{\nu X_i}\right) = \left[e^{-\nu\delta_m} E\left(e^{\nu X}\right)\right]^n, \qquad (2.1.191)$$

где X – одна из величин X_i. Параметр ν, который дает наиболее точную верхнюю границу получается дифференцированием (2.1.191) и приравниванием производной нулю. Это ведет к уравнению

$$E\left(Xe^{\nu X}\right) - \delta_m E\left(e^{\nu X}\right) = 0 \qquad (2.1.192)$$

Обозначим решение (2.1.192) через $\hat{\nu}$. Тогда граница для вероятности верхнего хвоста

$$P\left(\frac{1}{n}\sum_{i=1}^{n} X_i \geq \delta_m\right) \leq \left[e^{-\hat{\nu}\delta_m} E\left(e^{\hat{\nu}X}\right)\right]^n, \quad \delta_m > m_x. \tag{2.1.193}$$

Аналогично мы найдем, что вероятность нижнего хвоста имеет границу

$$P(Y \leq \delta_m) \leq \left[e^{-\hat{\nu}\ \delta_m} E\left(e^{\hat{\nu}\ X}\right)\right]^n, \quad \delta_m < m_x. \tag{2.1.194}$$

Пример 2.1.7. Пусть X_i, i=1, 2, ..., n – ряд статистически независимых случайных величин, определенных так:

$$X_i = \begin{cases} 1 & \text{с вероятностью } p < \frac{1}{2} \\ -1 & \text{с вероятностью } 1-p. \end{cases}$$

Мы хотим определить плотную верхнюю границу вероятности того, что сумма от X_i больше, чем нуль. Так как p<1/2, то сумма будет иметь отрицательное значение для математического ожидания (среднего), следовательно, будем искать вероятность верхнего хвоста. При $\delta_m = 0$ в (2.1.193) имеем

$$P\left(\sum_{i=1}^{n} X_i \geq 0\right) \leq \left[E\left(e^{\hat{\nu}X}\right)\right]^n, \tag{2.1.195}$$

где $\hat{\nu}$ – решение уравнения

$$E\left(Xe^{\nu X}\right) = 0. \tag{2.1.196}$$

Теперь

$$E\left(Xe^{\nu X}\right) = -(1-p)e^{-\nu} + pe^{\nu} = 0.$$

Следовательно,

$$\hat{\nu} = \ln\left(\sqrt{\frac{1-p}{p}}\right). \tag{2.1.197}$$

Далее

$$E\left(e^{\hat{\nu}X}\right) = pe^{\hat{\nu}} + (1-p)e^{-\hat{\nu}}$$

Следовательно, для границы в (2.1.195) получаем

$$P\left(\sum_{i=1}^{n} X_i \geq 0\right) \leq \left[pe^{\hat{\nu}} + (1-p)e^{-\hat{\nu}}\right]^n = \left[p\sqrt{\frac{1-p}{p}} + (1-p)\sqrt{\frac{1-p}{p}}\right]^n \leq \left[4p(1-p)\right]^{n/2}. \tag{2.1.198}$$

Мы видим, что верхняя граница уменьшается экспоненциально с n, как ожидалось. В противоположность этому согласно границе Чебышева вероятность хвоста уменьшается обратно пропорционально n.

Центральная предельная теорема. В этом разделе рассмотрим чрезвычайно полезную теорему, касающуюся ИФР суммы случайных величин в пределе, когда число слагаемых суммы неограниченно возрастает. Имеется несколько версий этой теоремы. Докажем теорему для случая, когда случайные суммируемые величины X_i, i=1, 2, ..., n, статистически независимы и одинаково распределены, каждая из них имеет ограниченное среднее m_x и ограниченную дисперсию σ_x^2.

Для удобства определим нормированную случайную величину

$$U_i = \frac{X_i - m_x}{\sigma_x}, \quad i = 1, 2, ..., n.$$

Таким образом, U_i имеет нулевое среднее и единичную дисперсию.

Теперь пусть

$$Y=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_i. \qquad (2.1.199)$$

Так как каждое слагаемое суммы имеет нулевое среднее и единичную дисперсию, нормированная (множителем $1/\sqrt{n}$) величина Y имеет нулевое среднее и единичную дисперсию. Мы хотим определить ИФР для Y в пределе, когда $n\to\infty$.

Характеристическая функция Y равна

$$\psi_Y(j\upsilon)=E\left(e^{j\upsilon Y}\right)=E\left[\exp\left(\frac{j\upsilon\sum_{i=1}^{n}U_i}{\sqrt{n}}\right)\right]=\prod_{i=1}^{n}\psi_{U_i}\left(\frac{j\upsilon}{\sqrt{n}}\right)=\left[\psi_U\left(\frac{j\upsilon}{\sqrt{n}}\right)\right]^n, \qquad (2.1.200)$$

где U означает одну из U_i, которые одинаково распределены. Теперь разложим характеристическую функцию для U в ряд Тейлора.

$$\psi_U\left(\frac{j\upsilon}{\sqrt{n}}\right)=1+\frac{j\upsilon}{\sqrt{n}}E(U)-\frac{\upsilon^2}{n2!}E\left(U^2\right)+\frac{(j\upsilon)^3}{\left(\sqrt{n}\right)^3 3!}E\left(U^3\right)-\ldots \qquad (2.1.201)$$

Так как $E(U)=0$ и $E\left(U^2\right)=1$, (2.1.201) упрощается:

$$\psi_U\left(\frac{j\upsilon}{\sqrt{n}}\right)=1-\frac{\upsilon^2}{2n}+\frac{1}{n}R(\upsilon,n), \qquad (2.1.202)$$

где $R(\upsilon,n)/n$ означает остаток. Заметим, что $R(\upsilon,n)/n$ приближается к нулю, когда $n\to\infty$.

Подставив (2.1.202) в (2.1.200), получим характеристическую функцию Y в виде

$$\psi_Y(j\upsilon)=\left[1-\frac{\upsilon^2}{2n}+\frac{R(\upsilon,n)}{n}\right]^n. \qquad (2.1.203)$$

Взяв натуральный логарифм от (2.1.203), получим

$$\ln\psi_Y(j\upsilon)=n\ln\left[1-\frac{\upsilon^2}{2n}+\frac{R(\upsilon,n)}{n}\right]. \qquad (2.1.204)$$

Для малых значений x функцию $\ln(1+x)$ можно представить степенным рядом $\ln(1+x)=x-\frac{1}{2}x^2+\frac{1}{3}x^3-\ldots$.

Подставив это представление в (2.1.204), получим

$$\ln\psi_Y(j\upsilon)=n\left[-\frac{\upsilon^2}{2n}+\frac{R(\upsilon,n)}{n}-\frac{1}{2}\left(-\frac{\upsilon^2}{2n}+\frac{R(\upsilon,n)}{n}\right)^2+\ldots\right]. \qquad (2.1.205)$$

Окончательно, когда определим предел при $n\to\infty$, (2.1.205) приводит к

$$\lim_{n\to\infty}\ln\psi_Y(j\upsilon)=-\upsilon^2/2,$$

или, что эквивалентно,

$$\lim_{n\to\infty}\psi_Y(j\upsilon)=e^{-\upsilon^2/2}. \qquad (2.1.206)$$

Но это как раз характеристическая функция гауссовской случайной величины с нулевым средним и единичной дисперсией. Таким образом, мы имеем важный результат: ФПВ суммы статистически независимых и одинаково распределенных случайных величин с ограниченным средним и дисперсией приближается к гауссовской при $n\to\infty$. Этот результат известен как *центральная предельная теорема*.

Хотя мы предположили, что случайные величины в сумме распределены одинаково, это предположение можно ослабить при условии, что определённые дополнительные

ограничения все же накладываются на свойства случайных суммируемых величин. Имеется одна разновидность теоремы, например когда отказываются от предположения об одинаковом распределении случайных величин в пользу условия, накладываемого на третий абсолютный момент случайных величин суммы. Для обсуждения этой и других версий центральной предельной теоремы читатель отсылается к книге Крамера (1946).

2.2. СЛУЧАЙНЫЕ ПРОЦЕССЫ

Множество случайных явлений, которые имеют место в природе, являются функциями времени. Например, метеорологические явления, такие как случайные флуктуации температуры воздуха и давления воздуха, являются функциями времени. Напряжение теплового шума, создаваемое в резисторах электронных устройств, таких как радиоприёмник, также является функцией времени. Подобным образом, сигнал на выходе источника, который выдает информацию, характеризуется как случайный сигнал, меняющийся во времени. Звуковой сигнал который передается в телефонном канале, является примером такого сигнала. Все это примеры стохастических (случайных) процессов. При изучении систем цифровой связи мы используем случайные процессы для характеристики и моделирования сигналов, создаваемых источниками информации, для характеристики каналов связи, используемых для передачи информации, для характеристики шумов, создаваемых в приёмнике, и при синтезе оптимального приёмника для обработки принимаемого случайного сигнала.

В заданный момент времени t величина случайного процесса, будь то величина напряжения шума в резисторе или амплитуда сигнала, создаваемого звуковым источником, является случайной величиной. Таким образом, мы можем рассматривать случайный процесс как случайную величину; индексируемую параметром t. Мы будем обозначать такой процесс $X(t)$. Вообще говоря, параметр t непрерывен, в то время как X может быть или непрерывным или дискретным, в зависимости от характеристик источника, который создает случайный процесс.

Шумовое напряжение, создаваемое единственным резистором, или сообщение, выдаваемое источником информации, представляет единственную реализацию случайного процесса. Поэтому их называют *выборочной функцией* случайного процесса. Ряд всех возможных выборочных функций, например ряд всех шумовых напряжений, создаваемых резисторами, определяют ансамбль выборочных функций или, что эквивалентно, случайный процесс $X(t)$. Вообще говоря, число выборочных функций (реализаций) в ансамбле может быть очень большим; часто оно бесконечно.

Определяя случайный процесс $X(t)$ как ансамбль реализаций, мы можем рассмотреть значения процесса в ряде моментов времени t_1, t_2, t_3, ..., t_n, где n – положительное целое число. В общем, случайные величины $X_{t_i} = X(t_i)$, $i = 1, 2, \ldots n$ характеризуются статистически их СФПВ $p(x_{t_1}, x_{t_2}, \ldots, x_{t_n})$. Все вероятностные соотношения, определенные в разд. 2.1 для многомерных случайных величин, распространяются на случайные величины X_{t_i}, $i = 1, 2, \ldots n$.

Стационарные случайные процессы. Как указано выше, случайные величины X_{t_i}, $i = 1, 2, \ldots, n$, полученные из случайного процесса $X(t)$ для ряда моментов времени t_1, t_2, t_3, ..., t_n при некотором n, характеризуется статистически СФПВ $p(x_{t_1}, x_{t_2}, \ldots, x_{t_n})$. Рассмотрим другой ряд n случайных величин $X_{t_i+t} = X(t_i + t)$, $i = 1, 2, \ldots, n$, где t – произвольный временной сдвиг, одинаковый для всех i. Эти случайные величины

характеризуются СФПВ $p\left(x_{t_1+t}, x_{t_2+t}, \ldots, x_{t_n+t}\right)$. СФПВ случайных величин X_{t_i} и X_{t_i+t}, $i = 1, 2, \ldots n$, могут быть одинаковыми или нет. Если они одинаковы, т.е. если

$$p\left(x_{t_1}, x_{t_2}, \ldots, x_{t_n}\right) = p\left(x_{t_1+t}, x_{t_2+t}, \ldots, x_{t_n+t}\right) \qquad (2.2.1)$$

для всех t и n, случайный процесс называется *стационарным в строгом смысле*. Это значит, что статистика стационарного случайного процесса инвариантна к произвольному смещению по оси времени. С другой стороны, если СФПВ различны, случайный процесс называют *нестационарным*.

2.2.1. Статистические средние

Так же, как мы определили статистические средние для случайных величин, мы можем определить статистические средние для случайного процесса. Такие средние также называют средними по ансамблю. Пусть $X(t)$ определяет случайный процесс и пусть $X_{t_i} = X(t_i)$. Тогда n-й момент случайной величины X_{t_i} определяется как

$$E\left(X_{t_i}^n\right) = \int_{-\infty}^{\infty} x_{t_i}^n p\left(x_{t_i}\right) d x_{t_i} \qquad (2.2.2)$$

Вообще говоря, значение n-го момента будет зависеть от времени t_i, если ФПВ для X_{t_i} зависит от t_i. Однако, если процесс стационарен, $p(x_{t_i+t}) = p(x_{t_i})$ для всех t, то ФПВ не зависит от времени и, как следствие n-й момент не зависит от времени.

Далее мы рассмотрим две случайные величины $X_{t_i} = X(t_i)$, i=1, 2. Корреляция между X_{t1} и X_{t2} измеряется совместным моментом

$$E\left(X_{t_1} X_{t_2}\right) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_{t_1} x_{t_2} p\left(x_{t_1}, x_{t_2}\right) d x_{t_1} d x_{t_2}. \qquad (2.2.3)$$

Так как этот совместный момент зависит от выбора t_1 и t_2, его обозначают $\phi(t_1,t_2)$. Функцию $\phi(t_1,t_2)$ называют *автокорреляционной функцией случайного процесса*. Если процесс $X(t)$ стационарен, СФПВ пары (X_{t_1}, X_{t_2}) идентична СФПВ пары (X_{t_1+t}, X_{t_2+t}) для произвольного t. Это означает, что функция автокорреляции $X(t)$ не зависит от конкретных значений t_1 и t_2, но зависит от их разности t_1–t_2. Таким образом, для стационарного случайного процесса совместный момент (2.2.3) равен

$$E\left(X_{t_1} X_{t_2}\right) = \phi\left(t_1, t_2\right) = \phi\left(t_1 - t_2\right) = \phi(\tau), \qquad (2.2.4)$$

где $\tau = t_1 - t_2$ или, что эквивалентно, $t_2 = t_1 - \tau$. Если положить $t_2 = t_1 + \tau$, то

$$\phi(-\tau) = E\left(X_{t_1} X_{t_1+\tau}\right) = E\left(X_{t_1'} X_{t_1'-\tau}\right) = \phi(\tau)$$

Следовательно, $\phi(\tau)$ является чётной функцией. Заметим также, что $\phi(0) = E(X_t^2)$ определяет среднюю мощность процесса $X(t)$.

Существуют нестационарные процессы со свойствами: среднее значение процесса не зависит от времени (константа), а функция автокорреляции удовлетворяет условию $\phi\left(t_1, t_2\right) = \phi\left(t_1 - t_2\right)$.

Такие процессы называют *стационарными в широком смысле*. Следовательно, стационарность в широком смысле – это менее строгое условие, чем стационарность в строгом смысле. Если делается ссылка на стационарный случайный процесс при последующих обсуждениях, в которых участвуют функции корреляции, то везде имеется в виду менее строгое условие (стационарность в широком смысле).

С функцией автокорреляции связана функция автоковариации случайного процесса

которая определяется так

$$\mu(t_1,t_2)=E\left\{\left[X_{t_1}-m(t_1)\right]\left[X_{t_2}-m(t_2)\right]\right\}=\phi(t_1,t_2)-m(t_1)m(t_2), \qquad (2.2.5)$$

где $m(t_1)$ и $m(t_2)$ – средние для X_{t1} и X_{t2} соответственно. Если процесс стационарен, функция автоковариации упрощается и зависит только от $\tau=t_1-t_2$:

$$\mu(t_1,t_2)=\mu(t_1-t_2)=\mu(\tau)=\phi(\tau)-m^2, \qquad (2.2.6)$$

Совместные моменты более высокого порядка для двух или более случайных величин, полученных из случайного процесса, определятся очевидным образом. За исключением гауссовского случайного процесса, для которого моменты более высокого порядка можно выразить через моменты первого и второго порядка, моменты высокого порядка встречаются на практике очень редко.

Средние для гауссовских процессов. Предположим, что $X(t)$ является гауссовским случайным процессом. Следовательно, в момент времени $t=t_i$, $i=1, 2, \ldots, n$, случайные величины X_{ti}, $i=1, 2, \ldots, n$, являются совместно гауссовскими со средними значениями $m(t_i)$, $i=1, 2, \ldots, n$, и с автоковариациями

$$\mu(t_i,t_j)=E\left\{\left[X_{t_i}-m(t_i)\right]\left[X_{t_j}-m(t_j)\right]\right\}, \quad i,j=1, 2, \ldots, n. \qquad (2.2.7)$$

Если мы обозначим $n\times n$ матрицу ковариаций с элементами $\mu(t_i,t_j)$ через **M** и вектор средних значений через $\mathbf{m}_x$, тогда СФПВ случайных величин X_{t_i}, $i=1, 2, \ldots, n$ определяется формулой (2.1.150). Если гауссовский процесс стационарен, то $m(t_i)=m$ для всех t_i и $\mu(t_i,t_j)=\mu(t_i-t_j)$. Гауссовский случайный процесс полностью определяется средними значениями и функцией автокорреляции. Так как совместное гауссовское ФПВ зависит только от этих двух моментов, то следует, что если гауссовский процесс стационарен в широком смысле, он также стационарен в строгом смысле. Конечно, обратное утверждение верно для любого случайного процесса.

Средние для совместных случайных процессов. Пусть $X(t)$ и $Y(t)$ – два случайных процесса и пусть $X_{t_i}\equiv X(t_i)$, $i=1,2,\ldots,n$, и $Y_{t_i}\equiv Y(t_i)$, $i=1,2,\ldots,n$, представляют случайные величины в моменты $t_1>t_2>t_3>\ldots>t_n$ и $t'_1>t'_2>t'_3>\ldots>t'_m$ соответственно. Эти два процесса характеризуются статистически их СФПВ

$$p(x_{t_1},x_{t_2},\ldots,x_{t_n},y_{t'_1},y_{t'_2},\ldots,y_{t'_m})$$

для ряда моментов $t_1,t_2,\ldots,t_n$, $t'_1,t'_2,\ldots,t'_m$ и для положительных целых значений n и m. *Функция взаимной (кросс-) корреляции* $X(t)$ и $Y(t)$, обозначаемая $\phi_{xy}(t_1,t_2)$, находится как совместный момент

$$\phi_{xy}(t_1,t_2)=E(X_{t_1}Y_{t_2})=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x_{t_1}y_{t_2}p(x_{t_1},y_{t_2})dx_{t_1}dy_{t_2}, \qquad (2.2.8)$$

и функция взаимных ковариаций

$$\mu_{xy}(t_1,t_2)=\phi_{xy}(t_1,t_2)-m_x(t_1)\,m_y(t_2). \qquad (2.2.9)$$

Когда процессы совместно и индивидуально стационарны, имеем $\phi_{xy}(t_1,t_2)=\phi_{xy}(t_1-t_2)$ и $\mu_{xy}(t_1,t_2)=\mu_{xy}(t_1-t_2)$. В этом случае

$$\phi_{xy}(-\tau)=E(X_{t_1}Y_{t_1+\tau})=E(X_{t'_1-\tau}Y_{t'_1})=\phi_{yx}(\tau). \qquad (2.2.10)$$

Случайные процессы $X(t)$ и $Y(t)$ называются *статистически независимыми*, если, и только если

$$p(x_{t_1},x_{t_2},\ldots,x_{t_n},y_{t'_1},y_{t'_2},\ldots,y_{t'_m})=p(x_{t_1},x_{t_2},\ldots,x_{t_n})p(y_{t'_1},y_{t'_2},\ldots,y_{t'_m})$$

для всех значений t_i и t_i' и для всех положительных целых n и m. Процессы называются *некоррелированными*, если $\phi_{xy}(t_1,t_2)=E(X_{t_1})E(Y_{t_2})$. Следовательно, $\mu_{xy}(t_1,t_2)=0$.

Комплексный случайный процесс $Z(t)$ определяется как

$$Z(t)=X(t)+jY(t), \tag{2.2.11}$$

где $X(t)$ и $Y(t)$ являются случайными процессами. СФПВ случайных величин $Z_{t_i}\equiv Z(t_i)$, $i=1, 2, \ldots, n$, дается СФПВ компонентов (X_{t_i}, Y_{t_i}), $i=1, 2, \ldots, n$. Так, ФПВ, которая характеризует Z_{t_i}, $i=1, 2, \ldots, n$, равна

$$p(x_{t_1}, x_{t_2}, \ldots, x_{t_n}, y_{t_1}, y_{t_2}, \ldots, y_{t_n}).$$

Комплексный случайный процесс $Z(t)$ встречается при представлении узкополосного шума на выходе полосового фильтра через его эквивалентные низкочастотные компоненты. Важной характеристикой такого процесса является его автокорреляционная функция. Эта функция определяется так:

$$\begin{aligned}\phi_{zz}(t_1,t_2)&=\tfrac{1}{2}E\left(Z_{t_1}Z_{t_2}^*\right)=\tfrac{1}{2}E\left[\left(X_{t_1}+jY_{t_1}\right)\left(X_{t_2}-jY_{t_2}\right)\right]=\\&=\tfrac{1}{2}\left\{\phi_{xx}(t_1,t_2)+\phi_{yy}(t_1,t_2)+j\left[\phi_{yx}(t_1,t_2)-\phi_{xy}(t_1,t_2)\right]\right\},\end{aligned} \tag{2.2.12}$$

где $\phi_{xx}(t_1,t_2)$ и $\phi_{yy}(t_1,t_2)$ – функции автокорреляции $X(t)$ и $Y(t)$ соответственно, $\phi_{yx}(t_1,t_2)$ и $\phi_{xy}(t_1,t_2)$ – функции взаимной корреляции. Множитель 1/2 при определении функции автокорреляции комплексного случайного процесса является произвольным, но он дает математически удобную нормировку, как мы покажем в нашем рассмотрении таких процессов в гл. 4.

Если $X(t)$ и $Y(t)$ являются совместно и индивидуально стационарными, функция автокорреляции $Z(t)$

$$\phi_{zz}(t_1,t_2)=\phi_{zz}(t_1-t_2)=\phi_{zz}(\tau),$$

где $t_2=t_1-\tau$. Комплексное сопряжение для (2.2.12)

$$\phi_{zz}^*(\tau)=\tfrac{1}{2}E\left(Z_{t_1}^*Z_{t_1-\tau}\right)=\tfrac{1}{2}E\left(Z_{t_1'+\tau}^*Z_{t_1'}\right)=\phi_{zz}(-\tau). \tag{2.2.13}$$

Следовательно, $\phi_{zz}(\tau)=\phi_{zz}^*(-\tau)$.

Теперь допустим, что $Z(t)=X(t)+jY(t)$ и $W(t)=U(t)+jV(t)$ – это два комплексных случайных процесса. Функции взаимной корреляции $Z(t)$ и $W(t)$ определяется как

$$\begin{aligned}\phi_{zw}(t_1,t_2)&=\tfrac{1}{2}E\left(Z_{t_1}W_{t_2}^*\right)=\tfrac{1}{2}E\left[\left(X_{t_1}+jY_{t_1}\right)\left(U_{t_2}-jV_{t_2}\right)\right]=\\&=\tfrac{1}{2}\left\{\phi_{xu}(t_1,t_2)+\phi_{yv}(t_1,t_2)+j\left[\phi_{yu}(t_1,t_2)-\phi_{xv}(t_1,t_2)\right]\right\}.\end{aligned} \tag{2.2.14}$$

Если $X(t)$, $Y(t)$, $U(t)$ и $V(t)$ попарно стационарны, функция взаимной корреляции (2.2.14) является функцией от разности времени $\tau=t_1-t_2$. Наконец,

$$\phi_{zw}^*(\tau)=\tfrac{1}{2}E\left(Z_{t_1}^*W_{t_1-\tau}\right)=\tfrac{1}{2}E\left(Z_{t_1'+\tau}^*W_{t_1'}\right)=\phi_{wz}(-\tau). \tag{2.2.15}$$

2.2.2. Спектральная плотность мощности

Частотный состав сигнала – его базовая характеристика, которой один сигнал отличается от другого. В общем, сигнал можно классифицировать как имеющий или финитную (ненулевую) среднюю мощность (и неограниченную энергию) или ограниченную энергию. Частотный состав сигнала с ограниченной энергией получается как преобразование Фурье соответствующей функции времени. Если сигнал периодический, его энергия не ограничена и, следовательно, его преобразование Фурье не

существует. Для спектрального анализа периодический сигнал представляют рядом Фурье. Посредством такого представления коэффициенты Фурье определяют распределение мощности на различных дискретных частотных компонентах.

Стационарный случайный процесс имеет неограниченную энергию и, следовательно, его преобразование Фурье не существует. Спектральные характеристики случайного сигнала можно получить путем вычисления преобразования Фурье автокорреляционной функции, т.е. распределение мощности по частотам определяется формулой

$$\Phi(f) = \int_{-\infty}^{\infty} \phi(\tau)\, e^{-j2\pi f t} d\tau . \qquad (2.2.16)$$

Обратное преобразование Фурье дает

$$\phi(\tau) = \int_{-\infty}^{\infty} \Phi(f) e^{j2\pi f t} df . \qquad (2.2.17)$$

Можно видеть, что

$$\phi(0) = \int_{-\infty}^{\infty} \Phi(f) df = E\left(|X_t|^2\right) \geq 0 . \qquad (2.2.18)$$

Поскольку $\Phi(0)$ определяет среднюю мощность случайного сигнала, которая равна площади под кривой $\Phi(f)$, то $\Phi(f)$ определяет распределение мощности как функция частоты. Поэтому $\Phi(f)$ называют *спектральной плотностью мощности* случайного процесса.

Если случайный процесс вещественный, $\phi(\tau)$ – вещественная и четная функция и, следовательно, $\Phi(f)$ – также вещественная и четная функция. С другой стороны, если процесс комплексный, $\phi(\tau) = \phi^*(-\tau)$ и, следовательно,

$$\begin{aligned} \Phi^*(f) &= \int_{-\infty}^{\infty} \phi^*(\tau)\, e^{j2\pi f t} d\tau = \\ &= \int_{-\infty}^{\infty} \phi^*(-\tau)\, e^{-j2\pi f t} d\tau = \int_{-\infty}^{\infty} \phi(\tau)\, e^{-j2\pi f t} d\tau = \Phi(f). \end{aligned} \qquad (2.2.19)$$

Значит, $\Phi(f)$ – вещественная функция.

Спектральную плотность мощности можно определить и для совместно стационарных процессов $X(t)$ и $Y(t)$, которые имеют взаимную функцию корреляции $\phi_{x,y}(\tau)$. Преобразование Фурье от $\phi_{x,y}(\tau)$, т.е.

$$\Phi_{xy}(f) = \int_{-\infty}^{\infty} \phi_{xy}(\tau)\, e^{-j2\pi f t} d\tau , \qquad (2.2.20)$$

называют *взаимной спектральной плотностью мощности*.

Если мы возьмем сопряженные значения двух частей (2.2.20), получим

$$\Phi_{xy}^*(f) = \int_{-\infty}^{\infty} \phi_{xy}^*(\tau)\, e^{j2\pi f t} d\tau = \int_{-\infty}^{\infty} \phi_{xy}^*(-\tau)\, e^{-j2\pi f t} d\tau = \int_{-\infty}^{\infty} \phi_{yx}(\tau)\, e^{-j2\pi f t} d\tau = \Phi_{yx}(f). \qquad (2.2.21)$$

Это соотношение справедливо в любом случае. Однако если $X(t)$ и $Y(t)$ – вещественные случайные процессы, то

$$\Phi_{xy}^*(f) = \int_{-\infty}^{\infty} \phi_{xy}(\tau)\, e^{j2\pi f t} d\tau = \Phi_{yx}(-f). \qquad (2.2.22)$$

Объединяя результаты (2.2.21) и (2.2.22), находим, что взаимная спектральная плотность мощности двух вещественных процессов удовлетворяет условию

$$\Phi_{yx}(f) = \Phi_{xy}(-f). \qquad (2.2.23)$$

2.2.3. Отклик линейной стационарной системы на случайный входной сигнал

Рассмотрим линейную стационарную систему (фильтр), которая характеризуется своей импульсной характеристикой $h(t)$ или, что эквивалентно, своей частотной характеристикой $H(f)$, где $h(t)$ и $H(f)$ связаны парой преобразования Фурье. Пусть $x(t)$ означает входной, а

$y(t)$ – выходной сигналы системы. Выход системы можно выразить интегралом свертки

$$y(t) = \int_{-\infty}^{\infty} h(\tau)\, x(t-\tau)\, d\tau . \qquad (2.2.24)$$

Теперь предположим, что $x(t)$ является реализацией стационарного случайного процесса $X(t)$. Тогда выход $y(t)$ является реализацией случайного процесса $Y(t)$. Мы хотим определить математическое ожидание и функцию корреляции выхода.

Поскольку свертка – это линейная операция над входным сигналом, математическое ожидание интеграла равно интегралу от математического ожидания подынтегральной функции. Таким образом, математическое ожидание $Y(t)$

$$\begin{aligned} m_y &= E[Y(t)] = \int_{-\infty}^{\infty} h(\tau)\, E[X(t-\tau)]\, d\tau = \\ &= m_x \int_{-\infty}^{\infty} h(\tau)\, d\tau = m_x\, H(0), \end{aligned} \qquad (2.2.25)$$

где $H(0)$ – коэффициент передачи (передаточная функция) линейной системы при $f=0$. Следовательно, среднее значение выходного процесса постоянно.

Функция корреляции выхода

$$\phi_{yy}(t_1, t_2) = \tfrac{1}{2} E\left(Y_{t_1} Y_{t_2}^*\right) = \frac{1}{2} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(\beta)\, h^*(\alpha)\, E\left[X(t_1-\beta) X^*(t_2-\alpha)\right] d\alpha\, d\beta =$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(\beta)\, h^*(\alpha)\, \phi_{xx}(t_1 - t_2 + \alpha - \beta)\, d\alpha\, d\beta .$$

Последнее выражение показывает, что двойной интеграл является функцией разности отсчетов времени t_1–t_2. Другими словами, если входной процесс стационарный, выходной процесс также стационарен. Следовательно,

$$\phi_{yy}(\tau) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h^*(\alpha)\, h(\beta)\, \phi_{xx}(\tau + \alpha - \beta)\, d\alpha\, d\beta . \qquad (2.2.26)$$

Взяв преобразование Фурье от обеих частей (2.2.26), получим спектральную плотность мощности выходного процесса в виде

$$\begin{aligned} \Phi_{yy}(f) &= \int_{-\infty}^{\infty} \phi_{yy}(\tau)\, e^{-j2\pi f\tau} d\tau = \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h^*(\alpha)\, h(\beta)\, \phi_{xx}(\tau + \alpha - \beta)\, e^{-j2\pi f\tau} d\tau\, d\alpha\, d\beta = \Phi_{xx}(f)\left|H(f)\right|^2 . \end{aligned} \qquad (2.2.27)$$

Таким образом, мы имеем важный результат, заключающийся в том, что спектральная плотность мощности выходного сигнала равна произведению спектральной плотности мощности входного сигнала и квадрата модуля частотной характеристики системы.

При расчёте автокорреляционной функции $\phi_{yy}(\tau)$ обычно легче определить спектральную плотность мощности $\Phi_{yy}(f)$ и затем вычислить обратное преобразование Фурье. Таким образом, имеем

$$\phi_{yy}(\tau) = \int_{-\infty}^{\infty} \Phi_{yy}(f)\, e^{j2\pi f\tau} df = \int_{-\infty}^{\infty} \Phi_{xx}(f)\left|H(f)\right|^2 e^{j2\pi f\tau} df . \qquad (2.2.28)$$

Видим, что средняя мощность выходного сигнала

$$\phi_{yy}(0) = \int_{-\infty}^{\infty} \Phi_{xx}(f)\left|H(f)\right|^2 df . \qquad (2.2.29)$$

Так как $\phi_{yy}(0) = E\left(\left|Y_t\right|^2\right)$, то

$$\int_{-\infty}^{\infty} \Phi_{xx}(f)\left|H(f)\right|^2 df \geq 0 .$$

Допустим, что $\left|H(f)\right|^2 = 1$ для некоторого малого интервала $f_1 \leq f \leq f_2$ и $H(f) = 0$ вн этого интервала. Тогда

$$\int_{f_1}^{f_2} \Phi_{xx}(f)\,df \geq 0.$$

Но это возможно тогда и только тогда, когда $\Phi_{xx}(f) \geq 0$ для всех f.

Пример 2.2.1. Предположим, что фильтр нижних частот (ФНЧ), показанный на рис. 2.2.1, находится под воздействием случайного процесса $X(t)$ со спектральной плотностью мощности

$$\Phi_{xx}(f) = \tfrac{1}{2}N_0 \quad \text{для всех } f.$$

Случайный процесс с одинаковой спектральной плотностью на всех частотах называется *белым шумом*. Определим спектральную плотность мощности выходного процесса. Передаточная функция ФНЧ

$$H(f) = \frac{R}{R + j2\pi fL} = \frac{1}{1 + j2\pi fL/R},$$

и, следовательно,

$$|H(f)|^2 = \frac{1}{1 + (2\pi L/R)^2 f^2}. \qquad (2.2.30)$$

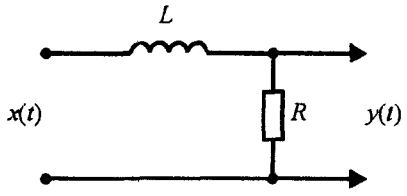


Рис. 2.2.1. Пример низкочастотного фильтра

Спектральная плотность мощности процесса на выходе

$$\Phi_{yy}(f) = \frac{N_0}{2} \cdot \frac{1}{1 + (2\pi L/R)^2 f^2}. \qquad (2.2.31)$$

Эту спектральную плотность иллюстрирует рис. 2.2.2.

Обратное преобразование Фурье определяет функцию автокорреляции

$$\phi_{yy}(\tau) = \int_{-\infty}^{\infty} \frac{N_0}{2} \cdot \frac{1}{1 + (2\pi L/R)^2 f^2} e^{j2\pi f\tau} df = \frac{RN_0}{4L} e^{-(R/L)|\tau|}. \qquad (2.2.32)$$

Автокорреляционная функция $\phi_{yy}(\tau)$ показана на рис. 2.2.3. Заметим, что второй момент процесса $Y(t)$ равен $\phi_{yy}(0) = RN_0/4L$.

В качестве заключительного упражнения определим взаимную корреляционную функцию между $Y(t)$ и $X(t)$, где $X(t)$ – сигнал на входе, а $Y(t)$ – сигнал на выходе линейной системы. Имеем

$$\phi_{yx}(t_1, t_2) = \tfrac{1}{2}E\left(Y_{t_1} X^*_{t_2}\right) = \tfrac{1}{2}\int_{-\infty}^{\infty} h(\alpha) E\left[X(t_1 - \alpha) X^*(t_2)\right] d\alpha =$$

$$= \int_{-\infty}^{\infty} h(\alpha)\phi_{xx}(t_1 - t_2 - \alpha)\,d\alpha = \phi_{yx}(t_1 - t_2).$$

Следовательно, случайные процессы $X(t)$ и $Y(t)$ совместно стационарны. Обозначив $t_1 - t_2 = \tau$, имеем

$$\phi_{yx}(\tau) \int_{-\infty}^{\infty} h(\alpha)\,\phi_{xx}(\tau - \alpha)\,d\alpha. \qquad (2.2.33)$$

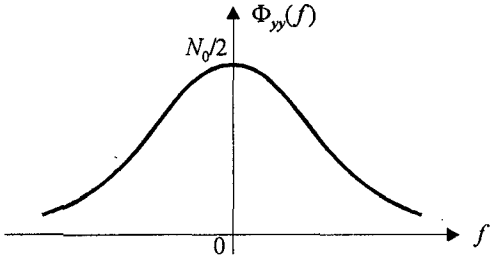


Рис. 2.2.2. Спектральная плотность мощности на выходе ФНЧ, когда на вход поступает белый шум

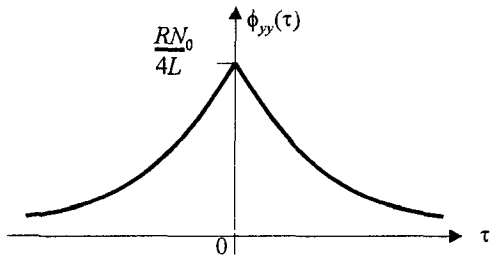


Рис. 2.2.3. Функция автокорреляции сигнала на выходе ФНЧ, когда на вход поступает белый шум

Заметим, что интеграл (2.2.33) – это интеграл свёртки. Следовательно, в частотной области из (2.2.33) следует соотношение

$$\Phi_{yx}(f) = \Phi_{xx}(f)H(f). \qquad (2.2.34)$$

Видно, что если на входе системы действует белый шум, то функция взаимной корреляции входа и выхода системы с точностью до масштабирующего коэффициента равна импульсному отклику $h(t)$.

2.2.4. Теорема отсчётов для частотно-ограниченных случайных процессов

Напомним, что детерминированный сигнал $s(t)$ с преобразованием Фурье $S(f)$ называется частотно-ограниченным, если $S(f)=0$ для $|f|>W$, где W – наивысшая частота, содержащаяся в $s(t)$. Такой сигнал однозначно определяется отсчётами $s(t)$, взятыми со скоростью $f_s \geq 2W$ отсч./с. Минимальная скорость $f_N = 2W$ отсч./с называется *скоростью Найквиста*. Представление сигнала через отсчёты, взятые со скоростью ниже скорости Найквиста, ведёт к ошибкам.

Частотно-ограниченный сигнал, представленный отсчётами, взятыми со скоростью Найквиста, может быть восстановлен по своим отсчётам интерполяционной формулой

$$s(t) = \sum_{n=-\infty}^{\infty} s(n/2W)\frac{\sin[2\pi W(t-n/2W)]}{2\pi W(t-n/2W)}, \qquad (2.2.35)$$

где $\{s(n/2W)\}$ – отсчёты $s(t)$, взятые в моменты времени $t=n/2W$, $n=0, \pm1, \pm2, \ldots$. Эквивалентным образом $s(t)$ можно реконструировать путём пропускания отсчётов дискретизированного сигнала через идеальный ФНЧ с импульсной характеристикой $h(t)=\sin(2\pi Wt)/2\pi Wt$. Рисунок 2.2.4 иллюстрирует процесс восстановления сигнала, основанный на идеальной интерполяции.

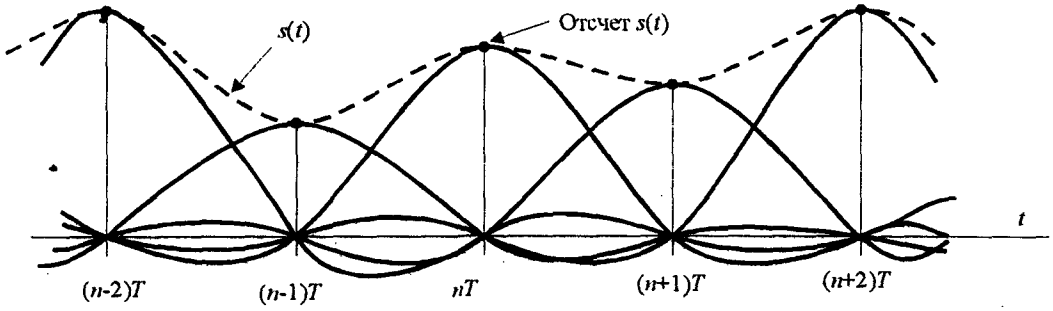


Рис. 2.2.4. Восстановление сигнала, основанное на идеальной интерполяции

Стационарный случайный процесс $X(t)$ называется частотно-ограниченным, если его спектральная плотность мощности $\Phi(f)=0$ для $|f|>W$. Поскольку $\Phi(f)$ является преобразованием Фурье автокорреляционной функции $\phi(\tau)$, то следует представление для $\phi(\tau)$:

$$\phi(\tau)=\sum_{n=-\infty}^{\infty}\phi\left(\frac{n}{2W}\right)\frac{\sin\left[2\pi W\left(\tau-\frac{n}{2W}\right)\right]}{2\pi W\left(\tau-\frac{n}{2W}\right)}, \qquad (2.2.36)$$

где $\{\phi(n/2W)\}$ – отсчёты $\phi(\tau)$, взятые при $\tau=n/2W$, $n=0, \pm1, \pm2,\ldots$.

Теперь, если $X(t)$ – частотно-ограниченный стационарный случайный процесс, то $X(t)$ можно представить в виде

$$X(t)=\sum_{n=-\infty}^{\infty}X\left(\frac{n}{2W}\right)\frac{\sin\left[2\pi W\left(t-\frac{n}{2W}\right)\right]}{2\pi W\left(t-\frac{n}{2W}\right)}, \qquad (2.2.37)$$

где $\{X(n/2W)\}$ – отсчёты $X(t)$, взятые при $t=n/2W$, $n=0, \pm1, \pm2, \ldots$.

Это – представление стационарного случайного процесса через его отсчёты.

Отсчёты являются случайными величинами, которые описываются статистически соответствующей СФПВ. Представление (2.2.37) легко устанавливается доказательством того (задача 2.17), что

$$E\left\{\left|X(t)-\sum_{n=-\infty}^{\infty}X\left(\frac{n}{2W}\right)\frac{\sin\left[2\pi W\left(t-\frac{n}{2W}\right)\right]}{2\pi W\left(t-\frac{n}{2W}\right)}\right|^2\right\}=0. \qquad (2.2.38)$$

Следовательно, равенство между представлением случайного процесса $X(t)$ через его отсчёты и самого процесса понимается в том смысле, что средний квадрат ошибки равен нулю.

2.2.5. Случайные сигналы и системы с дискретным временем

Описание случайных сигналов с непрерывным временем, данное выше, можно легко распространить на случайные сигналы с дискретным временем. Такие сигналы обычно получаются путем равномерной дискретизации во времени случайного процесса с непрерывным временем.

Случайный процесс с дискретным временем $X(n)$ состоит из множества реализаций последовательностей $\{x(n)\}$. Статистические свойства $X(n)=X_n$ сходны с теми, которые определены для $X(t)$, с тем ограничением, что n теперь целая переменная (дискретное время). Следовательно, m-й момент для $X(n)$ определяется как

$$E\left[X_n^m\right] = \int_{-\infty}^{\infty} x_n^m p\left(x_n\right) dx_n \qquad (2.2.39)$$

и автокорреляционная последовательность

$$\phi(n,k) = \tfrac{1}{2} E\left(X_n X_k^*\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_n x_k^* p\left(x_n, x_k\right) dx_n dx_k \,. \qquad (2.2.40)$$

Подобным образом определяется и автоковариационная последовательность

$$\mu(n,k) = \phi(n,k) - E\left(X_n\right) E\left(X_k^*\right). \qquad (2.2.41)$$

Для стационарного процесса имеем $\phi(n,k) \equiv \phi(n-k)$, $\mu(n,k) = \mu(n-k)$ и

$$\mu(n-k) = \phi(n-k) - \left|m_x\right|^2, \qquad (2.2.42)$$

где $m_x = E\left(X_n\right)$ – среднее значение.

Как и в случае случайного процесса с непрерывным временем стационарный процесс с дискретным временем имеет неограниченную энергию, но ограниченную среднюю мощность, которая определяется как

$$E\left(\left|X_n\right|^2\right) = \phi(0). \qquad (2.2.43)$$

Спектральная плотность мощности для случайного стационарного процесса с дискретным временем получается преобразованием Фурье от $\phi(n)$. Поскольку $\phi(n)$ – последовательность дискретного времени, преобразование Фурье определено в виде

$$\Phi\left(f\right) = \sum_{n=-\infty}^{\infty} \phi(n)\, \mathrm{e}^{-j2\pi fn}, \qquad (2.2.44)$$

а обратное преобразование – в виде

$$\phi(n) = \int_{-1/2}^{1/2} \Phi\left(f\right) \mathrm{e}^{j2\pi fn}\, df \,. \qquad (2.2.45)$$

Обратим внимание на то, что спектральная плотность мощности $\Phi(f)$ является периодической с периодом $f_p = 1$. Другими словами, $\Phi\left(f+k\right) = \Phi\left(f\right)$ для $k = \pm 1, \pm 2, \ldots$ Это характерно для преобразования Фурье дискретной во времени последовательности такой как $\phi(n)$.

В заключение рассмотрим отклик линейной стационарной системы с дискретным временем на стационарные случайные входные воздействия. Система характеризуется во временной области своей импульсной характеристикой $h(n)$ (откликом на единичный отсчет времени), а в частотной области – частотной характеристикой $H(f)$, где

$$H\left(f\right) = \sum_{n=-\infty}^{\infty} h(n)\, \mathrm{e}^{-j2\pi fn} \,. \qquad (2.2.46)$$

Отклик системы на стационарный случайный входной сигнал $X(n)$ определяется дискретной сверткой

$$y(n) = \sum_{n=-\infty}^{\infty} h(k)\, x(n-k) \,. \qquad (2.2.47)$$

Среднее значение выхода системы

$$m_y = E\left[y(n)\right] = \sum_{k=-\infty}^{\infty} h(k)\, E\left[x(n-k)\right];$$
$$m_y = m_x \sum_{k=-\infty}^{\infty} h(k) = m_x H(0). \qquad (2.2.48)$$

где $H(0)$ – передаточная функция системы на нулевой частоте.

Автокорреляционная последовательность для выходного процесса

$$\phi_{yy}(k) = \tfrac{1}{2}E\left[Y^*(n)\,Y(n+k)\right] = .$$

$$= \tfrac{1}{2}\sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} h^*(i)\,h(j)E\left[X^*(n-i)\,X(n+k-j)\right] = \qquad (2.2.49)$$

$$= \sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} h^*(i)\,h(j)\,\phi_{xx}(k-j+i).$$

Это общая форма для автокорреляционной последовательности выхода системы, выраженная через автокорреляционную функцию входа системы и импульсную характеристику системы. Производя преобразования Фурье над $\phi_{yy}(k)$ и учитывая (2.2.49), получаем соответствующее соотношение в частотной области

$$\Phi_{yy}(f) = \Phi_{xx}(f)\left|H(f)\right|^2, \qquad (2.2.50)$$

которое идентично (2.2.27), за исключением того, что в (2.2.50) спектральные плотности мощности $\Phi_{yy}(f)$ и $\Phi_{xx}(f)$ и частотная характеристика $H(f)$ являются периодическими функциями частоты с периодом $f_p = 1$.

2.2.6. Процессы с циклической стационарностью

При обработке сигналов, которые несут цифровую информацию, мы встречаемся со случайными процессами, которые имеют периодически повторяющиеся средние значения. Для конкретности рассмотрим случайный процесс вида

$$X(t) = \sum_{n=-\infty}^{\infty} a_n g(t-nT), \qquad (2.2.51)$$

где $\{a_n\}$ – последовательность (с дискретным временем) случайных величин со средним $m_a = E(a_n)$ для всех n и автокорреляционной последовательностью $\phi_{aa}(k) = \frac{1}{2}E\left(a_n^* a_{n+k}\right)$.

Сигнал $g(t)$ детерминирован. Случайный процесс $X(t)$ представляет сигнал для некоторых различных видов линейной модуляции, которые рассматриваются в гл. 4. Последовательность $\{a_n\}$ представляет цифровую информацию источника (символы), которая передается по каналу связи, а $1/T$ определяет скорость передачи информационных символов.

Определим среднее и автокорреляционную функцию $X(t)$. Сначала находим среднее значение

$$E\left[X(t)\right] = \sum_{n=-\infty}^{\infty} E(a_n)g(t-nT) = m_a\sum_{n=-\infty}^{\infty} g(t-nT). \qquad (2.2.52)$$

Видим, что среднее меняется во времени, но меняется периодически с периодом T. Автокорреляционная функция

$$\phi_{xx}(t+\tau,t) = \tfrac{1}{2}E\left[X(t+\tau)X^*(t)\right] =$$

$$= \tfrac{1}{2}\sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty} E\left(a_n^* a_m\right)g^*(t-nT)g(t+\tau-mT) = \qquad (2.2.53)$$

$$= \sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}\phi_{aa}(m-n)g^*(t-nT)g(t+\tau-mT).$$

Снова видим, что

$$\phi_{xx}(t+\tau+kT, t+kT) = \phi_{xx}(t+\tau,t), \qquad (2.2.54)$$

для $k = \pm 1, \pm 2, \ldots$. Следовательно, автокорреляционная функция $X(t)$ также является периодической с периодом T.

Такой случайный процесс назван *циклостационарным* или *периодически стационарным*. Поскольку автокорреляционная функция процесса зависит от обеих

переменных t и τ, его частотное представление требует двухмерного преобразования Фурье.

Поскольку крайне желательно характеризовать такие сигналы их спектральной плотностью мощности, альтернативный подход заключается в вычислении *средней во времени* за один период автокорреляционной функции, определяемой как

$$\bar{\phi}_{xx}(\tau) = \frac{1}{T}\int_{-T/2}^{T/2} \phi_{xx}(t+\tau, t)\,dt. \qquad (2.2.55)$$

Используя усредненную функцию автокорреляции, мы исключаем зависимость от времени. Теперь преобразование Фурье от $\bar{\phi}_{xx}(\tau)$ дает усредненную спектральную плотность мощности для циклически стационарного случайного процесса. Такой подход позволяет нам упростить характеристику циклически стационарного процесса в частотной области. Таким образом, спектральная плотность мощности определяется как

$$\Phi_{xx}(f) = \int_{-\infty}^{\infty} \bar{\phi}_{xx}(\tau) e^{-j2\pi f\tau} d\tau. \qquad (2.2.56)$$

2.3. БИБЛИОГРАФИЧЕСКИЕ ЗАМЕЧАНИЯ И ССЫЛКИ

В этой главе мы дали обзор базовых понятий и определений из теории вероятности и теории случайных процессов. Как отмечено в начале главы, эта теория является важным математическим инструментом при статистическом моделировании источников информации, каналов связи и при расчете цифровых систем связи. В частности, важной при оценке характеристик систем связи является граница Чернова. Эта граница часто используется для оценки вероятности ошибки цифровых систем связи при использовании кодирования при передаче информации. Наш обзор также осветил ряд распределений вероятностей и их свойств, которые часто используются при расчете систем цифровой связи.

Давенпорт и Рут (1958), Давенпорт (1970), Папулис (1984), Пеблес (1987), Хелстром (1991) и Леон-Гарсия (1994) дали в своих книгах инженерно-ориентированное рассмотрение теории вероятности и теории случайных процессов. Более глубокое математическое рассмотрение теории вероятности можно найти в книгах Лоэва (1955). Наконец, упомянем книгу Миллера (1962), который рассмотрел многомерные гауссовские распределения.[1]

ЗАДАЧИ

2.1. Один эксперимент имеет четыре взаимосвязанных результата A_i, i=1, 2, 3, 4, а второй эксперимент имеет три взаимосвязанных результата B_j, j=1, 2, 3. Совместные вероятности $P(A_i, B_j)$

$$\begin{array}{lll} P(A_1,B_1)=0{,}10 & P(A_1,B_2)=0{,}08 & P(A_1,B_3)=0{,}13 \\ P(A_2,B_1)=0{,}05 & P(A_2,B_2)=0{,}03 & P(A_2,B_3)=0{,}09 \\ P(A_3,B_1)=0{,}05 & P(A_3,B_2)=0{,}12 & P(A_3,B_3)=0{,}14 \\ P(A_4,B_1)=0{,}11 & P(A_4,B_2)=0{,}04 & P(A_4,B_3)=0{,}06. \end{array}$$

Определите вероятиость $P(A_i)$, $i = 1, 2, 3, 4$, и $P(B_j)$, $j = 1, 2, 3$.

2.2. Случайиые величиины X_i, $i = 1,2,\ldots,n$, имеют СФПВ $p(x_1, x_2, \ldots x_n)$. Докажите, что

[1] Первые монографии по теории вероятностей и теории случайиых процессов, ориеитированные на решение задач радиотехники, связи и управления, появились в России в 1957 г. и прииадлежат Б.Р. Левину [40], В.С. Пугачёву [41]. В 1966 г. появилась очередная книга Б.Р. Левина по этой тематике [27], а также пользующаяся большой популярностью книга В.И. Тихонова [42, см. также 43]

$$p(x_1,x_2,..,x_n) = p(x_n|x_{n-1},x_{n-2},..,x_1)p(x_{n-1}|x_{n-2},x_{n-3},..,x_1)...p(x_3|x_2,x_1)p(x_2|x_1)p(x_1).$$

2.3. Дана $p(x)$ – ФПВ случайной величины X. Случайная величина Y определяется как $Y = aX + b$, где $a<0$.

Определите ФПВ Y через ФПВ X.

2.4. Предположим, что X является гауссовской случайной величиной с нулевым средним и единичной дисперсией.

Пусть $Y = aX^3 + b, \quad a > 0$. Определите и постройте график ФПВ для Y.

2.5. а) Пусть X_r и X_i – статистически независимые гауссовские случайные величины с нулевыми средними и одинаковыми дисперсиями. Покажите, что преобразование (поворот) вида $Y_r + jY_i = (X_r + jX_i)e^{j\phi}$ порождает другую пару (Y_r, Y_i) гауссовских случайных величин, которые имеют ту же СФПВ, что и пара (X_r, X_i).

b) Заметим, что в п. а)

$\begin{bmatrix} Y_r \\ Y_i \end{bmatrix} = \mathbf{A}\begin{bmatrix} X_r \\ X_i \end{bmatrix}$, где $\mathbf{A}$ – матрица размерности 2×2. В порядке обобщения двухмериого преобразования гауссовских случайных величин из а) определите, какие свойства должны быть у матрицы (преобразования) $\mathbf{A}$ для того, чтобы ФПВ $\mathbf{X}$ и $\mathbf{Y}$, где $\mathbf{Y} = \mathbf{AX}$, $\mathbf{X} = (X_1 X_2 ... X_n)$ и $\mathbf{Y} = (Y_1 Y_2 ... Y_n)$, были бы одинаковыми.

2.6. Случайная величина Y определяется как

$Y = \sum_{i=1}^{n} X_i$, где X_i, i=1, 2, ... n – статистически независимые случайные величины, причем

$$X_i = \begin{cases} 1 & \text{с вероятностью } p, \\ 0 & \text{с вероятностью } 1-p. \end{cases}$$

а) Определите характеристическую функцию Y.

b) При помощи характеристической функции определите момент $E(Y)$ и $E(Y^2)$.

2.7. Четыре случайные величины X_1, X_2, X_3, X_4 являются совместно гауссовскими с нулевыми средними, с ковариацией $\mu_{ij} = E(X_i X_j)$ и характеристической функцией $\psi(j\upsilon_1, j\upsilon_2, j\upsilon_3, j\upsilon_4)$. Покажите, что $E(X_1X_2X_3X_4) = \mu_{12}\mu_{34} + \mu_{13}\mu_{24} + \mu_{14}\mu_{23}$.

2.8. При помощи характеристической функции для центрального и нецентрального хи-квадрат-распределения случайных величин, определяемых соответственно по формулам (2.1.109) и (2.1.117), определите соответствующие первые и вторые моменты (формулы (2.1.112) и (2.1.125)).

2.9. Случайная величина X распределена по Коши с ФПВ

$$p(x) = \frac{a/\pi}{x^2 + a^2}, -\infty < x < \infty.$$

а) Определите среднее и дисперсию X.

b) Определите характеристическую функцию X.

2.10. Случайная величина Y определена как

$Y = \frac{1}{n}\sum_{i=1}^{n} X_i$, где X_i, i=1, 2, ..., n – статистически независимые и одинаково распределенные случайные величины, каждая из которых имеет распределение Коши из задачи 2.9.

а) Определите характеристическую функцию Y.

b) Определите ФПВ для Y.

c) Рассмотрите ФПВ Y в пределе при $n \to \infty$. Работает ли центральная предельная теорема? Обоснуйте ваш ответ.

2.11. Предположим, что случайные процессы $X(t)$ и $Y(t)$ являются совместно и по отдельности стационарными.

а) Определите функцию автокорреляции $Z(t)=X(t)+Y(t)$.

b) Определите автокорреляционную функцию $Z(t)$ для случая, когда $X(t)$ и $Y(t)$ не коррелированы.

c) Определите автокорреляционную функцию для случая, когда $X(t)$ и $Y(t)$ являются некоррелированными и имеют нулевые средние.

2.12. Функция автокорреляции случайного процесса $X(t)$ определяется так: $\phi_{xx}(\tau) = \frac{1}{2}N_0\delta(\tau)$.

Такой процесс называется *белым шумом*. Пусть $X(t)$ является входом для идеального полосового фильтра с частотной характеристикой, показанной на рис. 2.12. Определите суммарную мощность шума на выходе фильтра.

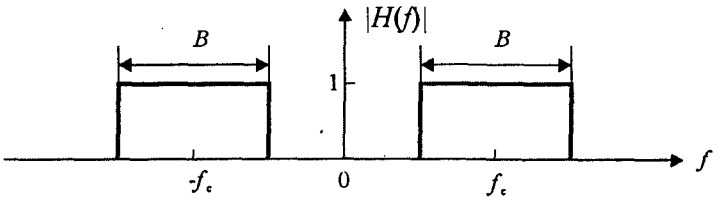


Рис. P2.12

2.13. Даиа ковариационная матрица $\begin{bmatrix} \mu_{11} & 0 & \mu_{13} \\ 0 & \mu_{22} & 0 \\ \mu_{31} & 0 & \mu_{33} \end{bmatrix}$ случайных величии X_1, X_2 и X_3.

Осуществлеио линейное преобразование **Y=AX**, где

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix}.$$

Определите ковариационную матрицу для **Y**.

2.14. Пусть $X(t)$ является вещественным стационарным гауссовским процессом с нулевым средним. Пусть новый процесс определен как $Y(t) = X^2(t)$. Определите автокорреляционную функцию $Y(t)$ через автокорреляционную функцию $X(t)$. Подсказка: используйте результат для гауссовских случайиых величин из задачи 2.7.

2.15. Для ФПВ Накагами (формула 2.1.147) определите нормированную случайную величину $X = R/N\Omega$. Найдите ФПВ для X.

2.16. Входиым воздействием цепи, показаниой на рис. 2.16, является случайный процесс $X(t)$ с $E[X(t)] = 0$. н $\phi_{xx}(\tau) = \sigma^2\delta(\tau)$, т.е. $X(t)$ является белым шумом.

a) Определите спектральную плотность мощности выхода $\Phi_{yy}(f)$.

b) Определите $\phi_{yy}(\tau)$ и $E[Y^2(t)]$.

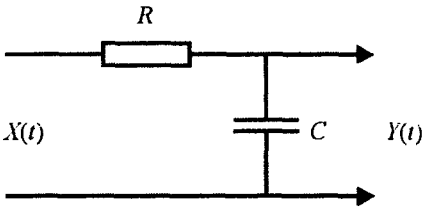


Рис. P2.16

2.17. Докажите справедливость (2.2.38).

2.18. Докажите, используя границу Чернова, что $Q(x) \le e^{-x^2/2}$, где $Q(x)$ определяется (2.1.97).

2.19. Определите среднее, автокорреляционную последовательность и спектральную плотность мощности для сигнала на выходе системы (цифрового фильтра) с импульсной характеристикой

$$h(n) = \begin{cases} 1 & (n = 0) \\ -2 & (n = 1) \\ 1 & (n = 2) \\ 0 & (\text{для других } n), \end{cases}$$

если входной случайный процесс $X(n)$ является белым шумом с дисперсией σ_x^2.

2.20. Автокорреляционная последовательность дискретного во времеии случайиого процесса равн $\phi(k) = \left(\frac{1}{2}\right)^{|k|}$.

Определите его спектральную плотность мощности.

2.21. Случайный процесс с дискретным временем $X(n) \equiv X(nT)$ получен периодически стробированием стационарного процесса $X(t)$ с непрерывиым временем и нулевым средним, где T – перио стробирования, т.е. $f_s = 1/T$ является скоростью выборки отсчетов.

a) Определите соотношения между функцией автокорреляции сигнала $X(t)$ и автокорреляционно последовательностью его отсчётов $X(n)$.

b) Выразите спектральную плотность мощности процесса $X(n)$ через спектральиую плотность мощнос

процесса $X(t)$.

c) Определите условия, при которых спектральная плотность мощности $X(n)$ равна спектральной плотности мощности $X(t)$.

2.22. Рассмотрим частотно-ограниченный стационарный случайный процесс $X(t)$ с нулевым средним и спектральной плотностью мощности (квазибелый случайный процесс)

$$\Phi(f)=\begin{cases}1, & |f|\le W,\\ 0, & |f|>W.\end{cases}$$

Для образования процесса с дискретным времеием $X(n)=X(nT)$ берутся отсчёты $X(t)$ со скоростью $f_s=1/T$.

a) Определите выражение для автокорреляциоиной последовательности $X(n)$.

b) Определите минимальное значение T, необходимое для получения «белой» последовательности (спектрально ровиой).

c) Повторите b) для случая, когда спектральная плотность для $X(t)$ определена как

$$\Phi(f)=\begin{cases}1-|f|/W, & |f|\le W,\\ 0, & |f|>W.\end{cases}$$

2.23. Покажите, что фуикции

$$f_k(t)=\frac{\sin\left[2\pi W\left(t-\frac{k}{2W}\right)\right]}{2\pi W\left(t-\frac{k}{2W}\right)},\quad k=0,\ \pm1,\ \pm2,...$$

являются ортогональными иа интервале $[-\infty,\infty]$, т. е.

$$\int_{-\infty}^{\infty}f_k(t)f_j(t)dt=\begin{cases}1/2W & (k=j),\\ 0 & (k\ne j).\end{cases}$$

Следовательно, формулу из теоремы отсчётов можно рассматривать как представлеиие частотно-ограиичеиной сигнала $s(t)$ обобщённым рядом Фурье, где веса разложеиия – это отсчёты сигнала $s(t)$, а $\{f_k(t)\}$ – ансамбль ортогональных функций, используемых в ортогональиом разложеиии.

2.24. Эквивалентная шумовая полоса частот системы определена как $B_{\text{экв}}=\frac{1}{G}\int_0^{\infty}|H(f)|^2df$, где $G=\max|H(f)|^2$. Используя это определение, найдите эквнвалеитную шумовую полосу идеального полосового фильтра, показанного на рисунке Р2.12, и низкочастотного фильтра, показанного на рисунке Р2.16.

3 КОДИРОВАНИЕ ИСТОЧНИКА

Системы связи предназначены для передачи информации, создаваемой источником, некоторого места назначения. Источники информации могут принимать множест различных форм. Например, в радиовещании источник выдает звуковой сигнал (речь и музыку). В телевизионном вещании выходом источника является, кроме звука, подвижн изображение. Выходы этих источников являются аналоговыми сигналами, и поэтому о называются *аналоговыми источниками*. В противоположность этому компьютеры устройства хранения информации, такие как магнитные или оптические диски, имею дискретный выход (обычно двоичные или ASCII[1] символы), и поэтому их называ *дискретными источниками*. В то время как источники являются аналоговыми и дискретными, цифровая система связи предназначается для передачи информации цифровой форме. Следовательно, выход источника должен быть преобразован в форма который может быть передан как цифровой. Это преобразование выхода источника цифровой формат обычно осуществляется кодером источника, выход которого может бы представлен последовательностью двоичных цифр.

В этой главе мы рассмотрим кодирование источника, основанное на математическ моделях источников информации и количественном измерении информации, выдаваем источником. Сначала мы рассмотрим кодирование дискретных источников и зат обсудим кодирования аналоговых источников. Мы начнем с рассмотрения математическ моделей для источников информации.

3.1. МАТЕМАТИЧЕСКИЕ МОДЕЛИ ДЛЯ ИСТОЧНИКОВ ИНФОРМАЦИИ

Произвольный источник информации создает выход, который является случайным, т выход источника характеризуется статистически. Действительно, если выход источни известен точно, то нет нужды его передавать. В этом разделе мы рассмотрим дискретные аналоговые источники информации и сформулируем математические модели для каждо типа источника.

Простейший тип дискретного источника – это такой, который выда последовательность букв (символов), выбираемых из определенного алфавита. Наприм *двоичный источник* выдает двоичную последовательность вида 100101110..., прич алфавит состоит из двух символов {0, 1}. В более общем случае источник дискретн информации с алфавитом из L символов, скажем $\{x_1, x_2, ..., x_L\}$, выдает последовательно букв, выбираемых из этого алфавита.

Чтобы конструировать математическую модель для дискретного источни предположим, что каждый символ алфавита $\{x_1, x_2, ..., x_L\}$ имеет заданную вероятно выбора p_k, т.е.

$$p_k=P(X=x_k),\ 1 \le k \le L,$$

[1] ASCII – американский стандартный код для информационного обмена (прп)

где $\sum_{k=1}^{L} p_k = 1$.

Мы рассмотрим две математические модели для дискретных источников. В первой мы предположим, что символы выходной последовательности источника статистически независимы, т.е. выбираемый текущий символ статистически независим от всех предыдущих и последующих. Источник, выход которого удовлетворяет условиям статистической независимости символов в выбранной последовательности, называется источником *без памяти*. Такой источник называется *дискретным источником без памяти* (ДИБП).

Если отдельные выходные символы дискретного источника статистически взаимозависимы, как, например, в английском тексте, мы можем сконструировать математическую модель, основанную на статической стационарности. По определению дискретный источник называется *стационарным*, если совместные вероятности двух последовательностей длины n, допустим a_1, a_2, ..., a_n и a_{1+m}, a_{2+m}, ..., a_{n+m} одинаковые для всех $n \geq 1$ и при всех сдвигах m. Другими словами, совместные вероятности для последовательностей источника произвольной длины инвариантны по отношению к произвольному сдвигу во времени.

Аналоговый источник выдает сигнал $x(t)$, который является реализацией случайного процесса $X(t)$. Предположим, что $X(t)$ – стационарный случайный процесс с автокорреляционной функцией $\phi_{xx}(\tau)$ и спектральной плотностью мощности $\Phi_{xx}(f)$. Если $X(t)$ – частотно-ограниченный случайный процесс, т.е. $\Phi_{xx}(f) = 0$ для $|f| \geq W$, можно использовать теорему отсчётов для представления $X(t)$ в виде

$$X(t) = \sum_{n=-\infty}^{\infty} X\left(\frac{n}{2W}\right) \frac{\sin[2\pi W(t - \frac{n}{2W})]}{2\pi W(t - \frac{n}{2W})}, \qquad (3.1.1)$$

где $X\{(n/2W)\}$ – отсчёты процесса $X(t)$, взятые со скоростью Найквиста $f_s = 2W$ 1/с. Используя теорему отсчётов, мы можем преобразовать аналоговый источник в эквивалентный источник с дискретным временем. После этого выход источника характеризуется совместной ФПВ $p(x_1, x_2, ..., x_m)$ для всех $m \geq 1$, где $X_n = X(n/2W)$, $1 \leq n \leq m$, – случайные величины, соответствующие отсчётам $X(t)$.

Заметим, что выходные отсчёты $X\{(n/2W)\}$ стационарного источника обычно непрерывны, и, следовательно, их нельзя представить в цифровой форме без потери точности представления. Например, мы можем квантовать каждый отсчёт рядом дискретных значений, но процесс квантования вносит потери в точность представления, и, следовательно, исходный сигнал не может быть восстановлен точно по квантованным отсчётам. Позже мы рассмотрим искажения, возникающие при квантовании уровней отсчётов аналогового источника.

3.2. ЛОГАРИФМИЧЕСКАЯ МЕРА ИНФОРМАЦИИ

Чтобы разработать подходящую меру для информации, рассмотрим две дискретные случайные величины X и Y с возможными значениями x_i, i=1, 2, ..., n, и y_j, j=1, 2, ..., m, соответственно. Допустим, мы наблюдаем некоторый выход Y=y_j и мы желаем количественно определить величину информации, которую обеспечивает выборка события Y=y_j относительно события X=x_i, i=1, 2, ..., n. Заметим, что если X и Y статистически не зависят друг от друга, выбор Y=y_j не даёт информации о выборе события X=x_i. С другой стороны, если X и Y полностью зависимы, так что выбор Y=y_j однозначно определяет выбор X=x_i, информационное содержание этого выбора точно такое же, как при выборе события

$X=x_i$. Подходящая мера информации, которая удовлетворяет указанным условиям, – это логарифм отношения условной вероятности

$$P(X = x_i \mid Y = y_j) = P(x_i \mid y_j)$$

к вероятности

$$P\left(X = x_i\right) \equiv P\left(x_i\right).$$

Это значит, что количество информации, полученное при появлении события $Y=y_j$ относительно события $X=x_i$ определяется как

$$I(x_i; y_j) = \log \frac{P(x_i \mid y_j)}{P(x_i)}. \qquad (3.2.1)$$

$I\left(x_i; y_j\right)$ названа *взаимной информацией* между x_i и y_j.

Единица измерения $I\left(x_i; y_j\right)$ определяется основанием логарифма, в качестве которого обычно выбирается или 2, или e. Когда основание логарифма равно 2, единицей измерения $I\left(x_i; y_j\right)$ является *бит*, а когда основание равно e, единицей измерения $I\left(x_i; y_j\right)$ является *нат* (натуральная единица). (Стандартная аббревиатура для $\log_e$ – это ln.) Так как

$$\ln a = \ln 2 \cdot \log_2 a = 0{,}69315 \cdot \log_2 a,$$

то количество информации, измеренное в натах, равно количеству информации, измеренной в битах, умноженному на $\ln 2$.

Когда случайные величины X и Y статистически независимы, то $P\left(x_i \middle| y_j\right) = P\left(x_i\right)$, и следовательно, $I\left(x_i; y_j\right) = 0$. С другой стороны, когда выбор события $Y=y_j$ полностью определён выбором события $X=x_i$, условная вероятность в числителе (3.21) равна единице и, следовательно,

$$I(x_i; y_j) = \log \frac{1}{P(x_i)} = -\log P(x_i) = I(x_i; x_i). \qquad (3.2.2)$$

Но (3.22) как раз определяет информацию о $X=x_i$. Исходя из этих соображений, её называют *собственной информацией* события $X=x_i$. Она обозначается так:

$$I(x_i) = \log \frac{1}{P(x_i)} = -\log P(x_i). \qquad (3.2.3)$$

Заметим, что событие, которое выбирается с высокой вероятностью, сообщает меньше информации, чем маловероятное событие. Действительно, если имеется единственное событие x с вероятностью $P(x)=1$, тогда $I(x)=0$.

Чтобы далее показать, что логарифмическая мера количества информации является единственно приемлемой для цифровой связи, рассмотрим несколько примеров.

Пример 3.2.1. Предположим, что имеется дискретный источник, который выдаёт двоичную цифру 0 или 1 с равной вероятностью каждые τ_s секунд. Количество информации при каждом появлении новой цифры

$$I(x_i) = -\log_2 P(x_i) = -\log_2 \tfrac{1}{2} = 1 \text{(бит)}, \quad x_i = 0, 1.$$

Теперь предположим, что последовательные цифры на выходе источника статистически независимы, т.е. источник не имеет памяти. Рассмотрим блок символов источника из k двоичных цифр, который существует на интервале $k\tau_s$. Имеется $M = 2^k$ таких возможных k-битовых блоков, каждый с равной вероятностью $1/M = 2^{-k}$. Собственная информация k-битового блока равна

$$I(x_i') = -\log_2 2^{-k} = k \text{ бит},$$

и она выдаётся на временном интервале $K\tau_s$. Таким образом, логарифмическая мера количества информации обладает желаемыми свойствами аддитивности, когда определённое число единичных выходов источника рассматривается как один блок.

Теперь вернёмся к определению взаимной информации, определяемой (3.2.1), и умножим числитель и знаменатель отношения вероятностей на $p(y_j)$:

$$\frac{P(x_i \mid y_j)}{P(x_i)} = \frac{P(x_i \mid y_j)P(y_j)}{P(x_i)P(y_j)} = \frac{P(x_i, y_j)}{P(x_i)P(y_j)} = \frac{P(y_j \mid x_i)}{P(y_j)}.$$

Отсюда делаем вывод

$$I(x_i; y_j) = I(y_j; x_i). \qquad (3.2.4)$$

Таким образом, информация, содержащаяся в выборе события $Y=y_j$ относительно события $X=x_i$, идентична информации, содержащейся в выборе события $X=x_i$ относительно события $Y=y_j$.

Пример 3.2.2. Предположим, что X и Y – двоичные $\{0,1\}$ случайные величины, представляющие вход и выход канала с двоичным входом и двоичным выходом. Входные символы равновероятны, а условные вероятности выходных символов при заданном входе определяются так:

$$P(Y=0 \mid X=0) = 1-p_0,$$
$$P(Y=1 \mid X=0) = p_0,$$
$$P(Y=1 \mid X=1) = 1-p_1,$$
$$P(Y=0 \mid X=1) = p_1.$$

Определим, сколько информации об $X = 0$ и $X = 1$ содержится в событии $Y = 0$. Из заданных вероятностей получим

$$P(Y=0) = P(Y=0 \mid X=0)P(X=0) + P(Y=0 \mid X=1)P(X=1) = \tfrac{1}{2}(1-p_0+p_1);$$
$$P(Y=1) = P(Y=1 \mid X=0)P(X=0) + P(Y=1 \mid X=1)P(X=1) = \tfrac{1}{2}(1-p_1+p_0).$$

Тогда взаимная информация о символе $X = 0$ при условии, что наблюдается $Y = 0$, равна

$$I(x_1; y_1) = I(0;0) = \log_2 \frac{P(Y=0 \mid X=0)}{P(Y=0)} = \log_2 \frac{2(1-p_0)}{1-p_0+p_1}$$

Аналогично взаимная информация о символе $X=1$ при условии, что наблюдается $Y = 0$, равна

$$I(x_2; y_1) = I(1;0) = \log_2 \frac{2p_1}{1-p_0+p_1}.$$

Рассмотрим несколько частных случаев. В первом, когда $p_0 = p_1 = 0$, канал называют каналом *без шумов* и $I(0;0) = \log_2 2 = 1$ бит.

Следовательно, когда выход точно определяет вход, нет потери информации. С другой стороны, если $p_0 = p_1 = 1/2$, канал становится *непригодным* [1], так как $I(0;0) = \log_2 1 = 0$.

Если $p_0 = p_1 = 1/4$, то

$$I(0;0) = \log_2 \tfrac{3}{2} = 0{,}587 \text{ бит};$$
$$I(0;1) = \log_2 \tfrac{1}{2} = -1 \text{ бит}.$$

[1] Этот случай называют «обрыв канала» (прп)

Помимо определения взаимной информации и собственной информации полезно определить *условную собственную информацию* как

$$I(x_i \mid y_j) = \log \frac{1}{P(x_i \mid y_j)} = -\log P(x_i \mid y_j). \qquad (3.2.5)$$

Тогда, комбинируя (3.2.1), (3.2.3) и (3.2.5), получаем соотношение

$$I(x_i; y_j) = I(x_i) - I(x_i \mid y_j). \qquad (3.2.6)$$

Мы интерпретируем $I(x_i \mid y_j)$ как собственную информацию о событии $X=x_i$ после наблюдения события $Y=y_j$. Из условия $I(x_i) \geq 0$ и $I(x_i \mid y_j) \geq 0$ следует, что $I(x_i, y_j) < 0$, когда $I(x_i \mid y_j) > I(x_i)$, и $I(x_i, y_j) > 0$, когда $I(x_i \mid y_j) < I(x_i)$. Следовательно, взаимная информация между парой событий может быть или положительной, или отрицательной или равной нулю.

3.2.1. Средняя взаимная информация и энтропия

Зная взаимную информацию, связанную с парой событий (x_i, y_j), которые являются возможной реализацией двух случайных величин X и Y, мы можем получить среднее значение взаимной информации простым взвешиванием $I(x_i, y_j)$ с вероятностью появления этой пары и суммированием по всем возможным событиям. Таким образом получим

$$I(X;Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) I(x_i, y_j) = \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) \log \frac{P(x_i, y_j)}{P(x_i) P(y_j)} = I(Y;X) \qquad (3.2.7)$$

как среднюю взаимную информацию между X и Y.

Видно, что $I(X,Y) = 0$, когда X и Y статистически независимы и $P(x_i, y_j) = P(x_i) P(y_j)$. Важным свойством средней взаимной информации является то, что $I(X,Y) \geq 0$ (см. задачу 3.4).

Аналогично определим среднюю собственную информацию, обозначенную $H(X)$:

$$H(X) = \sum_{i=1}^{n} P(x_i) I(x_i) = -\sum_{i=1}^{n} P(x_i) \log P(x_i). \qquad (3.2.8)$$

Если X представляет собой алфавит возможных символов источника, $H(X)$ представляет среднюю собственную информацию на символ источника, и её называют *энтропией*[1] *источника*. В частном случае, когда символы источника равновероятны, $P(x_i) = 1/n$ для всех i, и, следовательно,

$$H(X) = -\sum_{i=1}^{n} \frac{1}{n} \log \frac{1}{n} = \log n. \qquad (3.2.9)$$

В общем случае $H(X) \leq \log n$ (см. задачу 3.5) при любых заданных вероятностях символов источника. Другими словами, *энтропия источника максимальна, когда выходные символы равновероятны.*

Пример 3.2.3. Рассмотрим двоичный источник, который выдаёт последовательность независимых символов, причём выходной кодовый символ «0» с вероятностью q, а символ «1» с вероятностью $1-q$. Энтропия такого источника

[1] Термин «энтропия» взят из механики (термодинамики), где функция, похожая на (3.2.8), названа (термодинамической) энтропией.

$$H(X) = H(q) = -q\log q - (1-q)\log(1-q). \qquad (3.2.10)$$

Функцию $H(q)$ иллюстрирует рис. 3.2.1. Видно, что максимальное значение функции энтропии имеет место при $q = \frac{1}{2}$, причём $H\left(\frac{1}{2}\right) = 1$ бит.

Среднее значение условной собственной информации называется условной энтропией и определяется как

$$H(X \mid Y) = \sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i, y_j)\log\frac{1}{P(x_i \mid y_j)}. \qquad (3.2.11)$$

Мы интерпретируем $H(X|Y)$ как неопределённость X (дополнительную информацию, содержащуюся в X) после наблюдения Y

Комбинация (3.2.7), (3.2.8) и (3.2.11) даёт соотношение

$$I(X;Y) = H(X) - H(X \mid Y) = H(Y) - H(Y \mid X). \qquad (3.2.12)$$

Из условия $I(X,Y) \geq 0$ следует, что $H(X) \geq H(X|Y)$ и $H(Y) \geq H(Y|X)$, причём равенство имеет место тогда, и только тогда, когда X и Y статистически незави-симы. Если мы интерпретируем $H(X|Y)$ как среднее значение неопределённости (условной собственной информации) X после наблюдения Y и $H(X)$ как среднее значение априорной неопределённости (собственной информации), т.е. имевшейся до наблюдения, тогда $I(X,Y)$ определяет взаимную информацию (уменьшение среднего значения неопределённости, имеющейся относительно X после наблюдения Y). Так как $H(X) \geq H(X|Y)$, то ясно, что при условии наблюдения Y энтропия $H(X)$ не увеличится.

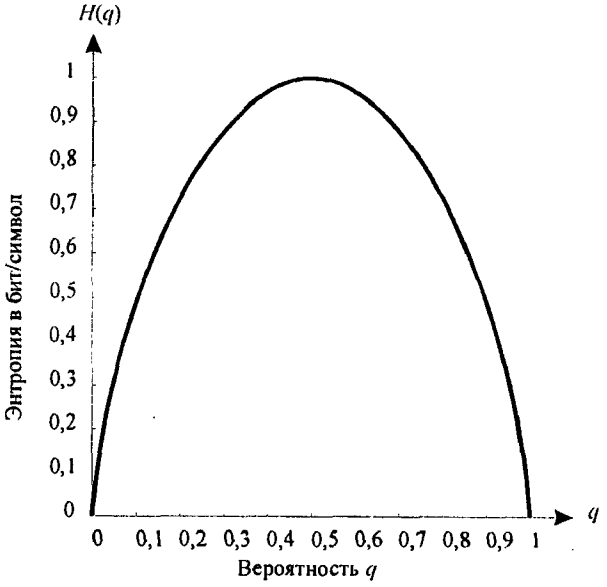


Рис. 3.2.1. Энтропия двоичного источника

Пример 3.2.4. Определим $H(X|Y)$ и $I(X,Y)$ для канала с двоичным входом и выходом, рассмотренного выше в примере 3.2.2, для случая, когда $p_0 = p_1 = p$. Пусть вероятность входных символов равна $P(X=0) = q$ и $P(X=1) = 1-q$.Тогда

$$H(X) \equiv H(q) = -q\lg q - (1-q)\lg(1-q),$$

где $H(q)$ – функция энтропии, а условная энтропия $H(X|Y)$ определяется (3.2.11). Зависимость $H(X|Y)$ в бит/символ как функция от q и параметра p показана на рис. 3.2.2. График средней взаимной информации $I(X,Y)$ в бит/символ дан на рис. 3.2.3.

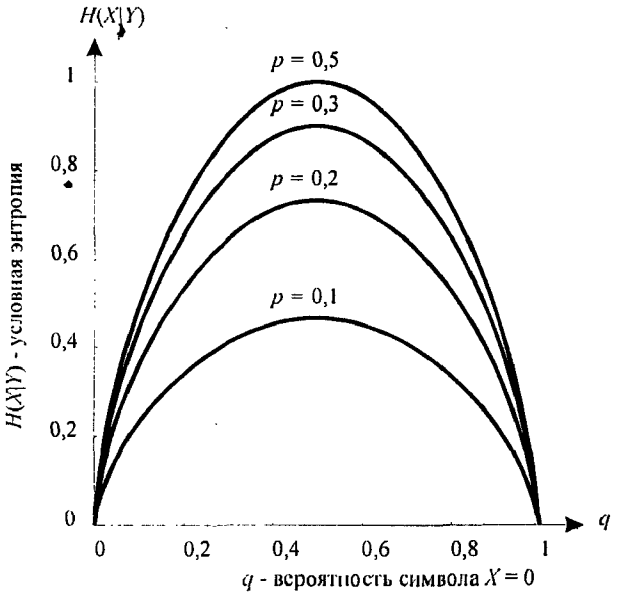


Рис. 3.2.2. Условная энтропия для двоичного симметричного канала

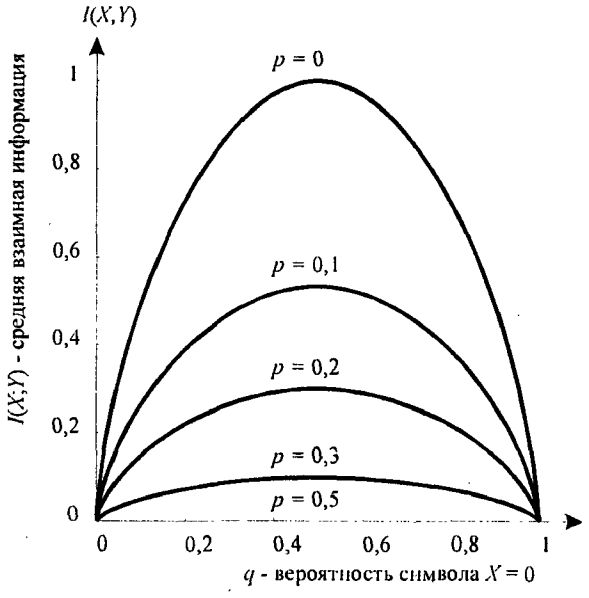


Рис.3.2.3. Средняя взаимная информация для двоичного симметричного канала

Когда условную энтропию $H(X|Y)$ рассматривают применительно к каналу с входом X и выходом Y, то $H(X|Y)$ называют ненадёжностью канала на символ и её интерпретируют как величину средней неопределённости, оставшейся в X после наблюдения Y [1].

Результаты, приведённые выше, легко обобщаются на случай произвольного числа случайных величин. В частности, предположим, что мы имеем блок из k случайных величин $X_1X_2X_3\ldots X_k$ с совместной вероятностью $P(X_1,X_2,\ldots,X_k)\equiv P(X_1=x_1,\ X_2=x_2,\ldots,X_k=x_k)$. Тогда энтропия определяется как

$$H(X_1X_2\ldots X_k)=-\sum_{j_1=1}^{n_1}\sum_{j_2=1}^{n_2}\cdots\sum_{j_k=1}^{n_k}P(x_{j_1}x_{j_2}\ldots x_{j_k})\log P(x_{j_1}x_{j_2}\ldots x_{j_k}). \qquad (3.2.13)$$

Поскольку совместную вероятность $P(X_1, X_2, \ldots, X_k)$ можно выразить в виде

$$P(x_1x_2\cdots x_k)=P(x_1)P(x_2\,|\,x_1)P(x_3\,|\,x_1x_2)\cdots P(x_k\,|\,x_1x_2\cdots x_{k-1}), \qquad (3.2.14)$$

то следует

$$H(X_1X_2X_3\ldots X_k)=H(X_1)+H(X_2\,|\,X_1)+H(X_3\,|\,X_1X_2)+$$
$$+\cdots+H(X_k\,|\,X_1\ldots X_{k-1})=\sum_{i=1}^{k}H(X_i\,|\,X_1X_2\ldots X_{i-1}). \qquad (3.2.15)$$

С учётом результата $H(X)\geq H(X|Y)$, где $X=X_m$ и $Y=X_1X_2\ldots X_{m-1}$, из (3.2.15) следует

$$H(X_1X_2\ldots X_k)\leq\sum_{m=1}^{k}H(X_m), \qquad (3.2.16)$$

причём равенство имеет место тогда, и только тогда, когда случайные величины $X_1, X_2, \ldots, X_k$ статистически независимы.

3.2.2. Измерение информации для непрерывных случайных величин

Определение взаимной информации, данное выше для дискретных случайных величин, можно непосредственно использовать для непрерывных случайных величин. В частности

[1] $H(Y|X)$ называют энтропией шума в канале (прп)

если X и Y – случайные величины с СПВ $p(x,y)$ и собственными ФПВ $p(x)$ и $p(y)$, то средняя взаимная информация между X и Y определяется как

$$I(X;Y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}p(x)p(y\mid x)\log\frac{p(y\mid x)p(x)}{p(x)p(y)}dxdy. \qquad (3.2.17)$$

Несмотря на то, что выражение для средней взаимной информации легко обобщается на непрерывные случайные величины, сделать это для собственной информации непрерывной случайной величины невозможно. Проблема в том, что непрерывные случайные величины требуют неограниченного числа двоичных цифр для их точного представления. Следовательно, энтропия непрерывной случайной величины также неограниченна. Всё же введём характеристику, которую назовём дифференциальной энтропией непрерывной случайной величины X:

$$h(X)=-\int_{-\infty}^{\infty}p(x)\log p(x)\,dx. \qquad (3.2.18)$$

Подчеркнём, что эта характеристика не имеет физического смысла собственной информации, хотя может показаться, что она является естественным обобщением определения энтропии для дискретной случайной величины (см. задачу 3.6). Определим среднюю условную дифференциальную энтропию X при заданном Y как

$$h(X\mid Y)=-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}p(x,y)\log p(x\mid y)dxdy. \qquad (3.2.19)$$

Тогда среднюю взаимную информацию можно выразить как

$$I(X;Y)=h(X)-h(X\mid Y)$$

или альтернативно как

$$I(X;Y)=h(Y)-h(Y\mid X).$$

В некоторых случаях, представляющих практический интерес, случайная величина X является дискретной, а Y – непрерывной. Для конкретности предположим, что X имеет возможные исходы x_i, i=1, 2, …, n, а Y определяется собственной ФПВ $p(y)$. Если X и Y статистически взаимосвязаны, мы можем выразить $p(y)$ так:

$$p(y)=\sum_{i=1}^{n}p(y\mid x_i)p(x_i).$$

Взаимная информация относительно события $X=x_i$ при наблюдении события $Y=y$ определяется как

$$I(x_i;y)=\log\frac{p(y\mid x_i)P(x_i)}{p(y)P(x_i)}=\log\frac{p(y\mid x_i)}{p(y)}. \qquad (3.2.20)$$

Тогда средняя взаимная информация между X и Y

$$I(X;Y)=\sum_{i=1}^{n}\int_{-\infty}^{\infty}p(y\mid x_i)P(x_i)\log\frac{p(y\mid x_i)}{p(y)}dy. \qquad (3.2.21)$$

Пример 3.2.5. Предположим, что X является дискретной случайной величиной с двумя равновероятными выходами $X_1 = A$ и $X_2 = -A$. Предположим, что условная ФПВ $p(y|x_i)$, i = 1, 2, является гауссовской со средним X_i и дисперсией σ^2, т.е.

$$p(y\mid A)=\frac{1}{\sqrt{2\pi}\sigma}e^{-(y-A)^2/2\sigma^2};$$
$$p(y\mid -A)=\frac{1}{\sqrt{2\pi}\sigma}e^{-(y+A)^2/2\sigma^2}. \qquad (3.2.22)$$

Средняя взаимная информация согласно (3.2.21) равна

$$I(X;Y)=\tfrac{1}{2}\int_{-\infty}^{\infty}\left[p(y\mid A)\log\frac{p(y\mid A)}{p(y)}+p(y\mid -A)\log\frac{p(y\mid -A)}{p(y)}\right]dy, \qquad (3.2.23)$$

$$p(y) = \tfrac{1}{2}\left[p(y \mid A) + p(y \mid -A)\right]. \qquad (3.2.24)$$

В гл. 7 мы покажем, что средняя взаимная информация $I(X;Y)$, определяемая (3.2.23), представляет пропускную способность канала с двоичным входом и аддитивным гауссовским шумом.

3.3. КОДИРОВАНИЕ ДЛЯ ДИСКРЕТНЫХ ИСТОЧНИКОВ

В разд. 3.2 мы ввели меру для информационного содержания дискретной случайной величины X. Когда X является выходом дискретного источника, энтропия $H(X)$ источника определяет среднее количество информации на символ, выдаваемой источником. В этом разделе мы рассмотрим процесс кодирования выхода источника, т.е. процесс представления выхода источника последовательностью двоичных цифр. Эффективность способа кодирования источника можно измерить путём сравнения среднего количества двоичных символов кодера на один символ источника и энтропии источника $H(X)$.

На первый взгляд может показаться, что кодирование дискретного источника с конечным объёмом алфавита является простой проблемой. Однако это верно, только если источник без памяти, т.е. когда последовательные символы источника статистически независимы и каждый символ кодируется отдельно. Дискретный источник без памяти (ДИБП) является простейшей моделью, которую можно предложить для физического источника. Эта идеализированная математическая модель подходит для немногих физических источников. Например, можно легко убедиться в том, что последовательно выдаваемые буквы устройством, печатающим осмысленный текст, статистически взаимосвязаны. С другой стороны, если печатается компьютерная программа на языке Фортран, то можно ожидать, что зависимость в последовательности выходных символов проявится значительно меньше. Во всяком случае, мы покажем, что всегда более эффективно кодировать блок символов источника вместо того, чтобы кодировать каждый символ отдельно. Если размер блока достаточно большой, то среднее количество символов кодера на один выходной символ источника можно сделать сколь угодно близким к энтропии источника.

3.3.1. Кодирование для дискретных источников без памяти

Предположим, что ДИБП выдает буквы или символы каждые τ_s секунд. Каждый символ выбирается из конечного алфавита x_i, $i=1, 2, \ldots, L$, с вероятностью $P(x_i)$, $i=1, 2, \ldots, L$. Энтропия ДИБП в битах на символ

$$H(X) = -\sum_{i=1}^{L} P(x_i)\log_2 P(x_i) \le \log_2 L, \qquad (3.3.1)$$

причем равенство имеет место, если все символы равновероятны. $H(X)$ определяет среднее число бит на символ источника, а производительность источника в битах определяется как $H(X)/\tau_s$.

Кодовые слова фиксированной длины. Сначала рассмотрим схему блокового кодирования, которая сопоставляет уникальный ряд из R двоичных символов с каждым символом источника. Поскольку имеется L возможных символов источника, то число двоичных символов кодера на один символ источника при уникальном кодировании

$$R = \log_2 L, \qquad (3.3.2)$$

когда L равно целой степени основания 2, и

$$R = \lfloor \log_2 L \rfloor + 1, \qquad (3.3.3)$$

когда L не равно целой степени основания 2.

Здесь $\lfloor x \rfloor$ означает наибольшее целое, меньшее, чем x.

R будем называть скоростью кодирования [1]. Она определяет число символов кодера на один символ источника. Поскольку $H(X) \le \log_2 L$, то $R \ge H(X)$.

Эффективность кодирования для ДИБП определяется отношением $H(X)/R$. Видим, что если L равно степени числа 2 и символы источника равновероятны, то $R=H(X)$. Следовательно, код фиксированной длины с R двоичными символами на символ источника в данном случае обеспечивает стопроцентную эффективность. Однако, если L не равно степени 2, но символы источника всё ещё равновероятны, R отличается от $H(X)$ самое большее на один бит на символ. Если $\log L >> 1$, эффективность такой схемы кодирования высока. С другой стороны, если L мало, эффективность кода с фиксированной длиной можно увеличить путем кодирования последовательности из J символов источника за время $J\tau_s$. Чтобы выполнить такое кодирование, мы должны выбрать L^J уникальных кодовых слов. Используя кодовую последовательность из N двоичных символов, мы можем образовать 2^N возможных кодовых слов. Число N должно быть выбрано так, чтобы

$$N \ge J \log_2 L .$$

Следовательно, требуется минимальное целое значение для N, равное

$$N = \lfloor J \log_2 L \rfloor + 1. \qquad (3.3.4)$$

Теперь среднее число символов кода на символ источника $R=N/J$, и, таким образом, неэффективность кодирования сокращается примерно в J раз по сравнению с посимвольным кодированием, описанным выше. Взяв J достаточно большим, можно сделать эффективность процедуры кодирования, измеренную отношением $JH(X)/N$, как угодно близкой к единице. Методы кодирования, описанные выше, не приводят к искажениям, так как кодирование символов источника или блоков таких символов в кодовые слова выполняется однозначно (уникально). Такие типы кодов названы бесшумными.

Теперь предположим, что мы пытаемся уменьшить скорость кодирования R путем смягчения условия однозначности процесса кодирования. Например, предположим, что только доля L^J блоков символов источника кодируется однозначно. Конкретно, выберем 2^N-1 наиболее вероятных J-символьных блоков и будем кодировать каждый из них однозначно, в то время как оставшиеся $L^J - (2^N - 1)$ блоков длины J будем представлять одним оставшимся кодовым словом. Эта процедура кодирования вызовет ошибку декодирования каждый раз, когда источник выдаст такой маловероятный блок. Пусть P_e означает вероятность такой ошибки. Отталкиваясь от этой процедуры кодирования, Шеннон (1948) доказал следующую теорему кодирования источника.

Теорема кодирования источника I. Пусть X – это ансамбль символов ДИБП с конечной энтропией $H(X)$. Блоки из J символов источника кодируются в двоичные кодовые слова длиной N. Для любого $\varepsilon>0$ вероятность P_e ошибки декодирования можно сделать сколь угодно малой, если

$$R = \frac{N}{J} \ge H(X) + \varepsilon \qquad (3.3.5)$$

J достаточно велико.

Наоборот, если

---

[1] Этот параметр не следует путать со скоростью передачи информации от двоичного источника, используемой, в частности, в гл. 4. По своему смыслу используемый здесь параметр R можно было бы назвать «траты (на кодирование)» (прп).

$$R \le H(X) - \varepsilon, \qquad (3.3.6)$$

тогда P_e сколь угодно близка к 1 при достаточно больших J.

Исходя из этой теоремы мы видим, что среднее число бит на символ источника, требуемое для кодирования выхода ДИБП с произвольно малой вероятностью ошибки декодирования, ограничено снизу энтропией источника $H(X)$. С другой стороны, если $R < H(X)$, вероятность ошибки декодирования приближается к 100 %, если J произвольно увеличивать.

Кодовые слова переменной длины. Если символы источника неравновероятны, более эффективный метод кодирования сводится к использованию кодовых слов переменной длины. Примером такого кодирования является код Морзе, который восходит к девятнадцатому веку. В коде Морзе символам, возникающим более часто, сопоставляются более короткие кодовые слова, а символам, возникающим менее часто, сопоставляются более длинные кодовые слова. Следуя этой общей идее, мы можем учесть вероятности различных символов источника при выборе кодовых слов. Проблема в том, чтобы предложить метод выбора кодовых слов для символов источника. Этот метод кодирования назван энтропийным кодированием.

Таблица 3.3.1. Коды переменной длины

Символ	$P(a_k)$	Код I	Код II	Код III
a_1	1/2	1	0	0
a_2	1/4	00	10	01
a_3	1/8	01	110	011
a_4	1/8	10	111	111

Для примера предположим, что выходные символы ДИБП a_1, a_2, a_3, a_4 с соответствующими вероятностями $P(a_1) = \frac{1}{2}$, $P(a_2) = \frac{1}{4}$, $P(a_3) = P(a_4) = \frac{1}{8}$ кодируются так, как показано в табл. 3.3.1. Код I имеет переменную длину и имеет принципиальный недостаток. Чтобы увидеть этот недостаток, предположим, что мы приняли последовательность 001001... Ясно, что 00 декодируется как a_2. Однако последующие четыре бита декодируются неоднозначно. Они могут декодироваться или как a_4a_3, или как $a_1a_2a_1$. Возможно, неоднозначность может быть разрешена путем ожидания последующих битов, но такое декодирование крайне нежелательно. Мы должны рассмотреть только коды, которые допускают немедленное декодирование, т.е. без задержки в декодере.

Код II в табл. 3.3.1 обеспечивает однозначное и немедленное декодирование. Удобно представлять кодовые слова этого кода графически как узлы на дереве, как показано на рис. 3.3.1. Видно, что 0 указывает на окончание кодового слова в первых трех кодовых словах. Эта характеристика вместе с тем обстоятельством, что ни одно кодовое слово не содержит более трех двоичных символов, делает этот код немедленно декодируемым. Заметим, что ни одно кодовое слово этого кода не является префиксом (началом) другого кодового слова. В общем, *префиксное условие* кода требует, чтобы для данного кодового слова $\mathbf{C}_k$ длины k с элементами $(b_1, b_2, \ldots b_k)$ не существовало других кодовых слов длины $l < k$ с элементами $(b_1, b_2, \ldots b_l)$ для $1 \le l \le k-1$.

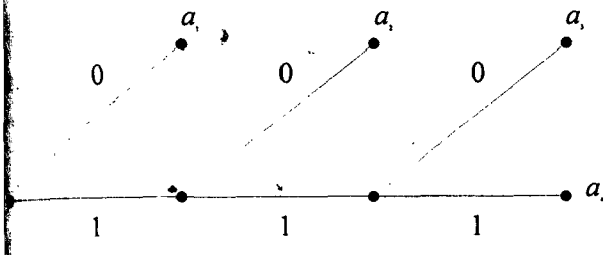


Рис. 3.3.1. Кодовое дерево для кода II в табл.3.3.1

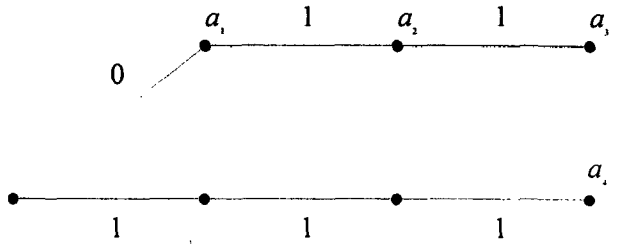


Рис. 3.3.2. Кодовое дерево для кода III в табл.3.3.1

Другими словами, нет кодовых слов длины $l<k$, которые совпадают с первыми l двоичными символами другого кодового слова длины $k>l$. Это свойство делает кодовые слова немедленно декодируемыми.

Код III из табл. 3.3.1 имеет кодовое дерево, показанное на рис. 3.3.2. Видим, что в этом случае имеет место однозначное декодирование, однако требующее задержки.

Ясно, что этот код не удовлетворяет префиксному условию.

Наша главная цель – создать систематическую процедуру для конструирования однозначных декодирующих кодов переменной длины, эффективных в том смысле, что среднее число бит на один символ источника, определяемое соотношением

$$\overline{R} = \sum_{k=1}^{L} n_k P(a_k), \qquad (3.3.7)$$

было бы минимальным. Условие существования кода переменной длины, которое удовлетворяет префиксному условию, дается неравенством Крафта.

Неравенство Крафта. Необходимым и достаточным условием существования двоичного кода с кодовыми символами длины $n_1 \le n_2 \le .... \le n_L$, удовлетворяющего условию префиксности, является

$$\sum_{k=1}^{L} 2^{-n^k} \le 1. \qquad (3.3.8)$$

Сначала мы докажем, что (3.3.8) является достаточным условием для существования префиксного кода. Чтобы построить такой код, мы начнем с полного двоичного дерева порядка $n = n_L$, которое имеет 2^n конечных узлов, причем от каждого узла порядка $k-1$ "растут" по два узла порядка k, $1 \le k \le n$.

Выберем некоторый узел порядка n_1 в качестве первого кодового слова $\mathbf{C}_1$. Этот выбор устраняет 2^{n-n_1} конечных узлов (т.е. долю 2^{-n_1} от 2^n конечных узлов). От остающихся доступных узлов порядка n_2 мы выбираем один узел для второго кодового слова $\mathbf{C}_2$. Этот выбор устраняет 2^{n-n_2} конечных узлов (т.е. долю 2^{-n_2} от 2^n конечных узлов). Этот процесс продолжается, пока последнее кодовое слово не определено в конечном узле $n = n_L$. Следовательно, в узле порядка $j < L$ доля числа отсечённых конечных узлов

$$\sum_{k=1}^{j} 2^{-n^k} < \sum_{k=1}^{L} 2^{-n^k} \le 1.$$

Всегда имеется узел порядка $k > j$, который может быть выбран для следующего слова. Таким образом, мы создали кодовое дерево, которое встроено в полное дерево из 2^n узлов, как иллюстрируется на рис. 3.3.3, для дерева, имеющего 16 конечных узлов, и источника, состоящего из пяти символов, отображаемых кодовыми словами длиной $n_1 = 1$, $n_2 = 2$, $n_3 = 3$ и $n_4 = n_5 = 4$.

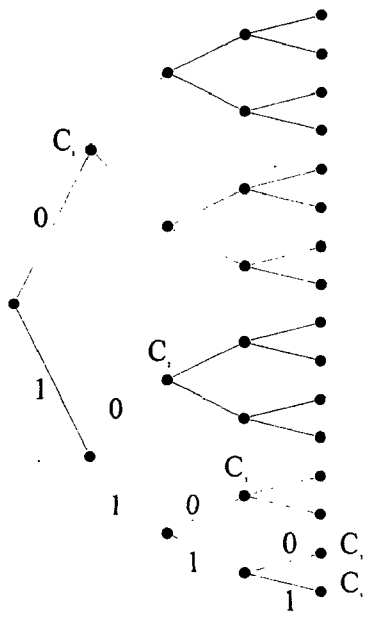


Рис. 3.3.3. Конструирование двоичного дерева, встроенного в полное дерево

Чтобы доказать, что (3.3.8) является необходимым условием, мы заметим, что в дер порядка $n = n_L$ число конечных узлов, отсечённых от общего числа 2^n конечных узл равно

$$\sum_{k=1}^{L} 2^{n-n_k} \le 2^n \text{. Следовательно,} \qquad \sum_{k=1}^{L} 2^{-n_k} \le 1,$$

и доказательство (3.3.8) закончено. Неравенство Крафта можно использовать д доказательства следующей теоремы кодирования источника (без шумов), котор применяется к кодам, удовлетворяющим префиксному условию.

Теорема кодирования источника II. Пусть X – ансамбль символов двоично источника без памяти с конечной энтропией $H(X)$ и выходными символами x_k, $1 \le k \le$ с соответствующими вероятностями выбора p_k, $1 \le k \le L$. Существует возможность созда код, который удовлетворяет префиксному условию и имеет среднюю длину $\overline{R}$, котор удовлетворяет неравенству

$$H(X) \le \overline{R} < H(X) + 1. \qquad (3.3.9)$$

Чтобы установить нижнюю границу в (3.3.9), обратим внимание на то, что для кодов слов, которые имеют длину n_k, $1 \le k \le L$, разность $H(X) - \overline{R}$ может быть выражена в вид

$$H(X) - \overline{R} = \sum_{k=1}^{L} p_k \log_2 \frac{1}{p_k} - \sum_{k=1}^{L} p_k n_k = \sum_{k=1}^{L} p_k \log_2 \frac{2^{-n_k}}{p_k}. \qquad (3.3.10)$$

Используя неравенство $\ln x \le x - 1$, из (3.3.10) находим

$$H(X) - \overline{R} \le (\log_2 e) \sum_{k=1}^{L} p_k \left(\frac{2^{-n_k}}{p_k} - 1 \right) \le (\log_2 e) \left(\sum_{k=1}^{L} 2^{-n_k} - 1 \right) \le 0,$$

где последнее неравенство следует из неравенства Крафта. Равенство имеет место, если только если $p_k = 2^{-n_k}$ для $1 \le k \le L$.

Верхняя граница в (3.3.9) может быть установлена при предположении что n_k, $1 \le k \le$ – целые числа, выбираемые из условия $2^{-n_k} \le p_k \le 2^{-n_k+1}$. Но если $p_k \ge 2$ просуммированы по $1 \le k \le L$, получаем неравенство Крафта, для которого демонстрировали, что там существует код, удовлетворяющий префиксному условию.

другой стороны, если мы берем логарифм $p_k < 2^{-n_k+1}$, получаем

$$\log p_k < -n_k + 1$$

или, что эквивалентно,

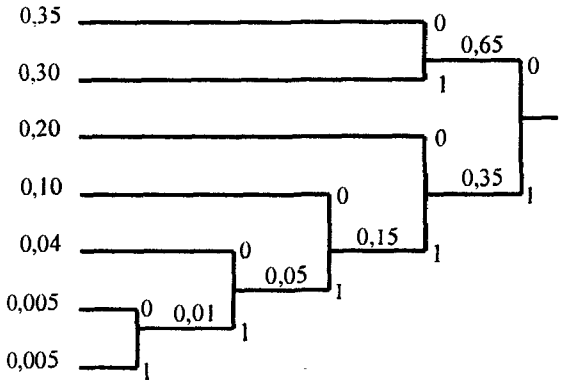
$$n_k < 1 - \log p_k . \qquad (3.3.11)$$

Если умножить обе части неравенства (3.3.11) на p_k и просуммировать по $1 \le k \le L$, получаем желательную верхнюю границу, данную в (3.3.9). Это завершает доказательство (3.3.9).

Мы установили, что коды переменной длины, которые удовлетворяют префиксному условию, – это эффективные коды для любого дискретного источника без памяти (ДИБП) с символами, имеющими различную априорную вероятность. Опишем теперь алгоритм для построения таких кодов.

Алгоритм кодирования Хаффмена. Хаффмен (1952) разработал алгоритм кодирования переменной длины, основанный на знании априорных вероятностей символов $P(x_i)$, $i = 1,2...,L$. Этот алгоритм оптимален в том смысле, что среднее число двоичных символов, требуемых для представления исходных символов, минимально. Получаемые кодовые слова удовлетворяют префиксному условию, определенному выше, что позволяет уникально и мгновенно декодировать полученную последовательность. Мы проиллюстрируем этот алгоритм кодирования посредством двух примеров.

Пример 3.3.1. Рассмотрим ДИБП с семью возможными символами $x_1, x_2, ..., x_7$, имеющими вероятности выбора, иллюстрируемые рис. 3.3.4.



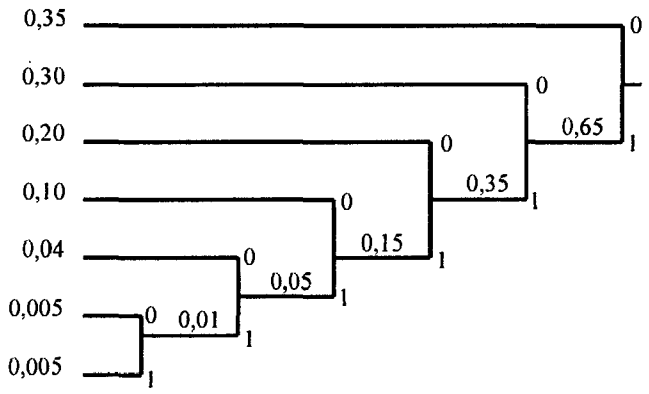
Символ	Вероятность	Собственная информация	Код
x_1	0,35	1,5146	00
x_2	0,30	1,7370	01
x_3	0,20	2,3219	10
x_4	0,10	3,3219	110
x_5	0,04	4,6439	1110
x_6	0,005	7,6439	11110
x_7	0,005	7,6439	11111
	$H(X) = 2,11$	$\bar{R} = 2,21$	

Рис. 3.3.4. Пример кодирования ДИБП кодом переменной длины

Мы упорядочили символы источника в порядке убывания вероятностей, т.е. $P(x_1) \ge P(x_2) \ge ... \ge P(x_7)$. Процесс кодирования начинаем с двух наименее вероятных символов x_6 и x_7. Эти два символа объединяем, как показано на рис. 3.3.4, причем верхнему ветвлению присваиваем «0», а нижнему – «1». Вероятности этих двух ветвей складываются, и общему узлу присваивается суммарная вероятность, равная в данном случае 0,01. Теперь мы имеем исходные символы $x_1, x_2, ..., x_5$ плюс новый символ, обозначим его x_6', полученный объединением x_6 и x_7. На следующем шаге снова объединяются два наименее вероятных символа из набора $x_1, x_2, x_3, x_4, x_5, x_6'$. Это x_5 и x_6', которые имеют объединенную вероятность 0,05. Переходу от x_5 присваиваем «0», а переходу от x_6' – «1». Эта процедура продолжается, пока мы не исчерпаем все возможные символы источника. Результат – кодовое дерево с ветвями, которые содержат требуемые

кодовые слова. Кодовые слова получаются, если двигаться от самого правого узла дерева переходя к самому левому узлу. Результирующие кодовые слова приведены на рис. 3.3. Среднее число двоичных элементов на символ этого кода $\overline{R} = 2{,}21$ бит/символ. Энтроп источника – 2,11 бит/символ.

Заметим, что полученный код не единственно возможный. Например, предпоследнем шаге процедуры кодирования мы имеем равный выбор между x_1 и x имеющими одинаковые вероятности. В этом пункте мы соединили x_1 и x_2. альтернативном коде мы можем соединить x_2 и x_3'. Результирующий код для этого случ иллюстрируется на рис. 3.3.5.



Символ	Код
x_1	0
x_2	10
x_3	110
x_4	1110
x_5	11110
x_6	111110
x_7	111111

$\overline{R} = 2{,}21$

Рис. 3.3.5. Альтернативный код для ДИБП в примере 3.3.1

Среднее число бит на символ для этого кода также равно 2.21. Следовательно полученные коды одинаково эффективны. Кроме того, назначение «0» верхнему переходу «1» нижнему (менее вероятному) переходу выбрано произвольно. Мы можем прос поменять местами 0 и 1 и получить ещё эффективный код, удовлетворяющий префиксно условию.

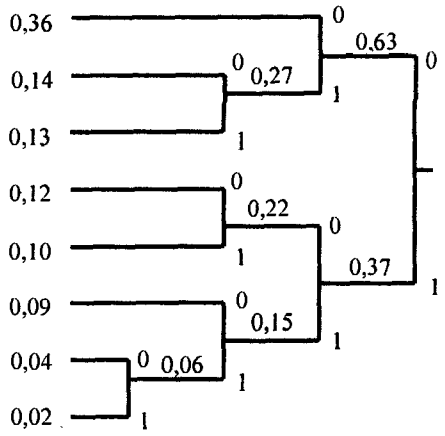
Пример 3.3.2. В качестве второго примера определим код Хаффмена для выхо ДИБП, иллюстрируемый на рис. 3.3.6. Энтропия этого источника $H(X) = 2{,}63$ бит/симво Код Хаффмена, показанный на рис. 3.3.6, имеет среднюю длину $\overline{R} = 2{,}70$ бит/симво Следовательно, его эффективность составляет 0,97.

Алгоритм кодирования переменной длины (Хаффмена), описанный в предыдущ примерах, генерирует префиксный код, имеющий среднюю длину $\overline{R}$, котор удовлетворяет (3.3.9). Однако вместо посимвольного кодирования более эффективно является процедура, основанная на кодировании блоков из J символов одновременно. таком случае границы в (3.3.9) в теореме кодирования источника II становятся другими:

$$JH(X) \le \overline{R}_J < JH(X) + 1, \qquad (3.3.12)$$

так как энтропия J-символьного блока от ДИБП равна $JH(X)$, и $\overline{R}_J$ – среднее число бит в J-символьном блоке. Если мы разделим (3.3.12) на J, то получим

$$H(X) \le \frac{\overline{R}_J}{J} < H(X) + \frac{1}{J}, \qquad (3.3.13)$$



Символ	Код
x_1	00
x_2	010
x_3	011
x_4	100
x_5	101
x_6	110
x_7	1110
x_8	1111
$H(X) = 2{,}63$	$\overline{R} = 2{,}70$

Рис. 3.3.6. Код Хаффмена для примера 3.3.2

где $\overline{R}_J / J \equiv \overline{R}$ – среднее число битов на исходный символ. Следовательно, $\overline{R}$ можно сделать как угодно близким к $H(X)$, выбирая J достаточно большим.

Пример 3.3.3. Выход ДИБП состоит из символов x_1, x_2 и x_3 с вероятностями 0,45, 0,35 и 0,20 соответственно. Энтропия этого источника $H(X)$=1,518 бит/символ. Код Хаффмена для этого источника, данный в табл. 3.3.2, требует $\overline{R}_1$=1,55 бит/символ и приводят к эффективности 97,9 %, Если посредством алгоритма Хаффмена символы закодированы парами, результирующий код выглядит так, как показано в табл. 3.3.3. Энтропия источника для пар символов $2H(X) = 3{,}036$ бит/пара символов. С другой стороны, код Хаффмена требует $\overline{R}_2 = 3{,}0675$ бит/пара символов. Таким образом, эффективность кодирования увеличилась до $2H(X)/\overline{R}_2 = 0{,}990$ (до 99,0 %).

Таблица 3.3.2. Код Хаффмена для примера 3.3.3

Символ	Вероятность	Собственная информация	Код
x_1	0,45	1,156	1
x_2	0,35	1,520	00
x_3	0,20	2,33	01
	$H(X)$=1,518 бит/символ		
	R_1=1,55 бит/символ		
	Эффективность 97,9 %		

Итак, мы продемонстрировали, что эффективное кодирование для ДИБП может быть выполнено на посимвольной основе, если использовать неравномерный код, основанный на алгоритме Хаффмена. Кроме того, эффективность процедуры кодирования увеличивается при кодировании блоков из J символов одновременно. Таким образом, выход ДИБП с энтропией $H(X)$ может быть закодирован неравномерным кодом со средним числом битов на исходный символ, которое может быть сделано как угодно близким к $H(X)$.

Таблица 3.3.3. Код Хаффмена для кодирования пар символов

Пара символов	Вероятность	Собственная информация	Код
$x_1\,x_1$	0,2025	2,312	10
$x_1\,x_2$	0,1575	2,676	001
$x_2\,x_1$	0,1575	2,676	010
$x_2\,x_2$	0,1225	3,039	011
$x_1\,x_3$	0,09	3,486	111
$x_3\,x_1$	0,09	3,486	0000
$x_2\,x_3$	0,07	3,850	0001
$x_3\,x_2$	0,07	3,850	1100
$x_3\,x_3$	0,04	4,660	1101

$2H(X)$=3,036 бит/пара символов; $\overline{R}_2 = 3{,}0675$ бит/пара символов

$\frac{1}{2}\overline{R}_2 = 1{,}534$ бит/символ; Эффективность 99,0 %

3.3.2. Дискретные стационарные источники

В предыдущем разделе мы описали эффективное кодирование выхода дискретного источника без памяти (ДИБП). В этом разделе мы рассмотрим дискретные источники, для которых последовательность символов выхода является статистически зависимой. Мы ограничим наше исследование источниками, которые являются статистически стационарными (однородными во времени).

Оценим энтропию некоторой последовательности символов от стационарного источника. Из определения в (3.2.13) и результата, данного в (3.2.15), энтропия блока случайных переменных $X_1 X_2 \ldots X_k$ равна

$$H(X_1X_2\ldots X_k) = \sum_{i=1}^{k} H(X_i \mid X_1X_2\ldots X_{i-1}), \qquad (3.3.14)$$

где $H(X_i \mid X_1X_2\ldots X_{i-1})$ – условная энтропия i-го символа при условии, что источник выдал предыдущие i–1 символов. Энтропия на символ для k-символьного блока определяется как

$$H_k(X) = \frac{1}{k} H(X_1X_2\ldots X_k). \qquad (3.3.15)$$

Мы определяем количество информации стационарного источника как энтропию на символ в (3.3.15) в пределе при $k\to\infty$, т.е.

$$H_\infty(X) = \lim_{k\to\infty} H_k(X) = \lim_{k\to\infty} \frac{1}{k} H(X_1X_2\ldots X_k). \qquad (3.3.16)$$

Существование этого предела установлено ниже.

В качестве альтернативы мы можем определять энтропию на символ источника как условную энтропию $H(X_k \mid X_1X_2\cdots X_{k-1})$ в пределе при $k\to\infty$. К счастью, этот предел также существует и идентичен пределу в (3.3.16). То есть

$$H_\infty(X) = \lim_{k\to\infty} H(X_k \mid X_1X_2\ldots X_{k-1}). \qquad (3.3.17)$$

Этот результат также установлен ниже. Наше изложение использует подход Галлагера (1968). Во-первых, мы покажем, что

$$H(X_k \mid X_1X_2\ldots X_{k-1}) \le H(X_{k-1} \mid X_1X_2\ldots X_{k-2}) \qquad (3.3.18)$$

для $k\ge 2$. С учётом предыдущего результата, согласно которому наложение условий на случайную переменную не может увеличивать её энтропию, мы имеем

$$H(X_k \mid X_1 X_2 \ldots X_{k-1}) \le H(X_k \mid X_2 X_3 \ldots X_{k-1}). \qquad (3.3.19)$$

В силу стационарности источника имеем

$$H(X_k \mid X_2 X_3 \ldots X_{k-1}) = H(X_{k-1} \mid X_1 X_2 \ldots X_{k-2}). \qquad (3.3.20)$$

Следовательно, (3.3.18) следует немедленно. Этот результат демонстрирует, что $H(X_k \mid X_1 X_2 \cdots X_{k-1})$ – не возрастающая последовательность (с ростом k).

Во-вторых, мы имеем результат

$$H_k(X) \ge H(X_k \mid X_1 X_2 \cdots X_{k-1}), \qquad (3.3.21)$$

оторый следует непосредственно из (3.3.14) и (3.3.15) и того факта, что последний член в умме (3.3.14) является нижней границей для каждого из остальных k–1 членов.

В-третьих, по определению $H_k(X)$ мы можем записать

$$H_k(X) = \frac{1}{k}[H(X_1 X_2 \ldots X_{k-1}) + H(X_k \mid X_1 \ldots X_{k-1})] =$$

$$= \frac{1}{k}[(k-1)H_{k-1}(X) + H(X_k \mid X_1 \ldots X_{k-1})] \le \frac{k-1}{k} H_{k-1}(X) + \frac{1}{k} H_k(X),$$

то приводит к

$$H_k(X) \le H_{k-1}(X). \qquad (3.3.22)$$

Следовательно, $H_k(X)$ – не возрастающая последовательность (с ростом k).

Поскольку $H_k(X)$ и условная энтропия $H(X_k \mid X_1 X_2 \ldots X_{k-1})$ не отрицательны и не озрастающие (с ростом k), оба предела должны существовать. Их предельные выражения огут быть установлены с использованием (3.3.14) и (3.3.15), чтобы выразить $H_{k+j}(X)$ как

$$H_{k+j}(X) = \frac{1}{k+j} H(X_1 X_2 \ldots X_{k-1}) +$$

$$+ \frac{1}{k+j}[H(X_k \mid X_1 \ldots X_{k-1}) + H(X_{k+1} \mid X_1 \ldots X_k) + \ldots + H(X_{k+j} \mid X_1 \ldots X_{k+j-1})].$$

Так как условная энтропия не возрастает, первый член в квадратных скобках является рхней границей для других слагаемых. Следовательно,

$$H_{k+j}(X) \le \frac{1}{k+j} H(X_1 X_2 \ldots X_{k-1}) + \frac{j+1}{k+j} H(X_k \mid X_1 X_2 \ldots X_{k-1}). \qquad (3.3.23)$$

Для фиксированного k в пределе для (3.3.23) при $j \to \infty$ получаем

$$H_\infty(X) \le H(X_k \mid X_1 X_2 \ldots X_{k-1}). \qquad (3.3.24)$$

Но (3.3.24) справедливо для всех k; следовательно, это справедливо и для $k \to \infty$. этому

$$H_\infty(X) \le \lim_{k\to\infty} H(X_k \mid X_1 X_2 \ldots X_{k-1}). \qquad (3.3.25)$$

С другой стороны, с учётом (3.3.21) мы получаем в пределе для $k \to \infty$

$$H_\infty(X) \ge \lim_{k\to\infty} H(X_k \mid X_1 X_2 \ldots X_{k-1}), \qquad (3.3.26)$$

устанавливает (3.3.17).

Теперь предположим, что мы имеем дискретный стационарный источник, который даёт J символов с энтропией на символ $H_J(X)$. Мы можем кодировать следовательность J символов кодом Хаффмена переменной длины, который овлетворяет префиксному условию при использовании процедуры, описанной в дыдущем разделе. Результирующий код имеет среднее число бит для блока с J

символами, который удовлетворяет условию

$$H(X_1 \ldots X_J) \le \overline{R_J} < H(X_1 \ldots X_J) + 1. \qquad (3.3.27)$$

Деля обе части (3.3.27) на J, мы получаем границы для среднего числа $\overline{R} = \overline{R_J}/J$ бит на исходный символ как

$$H_J(X) \le \overline{R} < H_J(X) + \frac{1}{J}. \qquad (3.3.28)$$

Увеличивая размер блока J, мы можем приближаться к $H_J(X)$ сколь угодно близко, и в пределе, когда $J \to \infty$, R удовлетворяет соотношению

$$H_\infty(X) \le \overline{R} < H_\infty(X) + \varepsilon, \qquad (3.3.29)$$

где ε стремится к нулю как $1/J$. Таким образом, эффективное кодирование стационарных источников может быть выполнено, если кодировать большие блоки символов в кодовые слова. Мы должны подчеркнуть, однако, что конструкция кода Хаффмена требует знания совместных ФПВ для J-символьных блоков.

3.3.3. Алгоритм Лемпела–Зива

Из нашего предшествующего обсуждения следует, что алгоритм кодирования Хаффмена приводит к оптимальному кодированию источника в том смысле, что кодовые слова удовлетворяют префиксному условию и средняя длина кодового блока минимальна. Конструируя код Хаффмена для ДИБП, мы должны знать вероятности появления всех исходных символов. В случае дискретного источника с памятью мы должны знать совместные вероятности всех блоков длины $n \ge 2$. Однако на практике статистика выхода источника чаще всего неизвестна. В принципе возможно оценить вероятности выхода дискретного источника, наблюдая длинную информационную последовательность, выдаваемую источником, и получая требуемые вероятности опытным путем. Такой метод пригоден для оценки вероятностей отдельных символов $\{p_k\}$. Однако вычислительная сложность оценки совместных вероятностей чрезвычайно высока. Следовательно, использование метода кодирования Хаффмена для многих реальных источников с памятью вообще непрактично.

В отличие от алгоритма кодирования Хаффмена алгоритм кодирования Лемпела–Зива разработан так, чтобы быть независимым от статистики источника. Следовательно, алгоритм Лемпела–Зива принадлежит классу *универсальных алгоритмов кодирования источника*. Это – алгоритм переменно-фиксированной длины, а кодирование выполняется так, как описано ниже.

В алгоритме Лемпела–Зива последовательность с выхода дискретного источника делится на блоки переменной длины, которые называются *фразами*. Каждая новая фраза представляет собой последовательность символов источника, отличающуюся от некоторой предыдущей фразы в последнем символе. Фразы перечислены в словаре, который сохраняет расположение существующих фраз. При кодировании новой фразы мы просто определяем адрес существующей фразы в словаре и добавляем в конец новый символ.

Как пример рассмотрим бинарную последовательность

101011010010011101010000110011101011000011011.

Деление последовательности, как описано выше, производит следующие фразы:

1, 0, 10, 11, 01, 00. 100, 111, 010, 1000, 011, 001, 110, 101, 10001, 1011.

Мы видим, что каждая фраза в последовательности – соединение одной из предыдущих фраз с новым выходным символом источника. Для кодирования фразы мы конструируем словарь, как показано в табл. 3.3.4.

Таблица 3.3.4. Словарь для алгоритма Лемпела–Зива

	Расположение в словаре	Содержимое словаря	Кодовое слово
1	0001	1	00001
2	0010	0	00000
3	0011	10	00010
4	0100	11	00011
5	0101	01	00101
6	0110	00	00100
7	0111	100	00110
8	1000	111	01001
9	1001	010	01010
10	1010	1000	01110
11	1011	011	01011
12	1100	001	01101
13	1101	110	01000
14	1110	101	00111
15	1111	10001	10101
16		1011	11101

Ячейки словаря пронумерованы последовательно, начиная с 1 и далее, в данном случае до 16, что является числом фраз в последовательности. Различные фразы, соответствующие каждой ячейке, также перечислены, как показано в таблице. Кодовые слова конструируются путём соединения двух частей. Первая часть представляет собой номер ячейки словаря (в двоичной форме) предыдущей фразы, которая соответствует новой фразе, кроме последнего символа. Вторая часть – это новый символ, выданный источником. Он добавляется в конец к первой части, т.е. к номеру ячейки предыдущей фразы. Первоначальный номер ячейки 0000 используется, чтобы кодировать «пустую» фразу.

Декодер источника создает идентичную таблицу на приемном конце системы связи и соответственно декодирует полученную последовательность.

Можно заметить, что таблица закодировала 44 исходных бита в 16 кодовых слов по пять битов каждый, что привело к 80 кодированным битам. Следовательно, алгоритм вообще не обеспечил никакое сжатие данных. Однако неэффективность является следствием того, что последовательность, которую мы рассмотрели, очень коротка. По мере увеличения длины последовательности процедура кодирования становится более эффективной и приводит к сжатию последовательности на выходе источника.

Как мы выбираем полную длину таблицы? Вообще, независимо от размера таблицы, она в конечном счёте переполнится. Чтобы решить проблему переполнения, кодер источника и декодер источника должны согласованно удалять фразы из соответствующих словарей, которые больше не используются и подставить новые фразы на их место.

Алгоритм Лемпела–Зива широко используется при сжатии компьютерных файлов. «Сжимающие» и «разжимающие» программы (утилиты) в операционной системе UNIX® и многочисленные алгоритмы в операционной системе MS DOS являются воплощениями различных версий этого алгоритма.

3.4. КОДИРОВАНИЕ ДЛЯ АНАЛОГОВЫХ ИСТОЧНИКОВ – ОПТИМАЛЬНОЕ КВАНТОВАНИЕ

Как отмечено в разд. 3.1, аналоговый источник выдаёт непрерывный сигнал $x(t)$, который является выборочной функцией случайного процесса $X(t)$. Если $X(t)$ является стационарным случайным процессом с ограниченной полосой, теорема отсчётов позволяет нам представить $X(t)$ последовательностью отсчётов, выбираемых равномерно со скоростью Найквиста.

Применяя теорему отсчётов, выход аналогового источника преобразуется в эквивалентную дискретную во времени последовательность отсчётов. Затем отсчёты квантуются по уровням и кодируются. Один тип простого кодирования – представление каждого дискретного уровня амплитуды последовательностью двоичных символов. Следовательно, если мы имеем L уровней, нам необходимы $R = \log_2 L$ бит/отсчёт (если L есть степень числа 2) или $R = \lfloor \log_2 L \rfloor + 1$ (в противном случае). Если уровни не равновероятны, но вероятности уровней на выходе источника известны, мы можем использовать процедуру кодирования Хаффмена (называемую также энтропийным кодированием), чтобы улучшить эффективность процесса кодирования.

Квантование амплитуд дискретизированного во времени сигнала обеспечивает сжатие данных, но это также приводит к некоторому искажению формы сигнала или потере его точности. Минимизация этих искажений является предметом рассмотрения в данном разделе. Многие результаты, данные в этом разделе, непосредственно применимы к дискретному во времени, непрерывному по амплитуде гауссовскому источнику без памяти. Такой источник служит хорошей моделью для нахождения остаточной ошибки в ряде методов кодирования источника, описанных в разд. 3.5.

3.4.1. Функция скорость-искажение $R(D)$

Начнём обсуждение квантования сигналов с рассмотрения погрешности представления отсчётов сигнала от информационного источника фиксированным числом символов (битов). Под термином «искажение» мы понимаем некоторую меру разности между фактическими выборками источника $\{x_k\}$ и соответствующими квантованными значениями $\tilde{x}_k$, которую мы обозначаем $d\{x_k, \tilde{x}_k\}$. Например, обычно используемая мера искажения – квадрат ошибки, определенная как

$$d(x_k, \tilde{x}_k) = (x_k - \tilde{x}_k)^2, \tag{3.4.1}$$

и используемое для определения ошибки квантования при ИКМ в разд. 3.5.1. Другие меры искажения могут принимать более общую форму:

$$d(x_k, \tilde{x}_k) = |x - \tilde{x}_k|^p, \tag{3.4.2}$$

где p принимает значения из ряда положительных целых чисел. Случай $p = 2$ имеет предпочтительную математическую трактовку.

Если $d\{x_k, \tilde{x}_k\}$ – мера искажения на отсчёт, искажение между последовательностью n отсчётов $\mathbf{X}_n$ и соответствующими n квантованными значениями $\tilde{\mathbf{X}}_n$ является средним значением искажения по n отсчётам, т.е.

$$d(\mathbf{X}_n, \tilde{\mathbf{X}}_n) = \frac{1}{n} \sum_{k=1}^{n} d(x_k, \tilde{x}_k). \tag{3.4.3}$$

На выходе источника имеет место случайный процесс, и, следовательно, n отсчётов в $\mathbf{X}_n$ являются случайными величинами. Поэтому $d(\mathbf{X}_n, \tilde{\mathbf{X}}_n)$ – случайная величина. Её математическое ожидание определяет искажение D, т.е.

$$D = E[d(\mathbf{X}_n, \widetilde{\mathbf{X}}_n)] = \frac{1}{n}\sum_{k=1}^{n} E[d(x_k, \widetilde{x}_k)] = E[d(x, \widetilde{x})], \qquad (3.4.4)$$

где последнее равенство следует из предположения, что исходный процесс является стационарным.

Теперь предположим, что мы имеем источник без памяти с непрерывно-амплитудным выходом $\mathbf{X}$, который имеет ФПВ отсчёта $p(x)$, квантованный амплитудный алфавит $\widetilde{\mathbf{X}}$ и меру искажения на отсчёт $d\{x_k, \widetilde{x}_k\}$, где $x \in \mathbf{X}$ и $\widetilde{x} \in \widetilde{\mathbf{X}}$. Тогда минимальная скорость в битах на отсчёт, требуемая для представления выхода $\mathbf{X}$ источника без памяти с искажением, меньшим или равным D, называется функцией скорость-искажение и определяется как

$$R(D) = \min_{p(\widetilde{x}|x): E[d(\mathbf{X},\widetilde{\mathbf{X}})] \le D} I(\mathbf{X}, \widetilde{\mathbf{X}}), \qquad (3.4.5)$$

где $I(\mathbf{X}, \widetilde{\mathbf{X}})$ – средняя взаимная информация между $\mathbf{X}$ и $\widetilde{\mathbf{X}}$. Вообще, скорость $R(D)$ уменьшается при увеличении D или, наоборот, $R(D)$ увеличивается при уменьшении D.

Для гауссовской модели непрерывного по амплитуде информационного источника без памяти Шеннон доказал следующую фундаментальную теорему.

Теорема: Функция скорость-искажение для гауссовского источника без памяти (Шеннон, 1959а). Минимальная скорость кодирования, необходимая для представления выхода дискретного во времени, непрерывного по амплитуде гауссовского источника без памяти, при использовании в качестве меры искажения среднеквадратической ошибки на символ (односимвольная мера искажения)

$$R_g(D) = \begin{cases} \frac{1}{2}\log_2(\sigma_x^2 / D) & (0 \le D \le \sigma_x^2), \\ 0 & (D > \sigma_x^2), \end{cases} \qquad (3.4.6)$$

где σ_x^2 – дисперсия выхода. гауссовского источника.

Заметим, что (3.4.6) подразумевает, что, если искажение $D \ge \sigma_x^2$, никакой информации передавать не нужно. Конкретно при $D = \sigma_x^2$ для реконструкции сигнала достаточно воспроизвести нули. При $D > \sigma_x^2$ для реконструкции сигнала мы можем использовать статистически независимые гауссовские шумовые выборки с дисперсией $D - \sigma_x^2$. График функции $R_g(D)$ представлен на рис. 3.4.1.

Функция скорость-искажение $R(D)$ источника связана со следующей основной теоремой кодирования источника в теории информации.

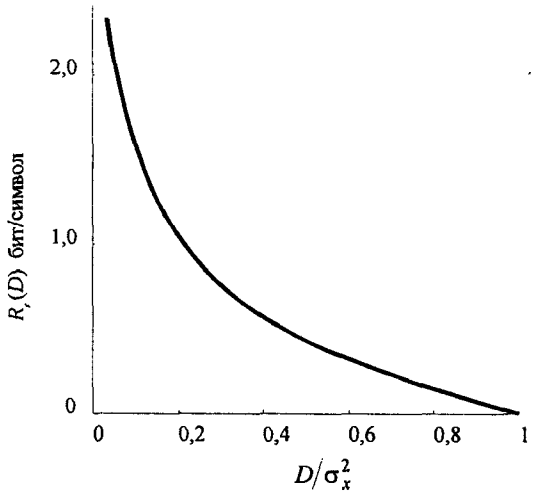


Рис. 3.4.1. Функция скорость-искажение для непрерывного по амплитуде гауссовского источника без памяти

Теорема: Кодирование источника с заданной мерой искажения (Шеннон, 1959а). Существует схема кодирования, которая отображает выход источника в кодовые слова так, что для любого данного искажения D минимальная скорость $R(D)$ бит на символ (на

отсчёт) источника является достаточной для восстановления исходного сигнала со средним искажением, которое является произвольно близким к D.

Это очевидно, потому что функция скорость-искажение $R(D)$ для любого источника представляет нижнюю границу скорости источника, которая является возможной для данного уровня искажения.

Вернёмся к результату в (3.4.6) для функции скорость-искажение гауссовского источника без памяти. Если мы поменяем функциональную зависимость между D и R, мы можем выразить D_g через R как

$$D_g(R) = 2^{-2R}\sigma_x^{\,2}. \qquad (3.4.7)$$

Эта функция называется функцией искажение-скорость для дискретного во времени гауссовского источника без памяти

Если искажение в (3.4.7) выразить в децибелах, мы получаем

$$10\log_{10} D_g(R) = -6R + 10\log_{10}\sigma_x^{\,2}. \qquad (3.4.8)$$

Заметим, что среднеквадратическое искажение уменьшается со скоростью 6 дБ/бит.

Явных выражений для функции скорость-искажение для негауссовских источников без памяти не существует. Однако имеются полезные верхние и нижние границы функции скорость-искажение для произвольного дискретного по времени, непрерывного по амплитуде источника без памяти. Верхняя граница даётся следующей теоремой.

Теорема: Верхняя граница для $R(D)$**.** Функция скорость-искажение непрерывного по амплитуде источника без памяти с нулевым средним и конечной дисперсией $\sigma_x^{\,2}$ при использовании среднеквадратичной меры искажений ограничена сверху величиной

$$R(D) \le \tfrac{1}{2}\log_2 \frac{\sigma_x^{\,2}}{D} \quad (0 \le D \le \sigma_x^{\,2}). \qquad (3.4.9)$$

Доказательство этой теоремы дано Бергером (1971). Подразумевается, что гауссовский источник требует максимальную скорость кодирования среди всех других источников при заданном уровне среднеквадратической ошибки. Следовательно, функция скорость-искажение $R(D)$ для произвольного непрерывного источника без памяти с нулевым средним и конечной дисперсией $\sigma_x^{\,2}$ удовлетворяет условию $R(D) \le R_g(D)$. Аналогично функция искажение-скорость того же источника удовлетворяет условию

$$D(R) \le D_g(R) = 2^{-2R}\sigma_x^{\,2}. \qquad (3.4.10)$$

Существует также нижняя граница функции скорость-искажение. Её называют нижней границей Шеннона для среднеквадратической ошибки искажения, и она определяется так:

$$R^*(D) = h(X) - \tfrac{1}{2}\log_2 2\pi e D, \qquad (3.4.11)$$

где $h(X)$ – дифференциальная энтропия источника без памяти с непрерывной амплитудой. Функция искажение-скорость, соответствующая (3.4.11), равна

$$D^*(R) = \frac{1}{2\pi e} 2^{-2[R-h(X)]}. \qquad (3.4.12)$$

Следовательно, функция скорость-искажение для произвольного источника без памяти с непрерывной амплитудой ограничена сверху и снизу:

$$R^*(D) \le R(D) \le R_g(D), \qquad (3.4.13)$$

и соответствующая функция искажение-скорость ограничена:

$$D^*(R) \le D(R) \le D_g(R). \qquad (3.4.14)$$

Дифференциальная энтропия гауссовского источника без памяти

$$h_g(X) = \tfrac{1}{2}\log_2 2\pi e\sigma_x^{\,2}, \tag{3.4.15}$$

так что нижняя граница $R^*(D)$ в (3.4.11) уменьшается до $R_g(D)$. Теперь, если выразить $D^*(R)$ в децибелах и нормировать к $\sigma_x^{\,2} = 1$ [или деля $D^*(R)$ на $\sigma_x^{\,2}$], мы получаем из (3.4.12)

$$10\log_{10} D^*(R) = -6R - 6[h_g(X) - h(X)] \tag{3.4.16}$$

или, что эквивалентно,

$$10\log_{10}\frac{D_g(R)}{D^*(R)} = 6[h_g(X) - h(X)]\ \text{дБ} = 6[R_g(D) - R^*(D)]\ \text{дБ}. \tag{3.4.17}$$

Соотношения в (3.4.16) и (3.4.17) позволяют сравнивать нижнюю границу искажений с верхней границей, которая определяет искажения для гауссовского источника. Обратим внимание, что $D^*(R)$ также уменьшается со скоростью -6 дБ/бит. Мы должны также отметить, что дифференциальная энтропия $h(X)$ ограничена сверху величиной $h_g(X)$, как показано Шенноном (1948b).

В табл. 3.4.1 даны четыре типа ФПВ, которые являются моделями распределения, обычно используемыми для источника сигнала. В таблице даны значения дифференциальной энтропии, различия в скорости (бит на отсчёт) и различия в искажении между верхней и нижней границами. Заметим, что гамма-распределение показывает самое большое отклонение от гауссовского. Распределение Лапласа наиболее близко к гауссовскому, а равномерное распределение занимает второе место по близости среди ФПВ, показанных в таблице. Эти результаты дают некоторое представление о различии между верхними и нижними границами искажений и скорости.

Перед завершением этого раздела рассмотрим гауссовский источник с ограниченной полосой частот со спектральной плотностью

$$\Phi(f) = \begin{cases} \sigma_x^{\,2}/2W & (|f| \le W), \\ 0 & (|f| > W). \end{cases} \tag{3.4.18}$$

Если выход этого источника дискретизирован с частотой Найквиста, его отсчёты некоррелированны и, так как источник гауссовский, они также статистически независимы.

Таблица 3.4.1. Дифференциальная энтропия и сравнение скорости и искажений четырёх распространённых ФПВ для моделей сигнала

ФПВ	$p(x)$	$h(X)$	$R_g(D) - R^*(D)$ бит/отсчёт	$D_g(R) - D^*(R)$ (дБ)
Гауссовское	$\frac{1}{\sqrt{2\pi}\sigma_x} e^{-x^2/2\sigma_x^2}$	$\frac{1}{2}\log_2 2\pi e\sigma_x^{\,2}$	0	0
Равномерное	$\frac{1}{2\sqrt{3}\sigma_x}$, $\lvert x\rvert \le \sqrt{3}\sigma_x$	$\frac{1}{2}\log_2 12\sigma_x^{\,2}$	0,255	1,53
Лапласа	$\frac{1}{\sqrt{2}\sigma_x} e^{-\sqrt{2}\lvert x\rvert/\sigma_x}$	$\frac{1}{2}\log_2 2e^2\sigma_x^{\,2}$	0,104	0,62
Гамма	$\frac{\sqrt[4]{3}}{\sqrt{8\pi\sigma_x\lvert x\rvert}} e^{-\sqrt{3}\lvert x\rvert/2\sigma_x}$	$\frac{1}{2}\log_2\left(4\pi e^{0{,}423}\sigma_x^{\,2}/3\right)$	0,709	4,25

Следовательно, эквивалентный дискретный во времени гауссовский источник является источником без памяти. Функция скорость-искажение для каждого отсчёта дается (3.4.6). Поэтому функция скорость-искажение для белого гауссовского источника с ограниченной полосой частот в бит/отсчёт равна

$$R_g(D) = W \log_2 \frac{\sigma_x^2}{D} \quad (0 \le D \le \sigma_x^2). \tag{3.4.19}$$

Соответствующая функция искажение-скорость

$$D_g(R) = 2^{-R/W} \sigma_x^2. \tag{3.4.20}$$

Выражая в децибелах и нормируя к σ_x^2, получаем

$$10 \log D_g(R)/\sigma_x^2 = -3R/W. \tag{3.4.21}$$

Большое количество случаев, в которых гауссовский процесс не является ни белым, ни с ограниченной полосой, было рассмотрено Галлагером (1968) и Гобликом и Холсингером (1967).

3.4.2. Скалярное квантование

При кодировании источника квантователь может быть оптимизирован, если известна ФПВ уровней сигнала на входе квантователя. Например, предположим, что последовательность $\{x_n\}$ на входе квантователя имеет ФПВ $p(x)$ и $L=2^R$ – желаемое число уровней квантования. Необходимо рассчитать оптимальный скалярный квантователь, который минимизирует некоторую функцию ошибки квантования $q = \tilde{x} - x$, где $\tilde{x}$ – квантованное значение x. Для дальнейшей разработки предположим, что $f(\tilde{x} - x)$ определяет желательную функцию ошибки. Тогда искажение, возникающее за счёт квантования сигнальных уровней, равно

$$D = \int_{-\infty}^{\infty} f(\tilde{x} - x) p(x) dx. \tag{3.4.22}$$

В общем, оптимальный квантователь минимизирует D путём оптимального выбора выходных уровней и входного диапазона для каждого выходного уровня. Эту оптимизационную проблему рассматривали Ллойд (1982) и Макс (1960), и полученный оптимальный квантователь назван квантователем Ллойда–Макса.

У равномерного квантователя выходные уровни определяются как $\tilde{x}_k = \frac{1}{2}(2k-1)\Delta$ для амплитуды входного сигнала в диапазоне $(k-1)\Delta \le x \le k\Delta$, где Δ – размер шага квантования. Если квантователь симметричен (относительно нуля) с конечным числом уровней, среднее искажение (3.4.22) может быть выражено в виде

$$D = 2 \sum_{k=1}^{L/2-1} \int_{(k-1)\Delta}^{k\Delta} f(\tfrac{1}{2}(2k-1)\Delta - x) p(x) dx + 2 \int_{(L/2-1)\Delta}^{\infty} f(\tfrac{1}{2}(2k-1)\Delta - x) p(x) dx. \tag{3.4.23}$$

В этом случае минимизация D выполняется с учётом параметра размера шага Δ.

Путём дифференцирования D по Δ получаем

$$\sum_{k=1}^{L/2-1} (2k-1) \int_{(k-1)\Delta}^{k\Delta} f(\tfrac{1}{2}(2k-1)\Delta - x) p(x) dx + (L-1) \int_{(L/2-1)\Delta}^{\infty} f'(\tfrac{1}{2}(L-1)\Delta - x) p(x) dx = 0, \tag{3.4.24}$$

где $f'(x)$ означает производную $f(x)$. При выборе критериальной функции ошибки $f(x)$ можно получить численное решение (3.4.24) для оптимального размера шага на компьютере для произвольной заданной ФПВ $p(x)$. Для среднеквадратичного критерия ошибки, кода $f(x)=x^2$, Макс(1960) рассчитал оптимальный размер шага $\Delta_{опт}$ и минимальное значение среднеквадратической ошибки, когда ФПВ $p(x)$ является гауссовской с нулевым средним и единичной дисперсией. Некоторые из этих результатов даны в табл. 3.4.2.

Таблица 3.4.2. Оптимальные размеры шага при равномерном квантовании гауссовских случайных величин

Число выходных уровней	Оптимальный размер шага $\Delta_{опт}$	Минимум СКО $D_{мин}$	10 lg $D_{мин}$ (дБ)
2	1,596	0,3634	–4,4
4	0,9957	0,1188	–9,25
8	0,5860	0,3744	–14,27
16	0,3352	0,01154	–19,38
32	0,1881	0,00349	–24,57

Видим, что минимальная среднеквадратическая ошибка D_{min} уменьшается немного больше, чем на 5 дБ, при каждом удвоении числа уровней L. Следовательно, каждый бит, который используется равномерным квантователем с оптимальным размером числа $\Delta_{опт}$ для гауссовского входного сигнала уменьшает искажение более чем на 5 дБ.

Если соблюдать условие, что квантователь равномерный, искажение можно дополнительно уменьшить. В этом случае мы выберем выходной уровень $\tilde{x} = \tilde{x}_k$, когда амплитуда входного сигнала находится в диапазоне $x_{k-1} \le x < x_k$. Для квантования с L уровнями крайними точками являются $x_0 = -\infty$ и $x_L = \infty$. Результирующее искажение

$$D = \sum_{k=1}^{L} \int_{x_{k-1}}^{x_k} f(\tilde{x}_k - x)\, p(x)\, dx \qquad (3.4.25)$$

снова минимизируется путём оптимального выбора $\{\tilde{x}_k\}$ и $\{x_k\}$.

Необходимые условия для минимальных искажений можно получить дифференцированием D по $\{x_k\}$ и $\{\tilde{x}_k\}$. Результат такой оптимизации выражается двумя уравнениями:

$$f(\tilde{x}_k - x_k) = f(\tilde{x}_{k+1} - x_k), \quad k = 1, 2, ..., L-1, \qquad (3.4.26)$$

$$\int_{x_{k-1}}^{x_k} f'(\tilde{x}_k - x)\, p(x)\, dx = 0, \quad k = 1, 2, ..., L. \qquad (3.4.27)$$

Как частный случай мы снова рассмотрим минимизацию среднеквадратических значений искажений. В этом случае, $f(x) = x^2$, и, следовательно, из (3.4.26) следует

$$x_k = \tfrac{1}{2}(\tilde{x}_k + \tilde{x}_{k+1}), \quad k = 1, 2, ..., L-1, \qquad (3.4.28)$$

что является среднеарифметическим $\tilde{x}_k$ и $\tilde{x}_{k+1}$. Соответствующие уравнения, определяющие $\{\tilde{x}_k\}$,

$$\int_{x_{k-1}}^{x_k} (\tilde{x}_k - x)\, p(x)\, dx = 0, \quad k = 1, 2, ..., L. \qquad (3.4.29)$$

Таким образом, $\tilde{x}_k$ является *центроидом* области $p(x)$ между x_{k-1} и x_k. Эти уравнения могут быть решены численно для произвольных ФПВ $p(x)$. Таблицы 3.4.3 и 3.4.4 дают результаты оптимизации Макса (1960) для оптимального четырёхуровневого и восьмиуровневого квантователя сигнала, распределённого по Гауссу с нулевым средним и единичной дисперсией.

Таблица 3.4.3. Оптимальный 4-уровневый квантователь для гауссовской случайной величины

Уровень k	x_k	$\tilde{x}_k$
1	–0,9816	–1,510
2	0,0	–0,4528
3	0,9816	0,4528
4	∞	1,510
$D_{\text{мин}}$ =0,1175		
10 lg $D_{\text{мин}}$ = –9,3 дБ		

Таблица 3.4.4. Оптимальный 8-уровневый квантизатор для гауссовской случайной величины (Макс, 1960)

Уровень k	x_k	$\tilde{x}_k$
1	–1,748	–2,152
2	–1,050	–1,344
3	–0,5006	–0,7560
4	0	–0,2451
5	0,5006	0,2451
6	1,050	0,7560
7	1,748	1,344
8	∞	2,152
$D_{\text{мин}}$ =0,03454		
10 lg $D_{\text{мин}}$ = –14,62 дБ		

В таблице 3.4.5 сравниваются минимальные среднеквадратические искажения для гауссовской амплитуды сигнала в равномерном и неравномерном квантователях. Из этой таблицы мы видим, что разница в характеристиках двух типов квантователей относительно мала для малых значений R (меньше чем 0,5 дБ для $R \le 3$), но она растёт с ростом R. Например: при R=5, неравномерный квантователь примерно на 1,5 дБ лучше равномерного.

Таблица 3.4.5. Сравнение оптимальных равномерного и неравномерного квантизаторов для гауссовской случайной величины (Макс, 1960; Паез и Глиссон, 1972)

R	10 lg $D_{\text{мин}}$	
(бит/отсчёт)	Равномерное (дБ)	Неравномерное (дБ)
1	–4,4	–4,4
2	–9,25	–9,30
3	–14,27	–14,62
4	–19,38	–20,22
5	–24,57	–26,02
6	–29,83	–31,89
7	–35,13	–37,81

Поучительно построить кривые зависимости минимальных искажений от битовой скорости $R = \log_2 L$ бит на отсчёт (на символ) источника для равномерного и неравномерного квантователей.

Эти кривые даны на рис. 3.4.2. Функциональную зависимость искажений D от битовой скорости R можно выразить как $D(R)$ – функцию искажение-скорость. Мы видим, что функция искажение-скорость для оптимального неравномерного квантователя лежит ниже, чем для равномерного квантователя. Поскольку квантователь превращает непрерывную амплитуду источника в дискретную, мы можем трактовать дискретные амплитуды как символы, скажем $\widetilde{x} = \{\widetilde{x}_k, 1 \le k \le L\}$ с соответствующими вероятностями $\{p_k\}$. Если отсчёты сигнала амплитуды статистически независимы, то на выходе квантователя имеем дискретный источник без памяти, и, следовательно, его энтропия

$$H(\widetilde{X}) = -\sum_{k=1}^{L} p_k \log_2 p_k . \qquad (3.4.30)$$

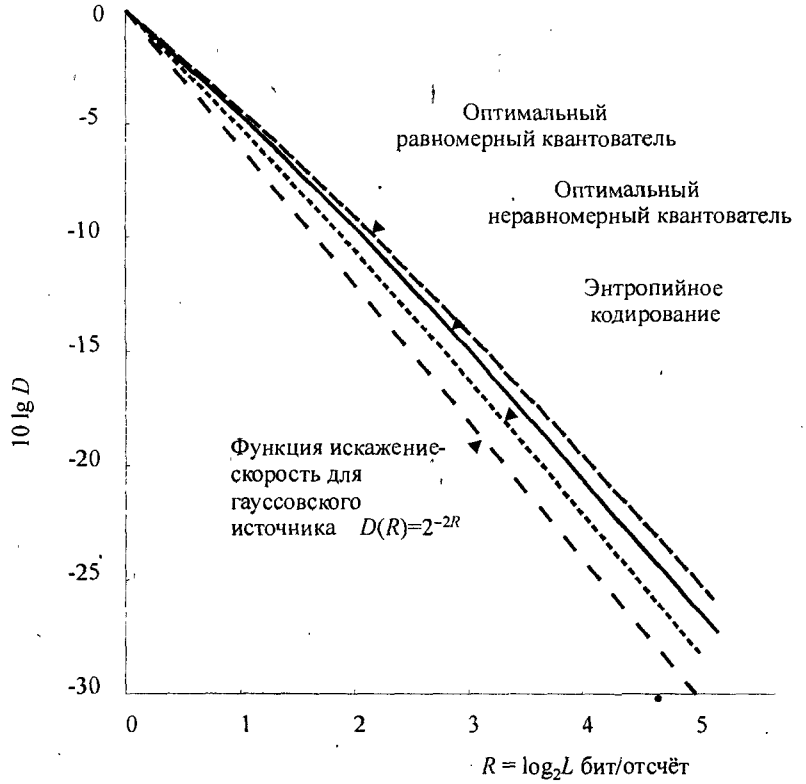


Рис. 3.4.2. Кривые зависимости искажение-скорость для гауссовского источника без памяти с дискретным временем

Для примера: оптимальный четырёхуровневый неравномерный квантователь для распределённой по Гауссу амплитуды приводит к вероятностям $p_1 = p_4 = 0{,}1635$ для двух внешних уровней и $p_2 = p_3 = 0{,}3365$ для двух внутренних уровней. В этом случае энтропия дискретного источника $H(\widetilde{X}) = 1{,}911$ бит/символ. Следовательно, при помощи энтропийного кодирования (кодирование Хаффмена) блоков выходных символов мы можем достичь минимальных искажений (–9,30 дБ) посредством 1,911 бит/символ вместо 2 бит/символ. Макс (1960) определил энтропию для дискретных символов источника после процесса квантования. Таблица 3.4.6 показывает значение энтропии при неравномерном квантовании. Зависимость $R(D)$ для этого случая также показана кривой на рис. 3.4.2 и обозначена как *энтропийное кодирование*.

Таблица 3.4.6. Энтропия выхода оптимального неравномерного квантователя гауссовской случайной величины (Макс, 1960)

R (бит/отсчёт)	Энтропия (бит/символ)	Искажения $10\lg D_{min}$
1	1,0	–4,4
2	1,911	–9,30
3	2,825	–14,62
4	3,765	–20,22
5	4,730	–26,02

Из этого обсуждения мы заключаем, что качество квантователя можно анализировать, когда известна ФПВ непрерывного выхода источника. Оптимальный квантователь с $L = 2^R$ уровнями обеспечивает минимальное искажение $D(R)$, где $R = \log_2 L$ бит/отсчёт. Такого уровня искажений можно достичь простым представлением каждого квантованного отсчёта R битами. Однако возможно более эффективное кодирование. Дискретные выходы квантователя характеризуются рядом вероятностей $\{p_k\}$, которые можно использовать для расчёта эффективных неравномерных кодов для выхода источника (энтропийное кодирование). Эффективность какого-либо метода кодирования можно сравнить с функцией искажение-скорость или, что эквивалентно, с функцией скорость-искажение для дискретного времени и непрерывных амплитуд источника, характеризуемого данной ФПВ. Если мы сравним характеристики оптимального неравномерного квантователя с функцией искажение-скорость, мы найдём, например, что для искажения в –26 дБ энтропийное кодирование требует скорость на 0,4 бит/отсчёт больше, чем минимальная скорость, даваемая (3.4.8), а простое блоковое кодирование каждого символа требует скорость на 0,68 бит/отсчёт больше, чем минимальная скорость. Мы также видим, что функция искажение-скорость для оптимального равномерного и неравномерного квантователей гауссовского источника асимптотически приближается к наклону –6 дБ/бит для больших R.

3.4.3. Векторное квантование

В предыдущих разделах мы рассмотрели квантование выходного сигнала непрерывного источника для случая, когда квантование выполняется последовательно по отдельным отсчётам, т.е. скалярное квантование. В этом разделе мы рассмотрим совместное квантование блока символьных отсчётов или блока сигнальных параметров. Этот вид квантования называется *блоковым* или *векторным квантованием*. Оно широко используется при кодировании речи в цифровых сотовых системах связи.

Фундаментальный результат теории искажения заключается в том, что лучшую характеристику можно достичь векторным, а не скалярным квантованием, даже если непрерывный источник без памяти. Если, кроме того, отсчёты сигнала или параметры сигнала статистически зависимы, мы можем использовать зависимость посредством совместного квантования блоков отсчётов или параметров и таким образом достичь большей эффективности (более низкой битовой скорости) по сравнению с той, которая достигается скалярным квантованием.

Проблему векторного квантования можно сформулировать так. Имеем n-мерный вектор $\mathbf{X} = \{x_1, x_2 \ldots x_n\}$ с n вещественными, непрерывными амплитудами компонент $\{x_k, 1 \le k \le n\}$, которые описываются СФПВ $p(x_1, x_2 .. x_n)$. Путём квантования вектор $\mathbf{X}$ превращается в другой n-мерный вектор $\widetilde{\mathbf{X}}$ с компонентами $\{\widetilde{x}_k, 1 \le k \le n\}$. Выразим операции квантования оператором $Q(.)$, так что

$$\widetilde{\mathbf{X}} = Q(\mathbf{X}), \qquad (3.4.31)$$

где $\widetilde{\mathbf{X}}$ – выход квантователя, когда на вход поступает вектор $\mathbf{X}$.

В принципе векторное квантование блоков данных можно рассматривать как проблему распознавания образов, включающую в себя классификацию блоков данных через дискретное количество категорий или ячеек в соответствии с некоторым критерием точности, таким, например, как среднеквадратическая погрешность. Для примера рассмотрим квантование двумерных векторов $\mathbf{X} = [x_1, x_2]$. Двумерное пространство разделяют на ячейки, как показано на рис. 3.4.3, где мы имеем произвольно выбранные шестиугольные ячейки $\{C_k\}$. Все входные векторы, которые попадают в ячейку C_k, квантуются в вектор $\widetilde{\mathbf{X}}_k$, который на рис. 3.4.3 отмечен как центр шестиугольника. В нашем примере иллюстрируются $L = 37$ векторов, один для каждой из 37 ячеек, на которые разбито двумерное пространство. Обозначим ряд возможных выходных векторов как $\{\widetilde{\mathbf{X}}_k, 1 < k \le L\}$.

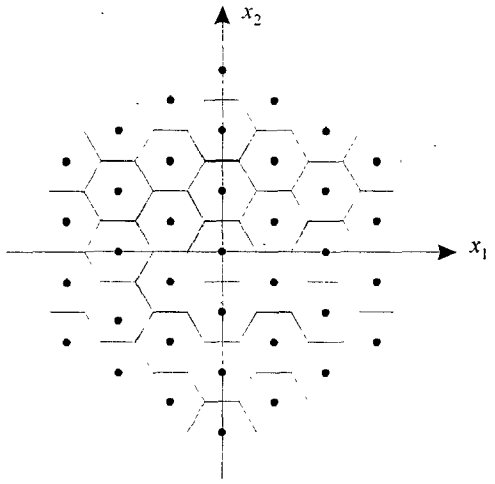


Рис. 3.4.3. Пример квантования в двухмерном пространстве

В общем, квантование n-мерного вектора $\mathbf{X}$ в n-мерный вектор $\widetilde{\mathbf{X}}$ ведёт к ошибке квантования или искажению $d(\mathbf{X}, \widetilde{\mathbf{X}})$. Среднее искажение по ряду входных векторов $\mathbf{X}$ равно

$$D = \sum_{k=1}^{L} P(\mathbf{X} \in C_k) E[d(\mathbf{X}, \widetilde{\mathbf{X}}_k) \mid \mathbf{X} \in C_k] = \sum_{k=1}^{L} P(\mathbf{X} \in C_k) \int_{X \in C_k} d(\mathbf{X}, \widetilde{\mathbf{X}}_k) p(\mathbf{X}) d\mathbf{X}^1, \qquad (3.4.32)$$

где $P(\mathbf{X} \in \mathrm{C_k})$ – вероятность того, что вектор $\mathbf{X}$ попадёт в ячейку C_k, а $p(\mathbf{X})$ – СФПВ n случайных величин. Как и в случае скалярного квантования, мы можем минимизировать D путём выбора ячеек $\{C_k, 1 < k \le L\}$ при заданной ФПВ $p(\mathbf{X})$. Обычно используемая мера искажений – среднеквадратическая ошибка (l_2 – норма) определяется как

$$d_2(\mathbf{X}, \widetilde{\mathbf{X}}) = \frac{1}{n}(\mathbf{X} - \widetilde{\mathbf{X}})^{\mathrm{T}}(\mathbf{X} - \widetilde{\mathbf{X}}) = \frac{1}{n}\sum_{k=1}^{n}(x_k - \widetilde{x}_k)^2 \qquad (3.4.33)$$

или, в более общем виде, взвешенная среднеквадратическая ошибка

---

[1] В интеграле (3.4.32) и далее обозначение $d\mathbf{X}$ следует понимать как $\prod_{k=1}^{n} dx_k$ – дифференциал объёма n-мерного пространства векторов $\mathbf{X}$, $\widetilde{\mathbf{X}}$, где x_k – элементы вектора $\mathbf{X}$ (прп).

$$d_{2W}(\mathbf{X},\tilde{\mathbf{X}}) = (\mathbf{X}-\tilde{\mathbf{X}})^{\mathrm{T}}\,\mathbf{W}(\mathbf{X}-\tilde{\mathbf{X}}), \qquad (3.4.34)$$

где $\mathbf{W}$ – положительно определённая взвешивающая матрица. Обычно мера $\mathbf{W}$ выбирается как обратная по отношению к матрице ковариаций входных данных $\mathbf{X}$.

Другая мера искажений, которая иногда используется, является частным случаем l_p нормы и определяется как

$$d_p(\mathbf{X},\tilde{\mathbf{X}}) = \frac{1}{n}\sum_{k=1}^{n} |\, x_k - \tilde{x}_k \,|^p . \qquad (3.4.35)$$

Частный случай, когда $p = 1$, часто используется как альтернатива случаю p=2.

Векторное квантование не ограничивается квантованием блока сигнальных отсчётов источника сигнала. Его можно использовать для квантования ряда параметров, извлечённых из данных. Например, при линейном кодировании с предсказанием (ЛКП), описанном в разделе 3.5.3, параметры, извлечённые из сигнала, являются коэффициентами предсказания, которые являются коэффициентами для всеполюсной фильтровой модели источника, который генерирует наблюдаемые данные. Эти параметры можно рассматривать как блок и квантовать как блок символов, используя некоторую подходящую меру искажений. В случае кодирования речи подходящей мерой искажений, которую предложили Итакура и Саити (1986, 1975), является взвешенная среднеквадратическая ошибка, где взвешивающая матрица $\mathbf{W}$ выбрана как нормированная матрица автоковариации Φ наблюдаемых данных.

При кодировании речи альтернативным рядом параметров, которые могут быть квантованы как блок и переданы к приёмнику, могут быть коэффициенты отражения (см. ниже) $\{a_{ij},\ 1 \le i \le m\}$.

Еще один ряд параметров, которые иногда используются для векторного квантования при линейном кодировании с предсказанием речи, содержит логарифмические отношения $\{r_k\}$, которые выражаются через коэффициенты отражения

$$r_k = \log\frac{1+a_{kk}}{1-a_{kk}}, \quad 1 \le k \le m . \qquad (3.4.36)$$

Теперь вернемся к математической формулировке векторного квантования и рассмотрим разбиение n-мерного пространства на L ячеек $\{C_k,\ 1 < k \le L\}$ с точки зрения минимизации среднего искажения по всем L-уровневым квантователям. Имеется два условия для минимизации. Первое заключается в том, что оптимальный квантователь использует селекцию по правилу ближайшего соседа, которое можно выразить математически как

$$Q(\mathbf{X}) = \tilde{\mathbf{X}}_k ,$$

если, и только если

$$D(\mathbf{X},\tilde{\mathbf{X}}_k) \le D(\mathbf{X},\tilde{\mathbf{X}}_j), \qquad k \ne j, \quad 1 \le j \le L . \qquad (3.4.37)$$

Второе условие, необходимое для оптимизации, заключается в том, что каждый выходной вектор $\tilde{\mathbf{X}}_k$ выбирается так, чтобы минимизировать среднее искажение в ячейке C_k. Другими словами, $\tilde{\mathbf{X}}_k$ – это вектор в C_k, который минимизирует

$$D_k = E[d(\mathbf{X},\tilde{\mathbf{X}}) \mid \mathbf{X} \in C_k] = \int_{\mathbf{X}\in C_k} d(\mathbf{X},\tilde{\mathbf{X}})\,p(\mathbf{X})\,d\mathbf{X} . \qquad (3.4.38)$$

Вектор $\tilde{\mathbf{X}}_k$, который минимизирует D_k, назван *центроидом* ячейки.

Таким образом, эти условия оптимизации определяют разбиение n-мерного

пространства на ячейки $\{C_k,\ 1 \le k \le L\}$, когда СФПВ $p(\mathbf{X})$ известна. Ясно, что указанные два условия обобщают задачу оптимального квантования скалярной величины оптимизации на случай квантования n-мерного вектора. В общем, мы ожидаем, что кодовые векторы более тесно группируются в областях, где СФПВ $p(\mathbf{X})$ велика, и, наоборот, разрежены в областях, где $p(\mathbf{X})$ мала.

В качестве верхней границы искажений векторного квантования мы можем использовать величину искажений оптимального скалярного квантователя, и эту границу можно применить для каждой компоненты вектора, как было описано в предыдущем разделе. С другой стороны, наилучшие характеристики, которые могут быть достигнуты оптимальным векторным квантователем, определяются функцией скорость-искажение или, что эквивалентно, функцией искажение-скорость. Функция искажение-скорость, которая была введена в предыдущем разделе, может быть определена в контексте векторного квантования следующим образом. Предположим, мы формируем вектор $\mathbf{X}$ размерности n из n последовательных отсчётов $\{x_k\}$. Вектор $\mathbf{X}$ квантуется в форму $\widetilde{\mathbf{X}} = Q(\mathbf{X})$, где $\widetilde{\mathbf{X}}$ – вектор, образованный рядом $\{\widetilde{\mathbf{X}}_m,\ 1 < m \le L\}$. Как было описано выше, среднее искажение D, получаемое при представлении $\mathbf{X}$ через $\widetilde{\mathbf{X}}$, равно $E\left[d(\mathbf{X},\widetilde{\mathbf{X}})\right]$, где $d(\mathbf{X},\widetilde{\mathbf{X}})$ – это искажение на одно измерение. Например,

$$d(\mathbf{X},\widetilde{\mathbf{X}}) = \frac{1}{n}\sum_{k=1}^{n} (x_k - \widetilde{x}_k)^2 .$$

Минимально достижимая средняя битовая скорость, с которой могут быть переданы векторы $\{\widetilde{\mathbf{X}}_m,\ 1 < m \le L\}$, равна

$$R = \frac{H(\widetilde{\mathbf{X}})}{n} \text{ бит/отсчет}, \qquad (3.4.39)$$

где $H(\widetilde{\mathbf{X}})$ – энтропия квантованного выхода источника, определяемая как

$$H(\widetilde{\mathbf{X}}) = -\sum_{k=1}^{L} p(\widetilde{\mathbf{X}}_l)\log_2 P(\widetilde{\mathbf{X}}_l) . \qquad (3.4.40)$$

Для данной средней скорости R минимально достижимое искажение

$$D_n(R) = \min_{Q(\mathbf{X})} E[d(\mathbf{X},\widetilde{\mathbf{X}})], \qquad (3.4.41)$$

где $R \ge H(\widetilde{\mathbf{X}})/n$ и минимум в (3.4.41) берётся по всем возможным отображениям $Q(\mathbf{X})$. В пределе, когда размерность n стремится к бесконечности, получаем

$$D(R) = \lim_{n\to\infty} D_n(R), \qquad (3.4.42)$$

где $D(R)$ – это функция искажение-скорость, которая была введена в предыдущем разделе. Из этого изложения очевидно, что функция искажение-скорость может быть как угодно приближена к пределу путём увеличения размерности n векторов.

Изложенный выше подход приемлем в предположении, что СФПВ $p(\mathbf{X})$ вектора данных известна. Однако на практике СФПВ $p(\mathbf{X})$ данных может быть неизвестна. В этом случае, возможно адаптивно выбрать квантованные выходные векторы с использованием ряда обучающих векторов $\mathbf{X}(m)$. Конкретнее, предположим, что мы имеем ряд из M векторов, причём M намного больше, чем L (M>>L). Итеративный групповой алгоритм, названный *алгоритмом K средних*, где в нашем случае K=L, может быть применён к обучающим векторам. Этот алгоритм итеративно делит M обучающих векторов на L групп так, что два необходимых условия оптимальности выполняются. Алгоритм K средних может быть описан так, как дано ниже [Макхоул и др. (1985)].

Алгоритм *K* средних

Шаг 1. Инициализируется начальный номер итерации i=0. Выбирается ряд выходных векторов $\tilde{\mathbf{X}}_k(0)$, $1 \le x \le L$.

Шаг 2. Обучающие векторы $\{\mathbf{X}(m), 1 < m \le M\}$ классифицируются в группы $\{C_k\}$ посредством правила ближайшего соседа:

$$\mathbf{X} \in C_k(i) \text{ если } D(\mathbf{X}, \tilde{\mathbf{X}}_k(i)) \le D(\mathbf{X}, \tilde{\mathbf{X}}_j(i)) \text{ для всех } k \ne j.$$

Шаг 3. Пересчитываются (для (i+1)-го шага) выходные векторы каждой группы путём вычисления центроида

$$\tilde{\mathbf{X}}_k(i) = \frac{1}{M_k} \sum_{\mathbf{X} \in C_k} \mathbf{X}(m), \quad 1 \le k \le L,$$

для обучающих векторов, которые попадают в каждую группу.

Кроме того, рассчитывается результирующее искажение $D(i)$ на i-й итерации.

Шаг 4. Заканчивается тестирование, если $D(i-1)$–$D(i)$ относительно мало. В противном случае следует идти к шагу 2.

Алгоритм K средних приводит к локальному минимуму (см. Андерберг, 1973; Линде и др., 1980). Начиная этот алгоритм различными рядами начальных выходных векторов $\{\mathbf{X}_k(0)\}$ и каждый раз выполняя оптимизацию, описанную алгоритмом K средних, можно найти глобальный оптимум. Однако вычислительные затраты этой поисковой процедуры могут ограничить поиск немногими инициализациями.

Если мы один раз выбрали выходные векторы $\{\tilde{\mathbf{X}}_k, 1 < k \le L\}$, каждый сигнальный вектор $\mathbf{X}(m)$ квантуется в выходной вектор, который является ближайшим к нему с точки зрения выбранной меры искажения. Если вычисление включает в себя оценку расстояния между $\mathbf{X}(m)$ и каждым из L возможных выходных векторов $\{\tilde{\mathbf{X}}_k\}$, процедура образует *полный поиск*. Если предположим, что каждое вычисление требует n умножений и сложений, то общее требуемое число вычислений для полного поиска равно

$$\mathscr{C} = nL \tag{3.4.43}$$

умножений и сложений на входной вектор.

Если мы выбрали L как степень 2, то $\log_2 L$ определяет число бит, требуемых для представления каждого вектора. Теперь, если R обозначает битовую скорость на отсчёт [на компоненту или на измерение $\mathbf{X}(m)$], имеем $nR = \log_2 L$ и, следовательно, вычислительные затраты

$$\mathscr{C} = n2^{nR}. \tag{3.4.44}$$

Заметим, что число вычислений растёт экспоненциально с параметром размерности n и битовой скорости R на измерение. Вследствие этого экспоненциального роста вычислительных затрат векторное квантование применяется в низкобитовых кодерах источника, таких как кодирование коэффициентов отражения или логарифмических отношений в линейном кодировании речи с предсказанием.

Вычислительные затраты, связанные с полным поиском, можно уменьшить при помощи изящного субоптимального алгоритма (см. Чанг и др., 1984; Гершо, 1982).

Чтобы продемонстрировать пользу векторного квантования по сравнению со скалярным квантованием, мы представим следующий пример, взятый у Макхоула и др. (1985).

Пример 3.4.1. Пусть X_1 и X_2 являются двумя случайными величинами с равномерной СФПВ:

$$p(x_1,x_2)=p(\mathbf{X})=\begin{cases}\dfrac{1}{ab} & (\mathbf{X}\in\mathbf{C}),\\ 0 & (\text{для других } \mathbf{X}),\end{cases} \tag{3.4.45}$$

где **C** – прямоугольная область, показанная на рис. 3.4.4. Заметим, что прямоугольник повёрнут на 45° относительно горизонтальной оси. На рис. 3.4.4 показаны также собственные плотности вероятности $p(x_1)$ и $p(x_2)$.

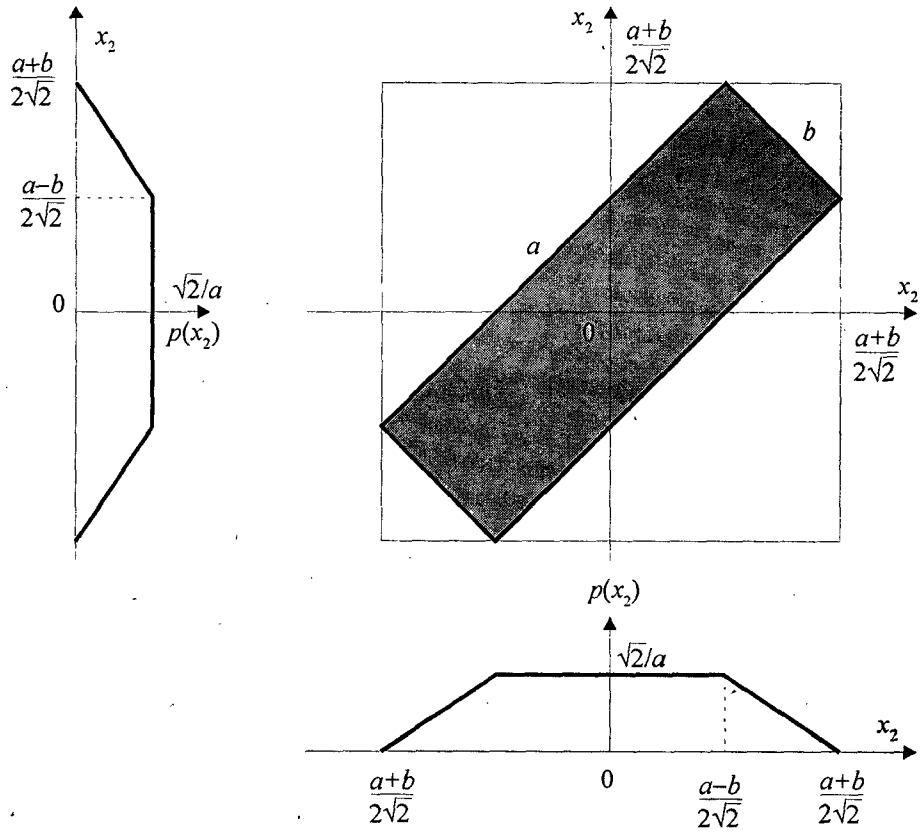


Рис. 3.4.4. Равномерная ФПВ в двух измерениях (Макхоул и др., 1985)

Если мы квантуем x_1 и x_2 раздельно, используя одинаковые интервалы квантования длины Δ, то требуемое число уровней квантования

$$L_1=L_2=\frac{a+b}{\sqrt{2}\Delta}. \tag{3.4.46}$$

Следовательно, для кодирования вектора $\mathbf{X}=[x_1\ x_2]$ потребуется число бит

$$R_x=R_1+R_2=\log_2 L_1+\log_2 L_2,$$

$$R_x=\log_2\frac{(a+b)^2}{2\Delta^2}. \tag{3.4.47}$$

Таким образом, скалярное квантование каждой компоненты эквивалентно векторному квантованию с общим числом уровней

$$L_x=L_1L_2=\frac{(a+b)^2}{2\Delta^2}. \tag{3.4.48}$$

Видим, что это приближение эквивалентно покрытию большой площади, которая

охватывает прямоугольник посредством квадратных ячеек, причём каждая ячейка представляет одну из L_x областей квантования. Поскольку $p(\mathbf{X})=0$, за исключением $\mathbf{X} \in C$, такое кодирование является расточительным и приводит к увеличению битовой скорости.

Если же мы покроем только область, где $p(\mathbf{X}) \neq 0$, квадратиками, имеющими площадь Δ^2, то общее число уровней, которые образуются, определяется площадью прямоугольника, делённой на Δ^2, т.е.

$$L_x' = \frac{ab}{\Delta^2}. \tag{3.4.49}$$

Следовательно, разница в битовой скорости при скалярном и векторном методах квантования равна

$$R_x - R_x' = \log_2 \frac{(a+b)^2}{2ab}. \tag{3.4.50}$$

Для случая, когда $a=4b$, разница в битовой скорости

$$R_x - R_x' = 1{,}64 \text{ бит/вектор.}$$

Следовательно, векторное квантование на 0,82 бит/отсчёт лучше, чем скалярное, при тех же искажениях.

Интересно заметить, что линейное преобразование (поворот на 45°) декоррелирует X_1 и X_2 и делает две случайные величины статистически независимыми. Тогда скалярное квантование и векторное квантование достигают одинаковой эффективности. Хотя линейное преобразование может декоррелировать вектор случайных величин, оно не приводит к статистически независимым случайным величинам в общем случае. Следовательно, векторное квантование будет всегда равняться или превосходить по характеристикам скалярный квантователь (см. задачу 3.40).

Векторное квантование применяется при различных методах кодирования речи, включая сигнальные методы и методы базовых моделей, которые рассматриваются в разд. 3.5. В методах, основанных на базовых моделях, таких как линейное кодирование с предсказанием, векторное квантование делает возможным кодирование речи на скоростях ниже 1000 бит/с (см. Бузо и др., 1980; Роукос и др., 1982; Пауль, 1983). Если использовать методы кодирования сигналов, возможно получить хорошее качество речи на скоростях передачи 16 000 бит/с, что эквивалентно скорости кодирования R=2 бит/отсчёт. За счёт дополнительных вычислительных усложнений в будущем станет возможным использовать сигнальные кодеры, обеспечивающие хорошее качество речи при скорости кодирования R=1 бит/отсчёт.

3.5. ТЕХНИКА КОДИРОВАНИЯ АНАЛОГОВЫХ ИСТОЧНИКОВ

За последние 40 лет было разработано много технических приёмов для кодирования аналоговых источников. Большинство из них использованы для кодирования речи и изображений. В этом разделе мы сжато опишем несколько из этих методов и используем кодирование речи как пример при оценивании их характеристик.

Удобно разделить методы кодирования аналоговых источников на три вида. Один вид назван *временное сигнальное кодирование*. При этом виде кодирования кодер источника проектируется так, чтобы представить в цифрах временные характеристики сигнала источника. Второй тип кодирования источника – *спектральное сигнальное кодирование*. В этом случае сигнал обычно подразделяется на различные частотные полоски и либо сигнал каждой полоски, либо его спектральные характеристики кодируются для передачи. Третий тип кодирования источника базируется на математической модели источника, и он называется *кодирование на базовой модели*.

3.5.1. Временное сигнальное кодирование

Имеется несколько технологических приёмов кодирования источника, которые используют временные характеристики сигнала. Наиболее широко использующийся метод описывается в этом разделе.

Импульсно-кодовая модуляция [1] **(ИКМ)**. Пусть $x(t)$ обозначает реализацию сигнала, выдаваемого источником, и пусть x_n обозначает отсчёт, взятый со скоростью стробирования $f_s \geq 2W$, где W – наивысшая частота в спектре $x(t)$.

В ИКМ каждый отсчёт сигнала квантуется в один из 2^R уровней, где R – число двоичных цифр, используемых для представления каждого отсчёта. Следовательно, скорость источника равна $R f_s$ бит/с.

Процесс квантования можно представить математически как

$$\tilde{x}_n = x_n + q_n, \qquad (3.5.1)$$

где $\tilde{x}_n$ представляет квантованное значение x_n, а q_n – ошибку квантования, которую мы трактуем как аддитивный шум. Предположим. что используется равномерное квантование, имеющее характеристику вход-выход, показанную на рис. 3.5.1, тогда шум квантования хорошо характеризуется статистически равномерной ФПВ

$$p(q) = \frac{1}{\Delta}, \qquad -\tfrac{1}{2}\Delta \leq q \leq \tfrac{1}{2}\Delta, \qquad (3.5.2)$$

где размер шага квантования $\Delta = 2^{-R}$.

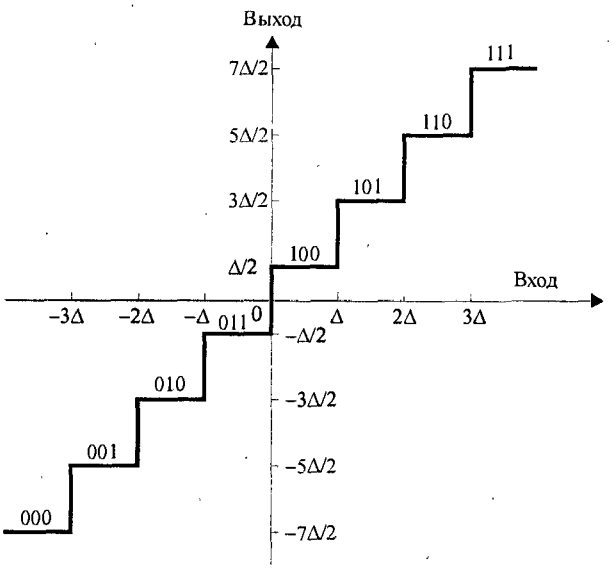


Рис. 3.5.1. Характеристика вход-выход для равномерного квантователя

Средний квадрат ошибки квантования

$$E(q^2) = \tfrac{1}{12}\Delta^2 = \tfrac{1}{12} \times 2^{-2R}. \qquad (3.5.3)$$

Средний квадрат ошибки в децибелах равен

$$10 \lg \tfrac{1}{12}\Delta^2 = 10 \lg(\tfrac{1}{12} \times 2^{-2R}) = -6R - 10{,}8 \text{ дБ}. \qquad (3.5.4)$$

Заметим, что шум квантования уменьшается на 6 дБ на каждый используемый в

[1] ИКМ, ДИКМ (дифференциальная ИКМ) и АДИКМ (адаптивная ДИКМ) относятся к технике кодирования источника. Они не являются методами цифровой модуляции.

квантователе бит. Например, 7-битовый квантователь вызывает мощность шума квантования в –52,8 дБ.

Для многих сигналов источника, таких как речевые сигналы, характерно то, что маленькие уровни сигнала появляются более часто, чем большие. Однако, равномерный квантователь обеспечивает одинаковые расстояния между последовательными уровнями во всём динамическом диапазоне сигнала. Лучший подход – это использовать неравномерный квантователь. Характеристики неравномерного квантователя обычно получают пропусканием сигнала через нелинейное устройство, которое сжимает уровни сигнала, поступающие затем на равномерный квантователь.

Например, логарифмический сжиматель (компрессор) имеет амплитудную характеристику вход-выход в виде

$$|y| = \frac{\log(1+\mu|x|)}{\log(1+\mu)}, \qquad (3.5.5)$$

где $|x| \le 1$ – амплитуда входа, $|y|$ – амплитуда выхода, μ – параметр, который выбирается так, чтобы получить требуемую характеристику компрессии. Рисунок 3.5.2 иллюстрирует характеристики компрессии для некоторых значений μ. Величина $\mu = 0$ соответствует случаю отсутствия компрессии.

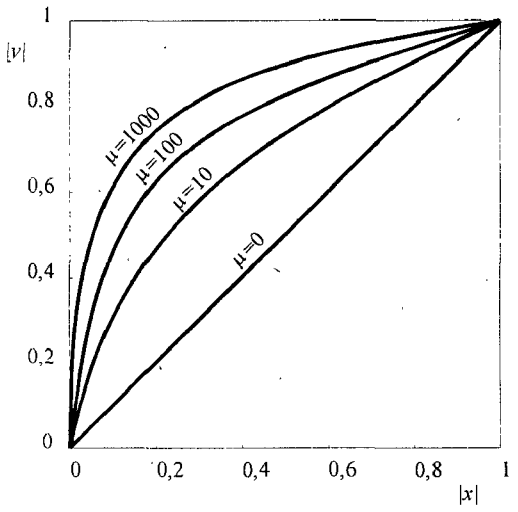


Рис.3.5.2. Амплитудная характеристика вход-выход для логарифмического компрессора

Величина $\mu = 255$ принята в качестве стандарта в США и Канаде при кодировании сигналов речи. Эта величина ведёт к уменьшению мощности шума квантования относительно равномерного квантования приблизительно на 24 дБ, как показано Джайантом (1974). Следовательно, 7-битовый квантователь, используемый совместно с логарифмическим компрессором с параметром $\mu = 255$, даёт мощность шума квантования примерно –77 дБ по сравнению с –53 дБ при равномерном квантовании.

При восстановлении сигнала по квантованным значениям используется преобразование (экспандирование), обратное логарифмическому, для декомпрессии амплитуды сигнала. Комбинированную пару компрессор–экспандер называют *компандером*.

Дифференциальная импульсно-кодовая модуляция (ДИКМ). В ИКМ каждый отсчёт кодируется независимо от других. Однако у многих источников сигнала при стробировании со скоростью Найквиста или быстрее проявляется значительная корреляция между последовательными отсчётами. Другими словами, изменения амплитуды между последовательными отсчётами в среднем относительно малы. Следовательно, схема кодирования, которая учитывает избыточность отсчётов, будет требовать более низкой битовой скорости кодирования для выхода источника.

Относительно простые решения получаются при кодировании разности между последовательными отсчётами, а не самих отсчётов. Поскольку можно ожидать, что разность между отсчётами сигнала меньше, чем действительные значения отсчётов, то потребуется меньшее число бит для представления разностного сигнала. Суть этого общего

подхода – в предсказании текущего значения отсчёта на основе предыдущих p отсчётов. Для конкретности предположим, что x_n означает текущий отсчёт источника, и пусть $\hat{x}_n$ обозначает предсказанное значение (оценку) для x_n, определяемое как

$$\hat{x}_n = \sum_{i=1}^{p} a_i x_{n-i} . \qquad (3.5.6)$$

Таким образом, $\hat{x}_n$ является взвешенной линейной комбинацией p отсчётов, а $\{a_i\}$ являются коэффициентами предсказания. Величины $\{a_i\}$ выбираются так, чтобы минимизировать некоторую функцию ошибки между x_n и $\hat{x}_n$.

Математически и практически подходящей функцией ошибок является среднеквадратическая ошибка (СКО). В этом случае мы выберем $\{a_i\}$ так, чтобы минимизировать

$$\begin{aligned} \mathscr{E}_p &= E(e_n^{\,2}) = E[(x_n - \sum_{i=1}^{p} a_i x_{n-i})^2] = \\ &= E(x_n^{\,2}) - 2\sum_{i=1}^{p} a_i E(x_n x_{n-i}) + \sum_{i=1}^{p}\sum_{j=1}^{p} a_i a_j E(x_{n-i} x_{n-j}). \end{aligned} \qquad (3.5.7)$$

В предположении, что выход источника стационарен (в широком смысле), мы можем выразить (3.5.7) в виде

$$\mathscr{E}_p = \phi(0) - 2\sum_{i=1}^{p} a_i \phi(i) + \sum_{i=1}^{p}\sum_{j=1}^{p} a_i a_j \phi(i-j), \qquad (3.5.8)$$

где $\phi(m)$ – автокорреляционная функция последовательности отсчётов сигнала x_n. Минимизация $\mathscr{E}_p$ по коэффициентам предсказания $\{a_i\}$ приводит к системе линейных уравнений

$$\sum_{i=1}^{p} a_i \phi(i-j) = \phi(j), \qquad j = 1, 2, ..., p . \qquad (3.5.9)$$

Таким образом, коэффициенты предсказания определены. Если автокорреляционная функция $\phi(j)$ априори не известна, она может быть определена по реализации отсчётов $\{x_n\}$ посредством соотношения [1]

$$\hat{\phi}(j) = \frac{1}{N}\sum_{i=1}^{N-j} x_i x_{i-j}, \qquad j = 1, 2, ..., p, \qquad (3.5.10)$$

и оценки $\hat{\phi}(j)$ используются в (3.5.9) для нахождения коэффициентов $\{a_i\}$. Заметим, что нормирующий множитель $1/N$ в (3.5.10) сокращается, когда $\hat{\phi}(n)$ подставляются в (3.5.9).

Линейные уравнения (3.5.9) для коэффициентов предсказателя называют *нормальными уравнениями* или уравнениями Юли–Волкера. Имеется алгоритм, разработанный Левинсоном (1974) и Дурбиным (1959) для эффективного решения этих уравнений. Он описывается в приложении А. Мы будем иметь дело с этими уравнениями более детально при последующем обсуждении линейного кодирования с предсказанием.

Имея метод определения коэффициентов предсказания, теперь рассмотрим блок–схему практической системы ДИКМ, показанную на рис. 3.5.3, *а*. В этой схеме предсказатель встроен в цепь обратной связи, охватывающей квантователь. Вход предсказателя обозначен $\tilde{x}_n$. Он представляет сигнальный отсчёт x_n, видоизменённый процессом

[1] Оценка автокорреляционной функции по конечному числу наблюдений заслуживает отдельного исследования, выходящего за пределы нашего обсуждения. Оценка (3.5.10) часто используется на практике.

квантования, а выход предсказателя равен

$$\hat{\tilde{x}}_n = \sum_{i=1}^{p} a_i \tilde{x}_{n-i} \,. \qquad (3.5.11)$$

• Разность

$$e_n = x_n - \hat{\tilde{x}}_n \qquad (3.5.12)$$

является входом квантователя, а $\tilde{e}_n$ обозначает его выход. Величина квантованной ошибки предсказания $\tilde{e}_n$ кодируется последовательностью двоичных символов и передаётся через канал в пункт приёма. Квантованная ошибка $\tilde{e}_n$ также суммируется с предсказанной величиной $\hat{\tilde{x}}_n$, чтобы получить $\tilde{x}_n$.

В месте приёма создаётся такой же предсказатель, как на передаче, а его выход $\hat{\tilde{x}}_n$ суммируется с $\tilde{e}_n$, чтобы получить $\tilde{x}_n$. Сигнал $\tilde{x}_n$ является входным воздействием для предсказателя и в то же время образует выходную последовательность, по которой с помощью фильтра НЧ, как показано на рис. 3.5.3, *b*, восстанавливается сигнал $\tilde{x}(t)$. Использование обратной связи вокруг квантователя обеспечивает то, что ошибка в $\tilde{x}_n$ – это просто ошибка квантования $q_n = \tilde{e}_n - e_n$ и что здесь нет накопления предыдущих ошибок квантования при декодировании. Имеем

$$q_n = \tilde{e}_n - e_n = \tilde{e}_n - \left(x_n - \hat{\tilde{x}}_n\right) = \tilde{x}_n - x_n \,. \qquad (3.5.13)$$

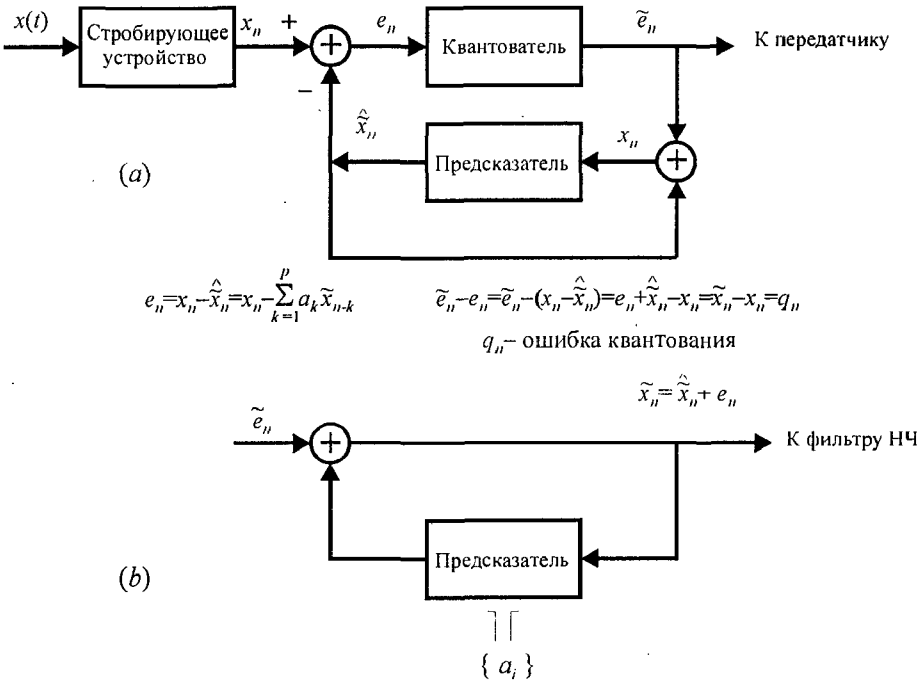


Рис. 3.5.3. (*a*) Блок-схема кодера ДИКМ; (*b*) Декодер ДИКМ в приёмнике

Следовательно, $\tilde{x}_n = x_n + q_n$. Это означает, что квантованный отсчёт $\tilde{x}_n$ отличается от входа x_n ошибкой квантования q_n независимо от использования предсказателя. Значит, ошибки квантования не накапливаются.

В системе ДИКМ, иллюстрированной рис. 3.5.3, оценка или предсказанная величина $\hat{\tilde{x}}_n$ отсчёта сигнала x_n получается посредством линейной комбинации предыдущих значений $\tilde{x}_{n-k}$, $k = 1, 2, \ldots, p$, как показано в (3.5.11). Улучшение качества оценки можно получить включением в оценку линейно отфильтрованных последних значений квантованной ошибки.

Конкретно, оценку $\hat{\tilde{x}}_n$ можно выразить так:

$$\hat{\tilde{x}}_n = \sum_{i=1}^{p} a_i \tilde{x}_{n-i} + \sum_{i=1}^{m} b_i \tilde{e}_{n-i}, \qquad (3.5.14)$$

где $\{b_i\}$ – коэффициенты фильтра для квантованной последовательности ошибок $\tilde{e}_n$. Блок-схемы кодера на передаче и декодера на приёме даны на рис. 3.5.4. Два ряда коэффициентов $\{a_i\}$ и $\{b_i\}$ выбираются так, чтобы минимизировать некоторую функцию ошибки $e_n = x_n - \hat{\tilde{x}}_n$, например среднеквадратическую ошибку.

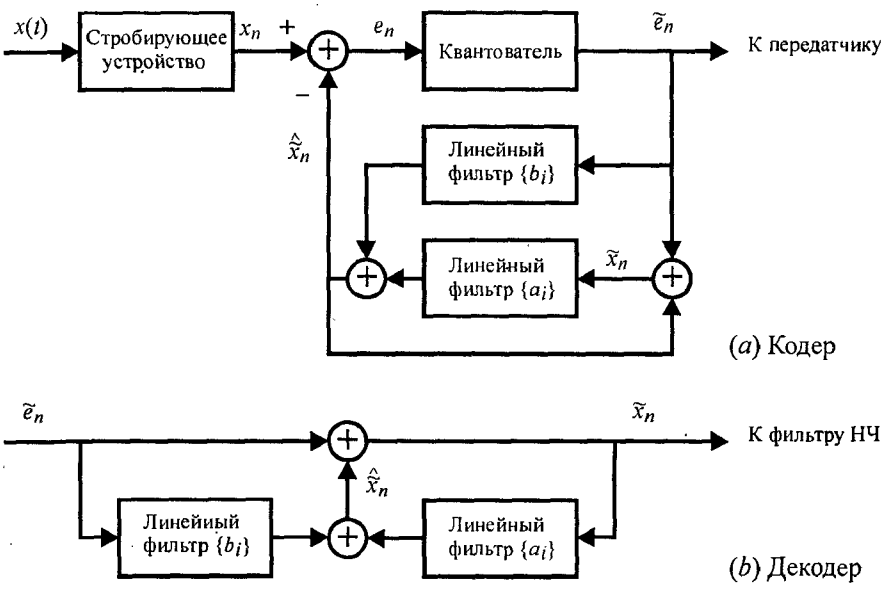


Рис. 3.5.4. Модифицированная ДИКМ посредством прибавления линейно отфильтрованной последовательности ошибок

Адаптивные ИКМ и ДИКМ. Многие реальные источники являются квазистационарными по своей природе. Одно из свойств квазистационарности характеристик случайного выхода источника заключается в том, что его дисперсия и автокорреляционная функция медленно меняются со временем. Кодеры ИКМ и ДИКМ, однако, проектируются в предположении, что выход источника стационарен. Эффективность и рабочие характеристики таких кодеров могут быть улучшены, если они будут адаптироваться к медленно меняющейся во времени статистике источника. Как в ИКМ, так и в ДИКМ ошибка квантования q_n, возникающая в равномерном квантователе, работающем с квазистационарным входным сигналом, будет иметь меняющуюся во времени дисперсию (мощность шума квантования). Одно улучшение, которое уменьшает динамический диапазон шума квантования, – это использование адаптивного квантователя. Хотя квантователь можно сделать адаптивным различными путями, относительно простой метод сводится к использованию равномерного квантователя, который меняет величину шага квантования в соответствии с дисперсией последних сигнальных отсчётов. Например, краткосрочная текущая оценка дисперсии x_n может быть рассчитана для входной последовательности $\{x_n\}$, и на основе такой оценки может быть установлен размер шага. В своём простейшем виде алгоритм для установки размера шага использует только предыдущий отсчёт сигнала. Такой алгоритм был успешно использован Джайантом (1974)

при кодировании сигналов речи. Рисунок 3.5.5 иллюстрирует такой (3-битовый) квантователь, в котором размер шага устанавливается рекуррентно согласно соотношению

$$\Delta_{n+1} = \Delta_n M(n), \qquad (3.5.15)$$

где $M(n)$ – множитель, величина которого зависит от уровня квантования отсчёта x_n, а Δ_n – размер шага квантования для обработки x_n. Величины множителей, оптимизированные для кодирования речи, были даны Джайантом (1974). Эти значения даны в табл. 3.5.1 для 2-, 3- и 4-битового адаптивного квантователя.

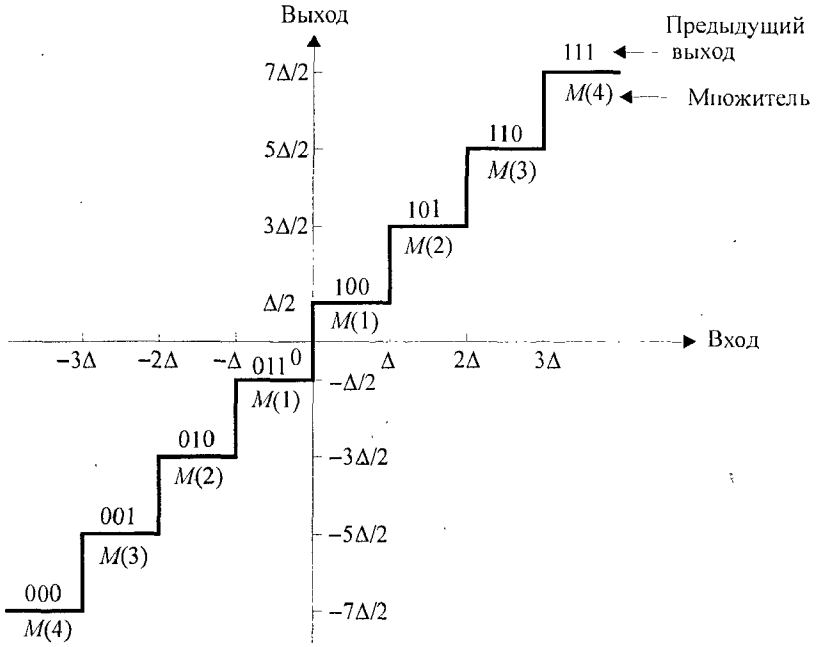


Рис. 3.5.5. Пример квантователя с адаптивным размером шага (Джайант, 1974)

Таблица 3.5.1. Коэффициенты умножения для адаптивной установки размера шага (Джайант, 1974).

	ИКМ			ДИКМ		
	2	3	4	2	3	4
$M(1)$	0,60	0,85	0,80	0,80	0,90	0,90
$M(2)$	2,20	1,00	0,80	1,60	0,90	0,90
$M(3)$		1,00	0,80		1,25	0,90
$M(4)$		1,50	0,80		1,70	0,90
$M(5)$			1,20			1,20
$M(6)$			1,60			1,60
$M(7)$			2,00			2,00
$M(8)$			2,40			2,40

Если выход источника квазистационарный, предсказатель в ДИКМ также можно сделать адаптивным. Коэффициенты предсказателя могут время от времени меняться, чтобы отразить меняющуюся статистику сигнала источника. Линейные уравнения (3.5.9) остаются справедливыми и с краткосрочной оценкой автокорреляционной функцией x_n, поставленной вместо оценки функции корреляции по ансамблю. Определённые таким

образом коэффициенты предсказателя могут быть вместе с ошибкой квантования $\tilde{e}_n$ переданы приёмнику, который использует такой же предсказатель. К сожалению, передача коэффициентов предсказателя приводит к увеличению необходимой битовой скорости, частично компенсируя снижение скорости, достигнутое посредством квантователя с немногими битами (немногими уровнями) для уменьшения динамического диапазона ошибки e_n, получаемой при адаптивном предсказании.

В качестве альтернативы предсказатель приёмника может вычислить свои собственные коэффициенты предсказания через $\tilde{e}_n$ и $\tilde{x}_n$, где

$$\tilde{x}_n = \tilde{e}_n + \sum_{i=1}^{p} a_i \tilde{x}_{n-i} . \qquad (3.5.16)$$

Если пренебречь шумом квантования, $\tilde{x}_n$ эквивалентно x_n. Следовательно, $\tilde{x}_n$ можно использовать для оценки автокорреляционной функции $\phi(n)$ в приёмнике, и результирующие оценки могут быть использованы в (3.5.9) вместо $\phi(n)$ при нахождении коэффициентов предсказателя. Для достаточно качественного квантования разность между x_n и $\tilde{x}_n$ очень мала. Следовательно, оценка $\phi(n)$, полученная через $\tilde{x}_n$, практически адекватна для определения коэффициентов предсказателя. Выполненный таким образом адаптивный предсказатель приводит к низкой скорости кодирования данных источника.

Вместо использования блоковой обработки для нахождения коэффициентов предсказателя $\{a_i\}$, как описано выше, мы можем адаптировать коэффициенты предсказателя поотсчётно, используя алгоритм градиентного типа, подобный адаптивному градиентному алгоритму выравнивания, который рассматривается в гл. 11. Похожий алгоритм градиентного типа также разработан для адаптации фильтровых коэффициентов $\{a_i\}$ и $\{b_i\}$ для системы ДИКМ, показанной на рис. 3.5.4. За подробностями такого алгоритма читатель может обратиться к книге Джайанта и Нолля (1984).

Дельта-модуляция (ДМ). Дельта-модуляцию можно рассматривать как простейшую форму ДИКМ, в которой используется двухуровневый (1-битовый) квантователь в соединении с фиксированным предсказателем первого порядка. Блок-схема кодера и декодера для ДМ показана на рис. 3.5.6, *а*. Заметим, что

$$\hat{\tilde{x}}_n = \tilde{x}_{n-i} = \hat{\tilde{x}}_{n-1} + \tilde{e}_{n-i} . \qquad (3.5.17)$$

Поскольку

$$q_n = \tilde{e}_n - e_n = \tilde{e}_n - \left(x_n - \hat{\tilde{x}}_n\right),$$

то следует

$$\hat{\tilde{x}}_n = x_{n-1} + q_{n-i} .$$

Таким образом, оцененное (предсказанное) значение x_n в действительности является предыдущим отсчётом x_{n-1}, изменённым шумом квантования q_{n-1}. Также заметим, что разностное уравнение (3.5.17) определяет интегратор со входом $\tilde{e}_n$. Следовательно, эквивалентной реализацией предсказателя первого порядка является интегратор со входом, равным квантованному сигналу ошибки $\tilde{e}_n$. В общем случае квантованный сигнал ошибки масштабируется некоторой величиной, скажем Δ_1, которая называется *размером шага.* Эквивалентная реализация показана на рис. 3.5.6, *b*. В результате кодер, показанный на рис. 3.5.6, аппроксимирует сигнал $x(t)$ посредством линейной ступенчатой функции. Для

относительно хорошей аппроксимации сигнал $x(t)$ должен меняться медленно относительно скорости стробирования. Это требование подразумевает, что скорость стробирования должна быть в несколько раз (не меньше 5) больше скорости Найквиста.

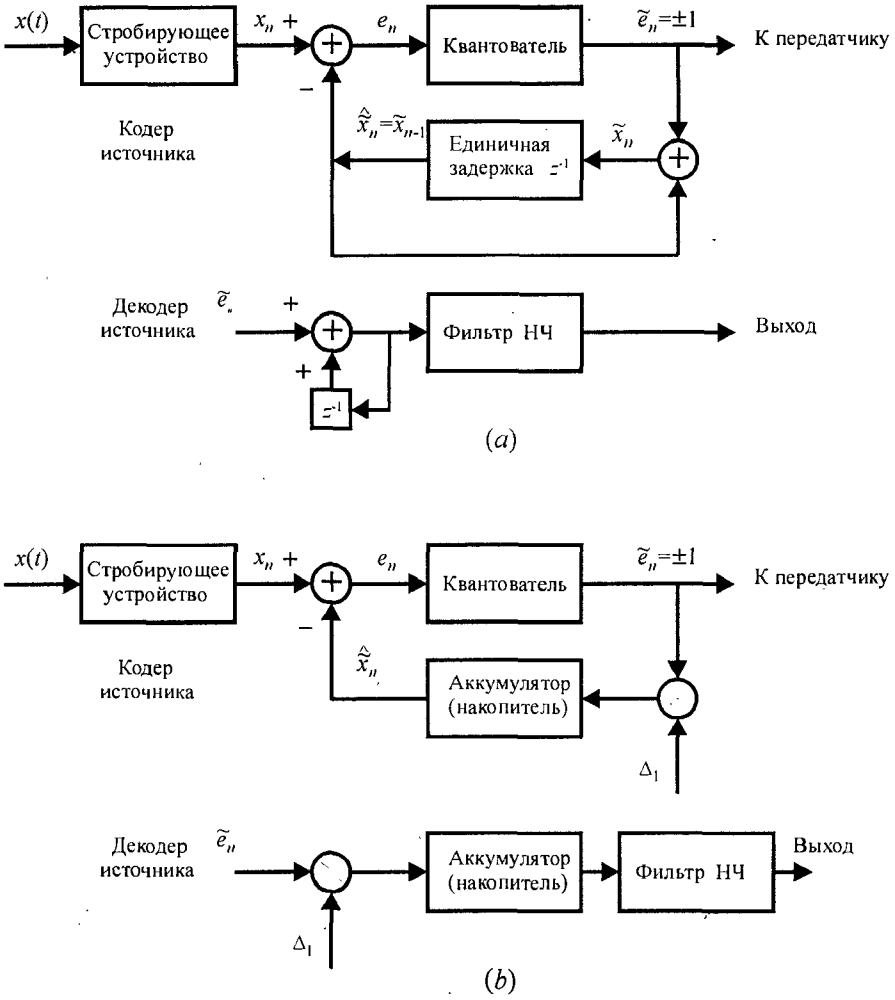


Рис. 3.5.6. (*a*) Блок-схема системы Δ-модуляции.
(*b*) Эквивалентная реализация системы Δ-модуляции

При заданной скорости стробирования характеристика качества кодера ДМ ограничена двумя видами искажений, как показано на рис. 3.5.7. Одно называется *перегрузкой по наклону*. Оно вызывается размером шага Δ_1, который слишком мал, чтобы следить за сигналом, который имеет крутой наклон. Второй тип искажений, названный *шумом дробления*, возникает от использования размера шага, который слишком велик для сигнала, имеющего малый наклон. Необходимость минимизации обоих видов искажения приводит к противоречивым требованиям для выбора размера шага Δ_1. Решение сводится к выбору Δ_1, который минимизирует сумму средних квадратов от двух видов искажений.

Даже если Δ_1 оптимизирован для минимизации общего среднего квадрата ошибки по наклону и шуму дробления, рабочая характеристика кодера с ДМ может всё ещё остаться неудовлетворительной. Альтернативное решение заключается в выборе переменного размера шага квантования, который адаптируется к кратковременным характеристикам сигнала источника, т.е. размер шага увеличивается, когда сигнал имеет крутой наклон, и уменьшается, когда сигнал имеет относительно малый наклон. Эта адаптивная характеристика иллюстрируется рис. 3.5.8.

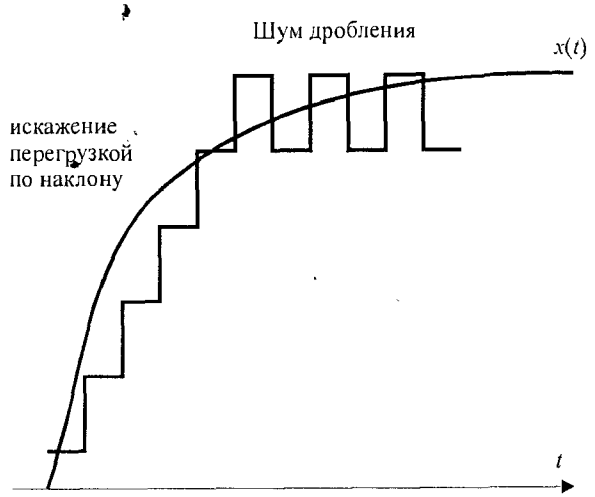


Рис. 3.5.7. Пример искажений перегрузкой по наклону и шума дробления в кодере дельта-модуляции

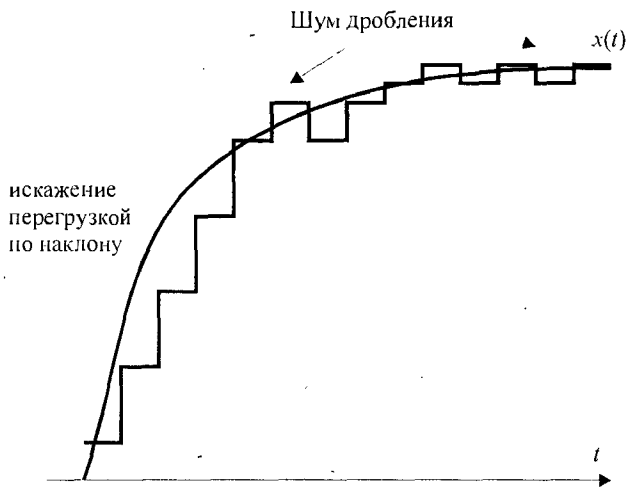


Рис. 3.5.8. Пример кодирования при дельта-модуляции с переменным размером шага

Для адаптивной установки размера шага на любой итерации могут быть использованы различные методы. Квантованная последовательность ошибок $\tilde{e}_n$ создаёт хороший критерий характеристик наклона кодируемого сигнала. Если у квантованной ошибки $\tilde{e}_n$ меняется знак между последовательными итерациями, это указывает на то, что наклон сигнала в этой области относительно мал. С другой стороны, когда сигнал имеет крутой наклон, последовательные значения ошибки $\tilde{e}_n$ имеет одинаковый знак. На основе этих наблюдений возможно создать алгоритм, который уменьшает или увеличивает размер шага в зависимости от последовательных значений $\tilde{e}_n$. Относительно простое правило, предложенное Джайантом (1970), сводится к адаптивному изменению размера шага согласно соотношению

$$\Delta_n = \Delta_{n-1} K^{\tilde{e}_n \cdot \tilde{e}_{n-1}}, \quad n = 1,2\ldots,$$

где $K \geq 1$ – константа, которая выбирается так, чтобы минимизировать суммарное искажение. Блок-схема кодера-декодера ДМ, которая включает этот адаптивный алгоритм, показана на рис. 3.5.9.

В технической литературе были исследованы и описаны насколько других вариантов адаптивного кодирования ДМ. Особенно эффективную и популярную технику, впервые предложенную Грифкесом (1970), называют *дельта-модуляцией с непрерывно меняющимся наклоном* (ДМ НМН). В ДМ НМН адаптивный параметр размера шага можно выразить так:

$$\Delta_n = \alpha \Delta_{n-1} + k_1,$$

если $\tilde{e}_n, \tilde{e}_{n-1}$ и $\tilde{e}_{n-2}$ имеют одинаковые знаки; в противном случае

$$\Delta_n = \alpha \Delta_{n-1} + k_2.$$

Параметры α, k_1, k_2 выбираются так, что $0 < \alpha \leq 1$ и $k_1 >> k_2 > 0$. Для более полного обсуждения этого и других вариантов адаптивный ДМ интересующемуся читателю следует обратиться к статьям Джайанта (1974) и Фланагана и др. (1979). Они содержат обширную библиографию. ИКМ, ДИКМ, адаптивные ИКМ, адаптивные ДИКМ и ДМ используют технику кодирования источника, которая пытается достоверно представить временную форму выходного сигнала источника. Следующий класс методов кодирования сигналов основывается на спектральных представлениях сигнала источника.

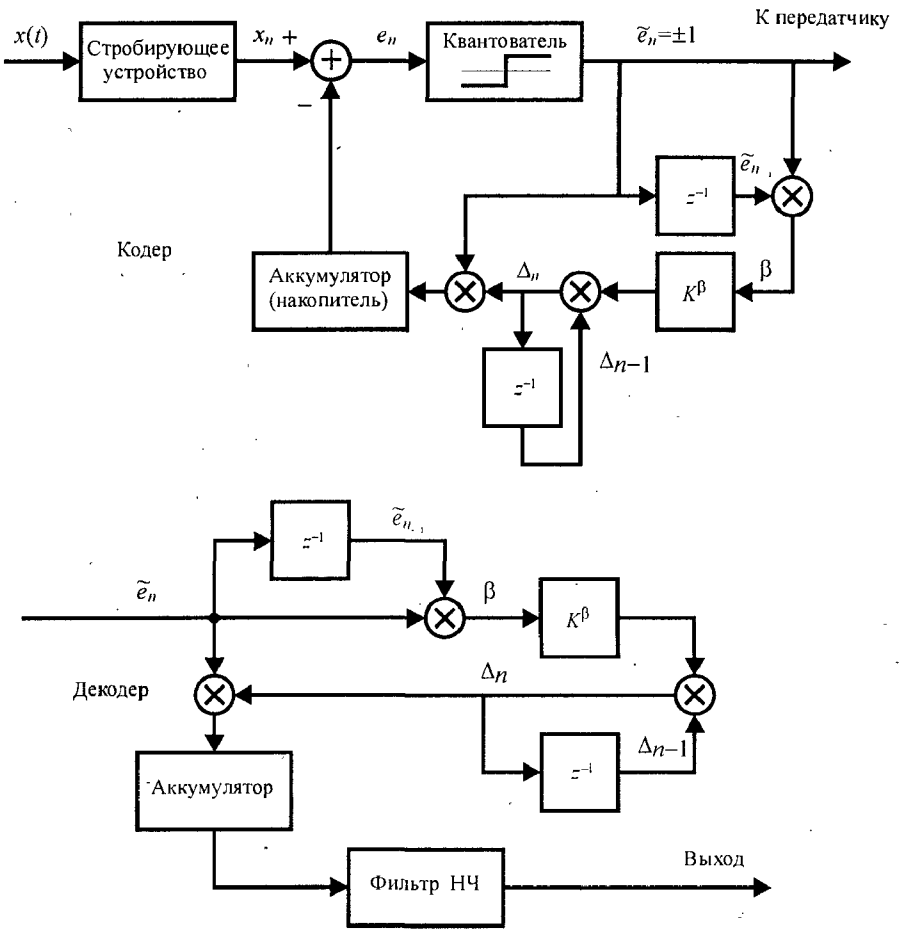


Рис. 3.5.9. Пример системы дельта-модуляции с адаптивным размером шага

3.5.2. Спектральное кодирование сигнала

В этом разделе мы кратко опишем методы кодирования сигнала, согласно которым фильтруют выход источника в определённом числе частотных полос или подполос и раздельно кодируют сигнал в каждой подполосе. Кодирование сигнала может быть выполнено во временной области в каждой подполоске или в частотной области, в которой представлен временной сигнал каждой подполоски.

Кодирование подполосок. При кодировании подполосок (КПП) сигналов речи и изображения суммарный сигнал разделяется на небольшое число частотных подполосок, а в каждой из них сигнал кодируется раздельно. При кодировании речи, например, низкочастотные полосы сигнала содержат большую часть спектральной энергии. В дополнение к этому шум квантования более заметен на слух в низкочастотной области. Следовательно, для представления сигнала в низкочастотных полосах надо использовать больше бит, а в высокочастотных – меньше.

Расчёт фильтров особенно важен для достижения хорошей рабочей характеристики КПП. На практике обычно используются квадратурно-зеркальные фильтры (КЗФ), так как они имеют наилучшую характеристику, определённую их совершенными конструктивными свойствами (см. Вайданатен, 1993). Используя КЗФ для КПП, низкочастотную полосу сигнала неоднократно делят пополам, что создаёт октавно-полосных фильтров. Выход каждого КЗФ подвергается децимации с коэффициентом 2 для уменьшения частоты стробирования. Например, предположим, что полоса частот сигнала речи простирается до

3200 Гц. Первая пара КЗФ перекрывает спектр в нижней полосе (0...1600 Гц) и верхней полосе (1600...3200 Гц). Затем нижняя полоса снова расщепляется на нижнюю (0...800 Гц) и верхнюю (800...1600 Гц) полосы путём использования другой пары КЗФ. Третье деление другой парой КЗФ может расщепить полосу 0...800 Гц на низкую (0...400 Гц) и высокую (400...800 Гц). Таким образом, тремя парами КЗФ мы получаем сигналы в частотных полосах 0...400, 400...800, 800...1600 и 1600...3200 Гц. Временной сигнал в каждой полосе может теперь кодироваться с различной точностью. На практике для кодирования сигнала в каждой подполоске используется адаптивная ИКМ.

Адаптивное преобразующее кодирование. При адаптивном преобразующем кодировании (АПК) сигнал источника стробируется и делится на группы из N_f отсчётов. Данные каждой группы преобразуются в спектральную область для кодирования и передачи. В декодере источника каждая группа спектральных отсчётов преобразуется обратно во временную область и пропускается через цифро-аналоговый преобразователь. Для достижения эффективного кодирования предусматривают больше бит для более важных спектральных коэффициентов и меньше бит для менее важных спектральных коэффициентов. Дополнительно при проектировании адаптивного распределения общего числа битов для спектральных коэффициентов мы можем адаптироваться к возможной меняющейся статистике сигнала источника. Целью выбора преобразования из временной области в частотную область является получение некоррелированных спектральных отсчётов. В этом смысле преобразование Карунена–Лоэва (ПКЛ) является оптимальным, поскольку оно даёт некоррелированные спектральные значения. Но ПКЛ в общем случае трудно выполнить (см. Винц, 1973). Дискретное преобразование Фурье (ДПФ) и *дискретное косинус–преобразование* ДКП являются приемлемыми альтернативами, хотя они субоптимальны. Из них ДКП даёт хорошую рабочую характеристику, сравнимую с ПКЛ, и оно обычно используется на практике (см. Кампанелла и Робинсон, 1971; Зелинский и Ноль, 1977).

При кодировании речи с использованием АПК возможно получить качественную передачу при скорости передачи около 9000 бит/с.

3.5.3. Модельное кодирование источника

В отличие от методов кодирования сигналов, описанных выше, модельное кодирование источника основано на совершенно ином подходе. В нём источник моделируется линейным порождающим фильтром, который при возбуждении подходящим входным сигналом выдаёт наблюдаемый выход реального источника. Вместо передачи отсчётов реальных сигналов к приёмнику передаются параметры порождающего фильтра вместе с подходящим возбуждающим сигналом. Если число параметров достаточно мало, методы модельного кодирования обеспечивают большое сжатие данных.

Наиболее широко используется метод кодирования источника, который называют линейным кодированием с предсказанием (ЛКП). В нём стробированная последовательность, обозначенная $\{x_n\}$, $n = 0, 1, \ldots, N-1$, предполагается выходом порождающего фильтра с дискретным временем и с передаточной функцией, имеющей только полюсы (всеполюсный фильтр):

$$H(z) = \frac{G}{1 - \sum_{i=1}^{p} a_k z^{-k}} \qquad (3.5.18)$$

Подходящими возбуждающими функциями для порождающего фильтра являются: импульс, последовательность импульсов или последовательность отсчётов белого гауссовского шума с единичной дисперсией. В любом случае предположим, что входная

последовательность обозначается υ_n, $n = 0,1,\ldots$. Тогда выходная последовательность порождающего фильтра удовлетворяет разностному уравнению

$$x_n = \sum_{k=1}^{p} a_k x_{n-k} + G\upsilon_n, \quad n = 0, 1, 2, \ldots \tag{3.5.19}$$

В общем, реально наблюдаемый выход источника x_n, $n = 0, 1, \ldots, N-1$ не удовлетворяет разностному уравнению (3.5.19), удовлетворяет этому уравнению только модель. Если вход является последовательностью отсчётов белого центрированного гауссовского шума, мы можем из (3.5.19) формировать оценку x_n при помощи взвешенной линейной комбинации

$$\hat{x}_n = \sum_{k=1}^{p} a_k x_{n-k}, \quad n > 0. \tag{3.5.20}$$

Разность между x_n и $\hat{x}_n$, а именно

$$e_n = x_n - \hat{x}_n = x_n - \sum_{k=1}^{p} a_k x_{n-k}, \tag{3.5.21}$$

определяет ошибку между наблюдаемым x_n и его оценкой $\hat{x}_n$. Фильтровые коэффициенты $\{a_k\}$ можно выбрать так, чтобы минимизировать средний квадрат этой ошибки.

Предположим, что входом $\{\upsilon_n\}$ является последовательность отсчётов белого центрированного шума с единичной дисперсией. Тогда выход фильтра x_n является случайной последовательностью и такой же является разность $e_n = x_n - \hat{x}_n$. Средний по ансамблю квадрат ошибки равен

$$\mathscr{E}_p = E(e_n^2) = E[(x_n - \sum_{k=1}^{p} a_k x_{n-k})^2] = \phi(0) - 2\sum_{k=1}^{p} a_k \phi(k) + \sum_{k=1}^{p}\sum_{m=1}^{p} a_k a_m \phi(k-m), \tag{3.5.22}$$

где $\phi(m)$ – автокорреляционная функция последовательности $\{x_n\}$, $n = 0, 1, \ldots, N-1$. Но $\mathscr{E}_p$ идентична СКО, определённой (3.5.8) для предсказателя, используемого в ДИКМ. Следовательно, минимизация $\mathscr{E}_p$ в (3.5.22) даёт систему линейных уравнений, данных ранее формулой (3.5.9). Для полного описания системной функции фильтра $H(z)$ мы должны определить помимо $\{a_k\}$ коэффициент усиления фильтра G. Из (3.5.19) и (3.5.21) имеем

$$E[(G^2\upsilon_n^2)] = G^2 E(\upsilon_n^2) = G^2 = E[(x_n - \sum_{k=1}^{p} a_k x_{n-k})^2] = \mathscr{E}_{p0}, \tag{3.5.23}$$

где $\mathscr{E}_{p0}$ – минимальная (остаточная) среднеквадратическая ошибка (СКО) предсказания, получаемая из (3.5.22) путём подстановки оптимальных предсказанных коэффициентов, которые следуют из решения (3.5.9). С помощью этой подстановки выражение для $\mathscr{E}_{p0}$ и, следовательно, для G^2 упрощается:

$$\mathscr{E}_{p0} = G^2 = \phi(0) - \sum_{k=1}^{p} a_k \phi(k). \tag{3.5.24}$$

На практике мы не знаем точно априори действительную автокорреляционную функцию выхода источника. Следовательно, вместо $\phi(m)$ мы подставим оценки $\hat{\phi}(m)$, даваемые (3.5.10), которые получены из ряда отсчётов x_n, $n = 0, 1, \ldots, N-1$, выдаваемых источником. Как указано ранее, алгоритм Левинсона–Дурбина, приведённый в приложении А, можно использовать для итеративного определения коэффициентов предсказания $\{a_k\}$,

начиная с предсказания первого порядка и выполняя итерацию до порядка предсказания p. Рекуррентные уравнения для $\{a_k\}$ можно выразить следующим образом:

$$\begin{aligned} a_{ii} &= \frac{\hat{\phi}(i) - \sum_{k=1}^{i-1} a_{i-1}\hat{\phi}(i-k)}{\hat{\mathcal{E}}_{i-1}}, \quad i = 2, 3, \ldots, p, \\ a_{ik} &= a_{i-1\,k} - a_{ii}a_{i-1\,k-1}, \qquad 1 \le k \le i-1, \\ \hat{\mathcal{E}}_i &= \left(1 - a_{ii}\right)\hat{\mathcal{E}}_{i-1}, \\ a_{11} &= \frac{\hat{\phi}(1)}{\hat{\phi}(0)}, \qquad \hat{\mathcal{E}}_0 = \hat{\phi}(0), \end{aligned} \tag{3.5.25}$$

где a_{ik}, $k = 1, 2, \ldots, i$, – коэффициенты предсказателя i-го порядка. Определяемые коэффициенты для предсказателя порядка p равны

$$a_k \equiv a_{pk}, \quad k = 1, 2, \ldots, p, \tag{3.5.26}$$

и остаточная СКО равна

$$\hat{\mathcal{E}} = G^2 = \hat{\phi}(0) - \sum_{k=1}^{p} a_k\hat{\phi}(k) = \hat{\phi}(0)\prod_{i=1}^{p}\left(1 - a_{ii}^2\right). \tag{3.5.27}$$

Заметим, что рекуррентные соотношения (3.5.25) дают нам не только коэффициенты предсказателя порядка p, но также коэффициенты предсказателя всех порядков, меньших p.

Остаточная СКО $\hat{\mathcal{E}}_i$, $i = 1, 2, \ldots, p$, формирует монотонно убывающую последовательность, т.е. $\hat{\mathcal{E}}_p \le \hat{\mathcal{E}}_{p-1} \le \ldots \hat{\mathcal{E}}_1 \le \hat{\mathcal{E}}_0$ и коэффициенты предсказания a_{ij} удовлетворяют условию

$$|a_{ii}| < 1, \quad i = 1, 2, \ldots, p. \tag{3.5.28}$$

Это условие необходимо и достаточно для того, чтобы все полюсы передаточной функции $H(z)$ находились внутри единичной окружности с центром в начале координат. Таким образом, условие (3.5.28) обеспечивает и устойчивость модели.

ЛКП успешно используется при моделировании источников речи. В этом случае коэффициенты a_{ii}, $i = 1, 2, \ldots, p$, названы *коэффициентами отражения* вследствие их соответствия коэффициентам отражения в акустической трубной модели голосового тракта (см. Рабинер и Шафер, 1978; Деллер и др., 1993).

Когда коэффициенты порождающего фильтра и усиление G оценены по выходам источника $\{x_n\}$, каждый из этих параметров кодируется последовательностью двоичных символов и передаётся к приёмнику. Декодирование источника или синтез сигналов речи могут быть выполнены в приёмнике, как показано на рис. 3.5.10. Генератор сигнала используется для создания отсчётов возбуждения $\{\upsilon_n\}$, которые масштабируются посредством G для получения необходимого входа фильтра с передаточной функцией $H(z)$, содержащей только полюсы и синтезированной по принимаемым коэффициентам отражения. Аналоговый сигнал источника может быть восстановлен и путём пропускания выхода $H(z)$ через аналоговый фильтр, который выполняет функцию интерполяции сигнала между отсчётными точками. В этой реализации синтезатора сигнала источника возбуждающая функция и параметр усиления G должны быть переданы вместе с коэффициентами отражения к приёмнику.

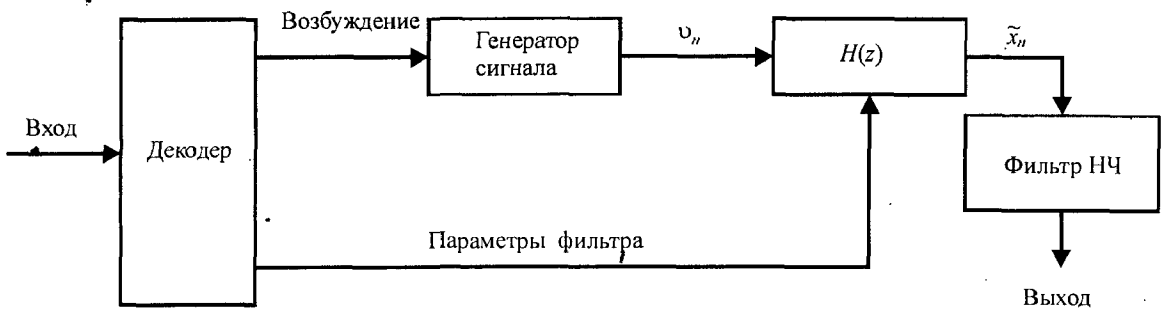


Рис. 3.5.10. Блок-схема синтезатора сигнала (декодера источника) для ЛКП системы

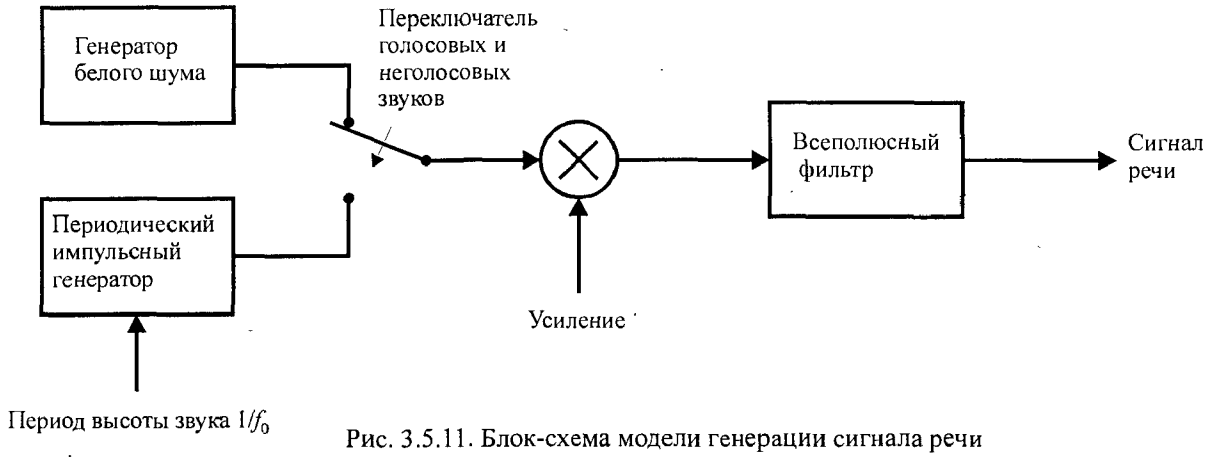


Рис. 3.5.11. Блок-схема модели генерации сигнала речи

Когда выход источника стационарен, параметры порождающего фильтра должны быть определены лишь однажды. Однако большинство источников, встречающихся на практике, в лучшем случае квазистационарны. В связи с этим обстоятельством необходимо периодически получать новые оценки для коэффициентов фильтра, для усиления G, вида возбуждающей функции и передавать эти данные к приёмнику.

Пример 3.5.1. Блок-схема рис. 3.5.11 иллюстрирует модель источника речи. Здесь имеются две взаимоисключающе возбуждающие функции для моделирования голосовых (вокализованных) и неголосовых (невокализованных) звуков речи. В пределах короткого интервала времени голосовая речь является периодической с основной частотой f_0 или с периодом повторения $1/f_0$ (основной тон), который зависит от говорящего. Таким образом, речь генерируется возбуждением модели (голосового тракта) фильтра с одними полюсами посредством периодической импульсной последовательности с периодом, равным требуемому периоду повторения. Невокализованные звуки генерируются путём возбуждения модели фильтра случайным шумом. Кодер речи в передатчике должен определить правильный тип возбуждающей функции, основной тон, параметр усиления G и коэффициенты предсказания. Эти параметры кодируются двоичными символами и передаются приёмнику. Как правило, информация о типе звука (вокализованный или невокализованный) требует для передачи 1 бит, период повторения основного тона адекватно представляется 6 битами, параметр усиления G может быть представлен 5 битами после того, как его динамический диапазон логарифмически сжат. Коэффициенты предсказания требуют 8...10 бит/коэфф. для адекватного представления (см. Рабинер и

Шафер, 1978). Смысл в такой высокой точности представления тот, что относительно малые изменения в коэффициентах предсказания ведут к большим изменениям в положении полюсов фильтра $H(z)$. Требования к точности могут быть ослаблены путём передачи коэффициентов отражения a_n, которые имеют меньший динамический диапазон. Они могут быть адекватно представлены 6 битами. Таким образом, для предсказателя порядка $p=10$ [пять полюсов в $H(z)$] общее число битов равно 72. С учётом квазистационарной природы сигналов речи линейная модель системы должна обновляться периодически, как правило, один раз каждые 15...30 мс. Поэтому битовая скорость кодера источника находится в диапазоне 4800...2400 бит/с.

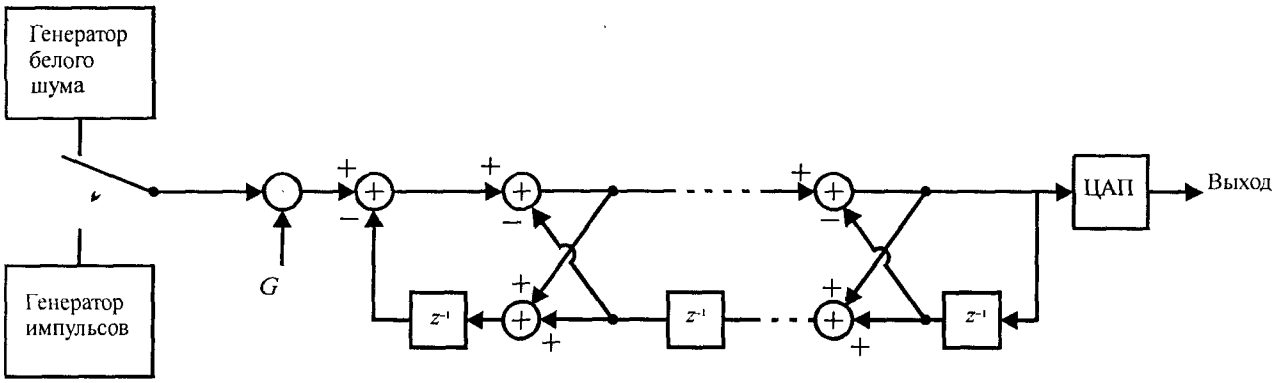


Рис. 3.5.12. Всеполюсный решётчатый фильтр для синтеза сигнала речи

Если к декодеру передаются коэффициенты отражения, нет надобности пересчитывать их в коэффициенты предсказания для того, чтобы реализовать синтезатор речи. Синтез выполняется путём реализации лестничного фильтра, показанного на рис. 3.5.12, который использует коэффициенты отражения непосредственно и который эквивалентен линейному фильтру предсказания.

Линейная модель с одними полюсами, для которой коэффициенты фильтра оцениваются посредством линейного предсказания, являются простейшей линейной моделью для источника. Более общая модель источника – линейный фильтр, который содержит и полюса, и нули. В нуль-полюсной модели выход источника x_n удовлетворяет разностному уравнению

$$x_n = \sum_{k=1}^{p} a_k x_{n-k} + \sum_{k=0}^{q} b_k \upsilon_{n-k},$$

где υ_n – входная возбуждающая последовательность. Задача теперь заключается в оценке параметров фильтра $\{a_k\}$ и $\{b_k\}$ на основе данных x_i, $i=0,1,\ldots,N-1$, выдаваемых источником. Однако критерий СКО, использованный для минимизации ошибки $e_n = x_n - \hat{x}_n$, где x_n – оценка $\hat{x}_n$, сводится теперь к решению совокупности нелинейных уравнений относительно параметров $\{a_k\}$ и $\{b_k\}$, что выполняется громоздко и математически трудно. Чтобы избежать решения нелинейных уравнений, разработан ряд субоптимальных методов для нуль-полюсного моделирования. Обсуждение этой техники. однако, уведёт нас далеко от предмета.

ЛКП, описанные выше, формируют основу для более сложных модельных методов кодирования источника. Модельные методы, используемые для кодирования речи, обычно называют *вокодерами* (voice coders). В дополнение к традиционным ЛКП вокодерам, описанным выше, разработаны другие типы вокодеров: остаточно возбуждённые ЛКП (ЛКОВ), многоимпульсные ЛКП вокодеры, кодовозбуждаемый ЛКП (ЛККВ) вокодер, и

векторно-суммарно-возбуждаемый ЛКП (ЛКВСВ) вокодер. ЛККВ и ЛКВСВ вокодеры используют векторно-квантованные возбуждающие кодовые слова для достижения высокого качества передачи речи при низкой битовой скорости кодирования.

Перед тем как закончить этот раздел, рассмотрим использование кодирования сигналов и ЛКП для кодирования сигналов речи и сравним битовые скорости этих методов кодирования.

Методы кодирования, применяемые для речевых сигналов. Передача речевых сигналов по телефонным линиям, радиоканалам и спутниковым каналам составляет наибольшую часть наших ежедневных связей. Поэтому понятно, что за последние три десятилетия большинство исследований было направлено на кодирование речи, а не на другие типы сигналов, передающих информацию. Действительно, вся техника кодирования, описанная в этом разделе, была использована для кодирования сигналов речи. Поэтому полезно сравнить эффективность этих методов через битовую скорость, требуемую для передачи сигнала речи. Предполагается, что сигнал речи ограничен полосой частот 200...3200 Гц и стробируется с номинальной скоростью 8000 отсч./с для всех кодеров, исключая ДМ, для которой скорость стробирования f_s равна битовой скорости. Предполагается, что в ЛКП кодере используются параметры, данные в примере 3.5.1.

Таблица 3.5.2 суммирует средние характеристики методов кодирования, описанные в этом разделе, и требуемые битовые скорости.

Таблица 3.5.2. Техника кодирования, применяемая для сигналов речи

Метод кодирования	Квантователь	Кодер	Скорость передачи, бит/с
ИКМ	Линейный	12 бит	96 000
ЛогИКМ	Логарифмический	7...8 бит	56 000...64 000
ДИКМ	Логарифмический	4...6 бит	32 000...48 000
АДИКМ	Адаптивный	3...4 бит	24 000...32 000
ДМ	Двоичный	1 бит	32 000...64 000
АДМ	Адаптивный двоичный	1 бит	16 000...32 000
ЛКП			2400...4800

С учётом качества синтеза сигнала речи в приёмнике посредством двоичных последовательностей, переданных по каналу без ошибок, все методы кодирования речи (ИКМ, ДИКМ, АДИКМ, ДМ, АДМ) обеспечивают по телефону качественную речь. Другими словами, слушателю будет затруднительно заметить разницу между цифровой речью и аналоговой формой речи. АДИКМ и АДМ являются особенно эффективной техникой кодирования. Статистическое кодирование позволяет снизить скорость передачи до 9600 бит/с с заметным искажением. Действительно, на скоростях ниже 16 000 бит/с искажения, обусловленные кодерами сигнала, возрастают существенно. Следовательно, эта техника не используется на скоростях ниже 9600 бит/с. Для скоростей ниже 9600 бит/с обычно используется техника кодирования типа ЛКП, которая базируется на линейных моделях источника. Синтезированная речь, полученная посредством техники кодирования этого класса, понятна. Однако сигнал речи имеет синтетическое качество, и искажения заметны.

3.6. БИБЛИОГРАФИЧЕСКИЕ ЗАМЕЧАНИЯ И ССЫЛКИ

Кодирование источника является областью интенсивной исследовательской деятельности, начиная с публикаций классических статей Шеннона в 1948 г. и статьи Хаффмена (1952). С годами были получены важные достижения в разработке высокоэффективных алгоритмов сжатия данных источника. В частности, значительными являются научные исследования универсальных кодеров источника и универсальных квантователей, опубликованные Зивом (1985), Зивом и Лемпелом (1977, 1978), Дависсоном (1973), Греем (1975), Дависоном и др. (1981).

Разработки по теории функций скорость-искажение имеются в книгах Галлагера (1968), Бергера (1971), Витерби и Омура (1979), Блейхута (1987) и Грея (1990).

Много работ было выполнено за несколько последних десятилетий по методам кодирования речи. Мы дали здесь обзор этих важных тем. Более исчерпывающая разработка дана в книгах Рабинера и Шафера (1978), Джайанта и Ноля (1984), Деллера и др.(1993). В дополнение к этим публикациям имеются специальные исследования в журнале IEEE Transactions on Communications (апрель 1979 и апрель 1982) и более новые в IEEE Journal on Selected Ereas in Communication (февраль 1988), посвящённые кодированию речи. Мы хотим также упомянуть публикацию в IEEE Press книги, содержащей репринты опубликованных статей по кодированию и квантованию сигналов, отредактированные Джайантом (1976).

В последнем десятилетии мы также увидели ряд важных достижений в области векторного квантования. Наша разработка этой темы основывалась на доходчивой работе Макхоула и др.(1985). Всесторонняя разработка по векторному квантованию и сжатию сигналов имеется в книге Гершо и Грея (1992).

ЗАДАЧИ

3.1. Рассмотрим совместный эксперимент из задачи 2.1 с заданной совместной вероятностью $p(A,B)$. Допустим, мы наблюдаем выходы A_i, $i = 1, 2, 3, 4$, эксперимент A.

a. Определите взаимную информацию $I(B_j; A_i)$ для $j = 1, 2, 3$ и $i = 1, 2, 3, 4$ в битах.

b. Определите среднюю взаимную информацию $I(B; A)$.

3.2. Предположим, что выходы B_j, $j = 1, 2, 3$, в задаче 3.1 представляют три возможных выходных слова ДИБП. Определите энтропию источника.

3.3. Докажите, что $\ln n \le n-1$ и продемонстрируйте законность этого неравенства, построив кривые $\ln n$ и $n-1$.

3.4. X и Y являются двумя дискретными случайными величинами с вероятностями $P(X = x, Y = y) = P(x, y)$. Покажите, что $I(X,Y) \ge 0$, причём равенство имеет место тогда, и только тогда, когда X и Y статистически независимы.

[Подсказка: используйте неравенство $\ln u \le u-1$ для $0 < u < 1$, чтобы доказать, что $-I(X,Y) \le 0$.]

3.5. Выход ДИБП состоит из возможных символов $x_1, x_2, \ldots, x_n$, которые появляются с вероятностями $p_1, p_2, \ldots p_n$ соответственно. Докажите, что энтропия $H(X)$ источника не превышает $\log n$.

3.6. Определите дифференциальную энтропию $h(X)$ равномерно распределённой случайной величины X с ФПВ $p(x) = \begin{cases} a^{-1} & (0 < x \le a), \\ 0 & (\text{вне этого интервала}) \end{cases}$

для следующих трёх случаев:

a) a=1;

b) a=4;

c) a=1/4.

Обратите внимание, что из расчётов следует, что $h(X)$ является не абсолютной, а только относительной мерой неопределённости.

3.7. ДИБП имеет алфавит из восьми символов x_i, $i = 1, 2, \ldots, 8$, с вероятностями 0,25; 0,2; 0,15; 0,12; 0,10; 0,08; 0,05 и 0,05.

a) Используйте процедуру кодирования Хаффмена, чтобы определить двоичный код для выхода источника.

b) Определите среднее число $\overline{R}$ двоичных символов на символ источника.

c) Определите энтропию источника и сравните с $\overline{R}$.

3.8. ДИБП источника имеет алфавит из пяти символов y_i, $i = 1, 2, \ldots, 5$, каждый из которых появляется с вероятностью 1/5. Вычислите эффективность равномерного двоичного кода, если:

a) Каждый символ кодируется отдельно в двоичную последовательность.

b) Два символа вместе кодируются в двоичную последовательность.

c) Три символа вместе кодируются в двоичную последовательность.

3.9. Напомним (3.2.6)

$$I(x_i; y_j) = I(x_i) - I(x_i|y_j).$$

Докажите, что

a) $I(x_i; y_j) = I(y_j) - I(y_j|x_i)$;

b) $I(x_i; y_j) = I(x_i) + I(y_j) - I(x_i, y_j)$, где $I(x_i; y_j) = -\log P(x_i, y_j)$.

3.10. Пусть X – геометрически распределённая случайная величина, т.е. $p(X = k) = p(1-p)^{k-1}$, $k = 1, 2, 3\ldots$

a) Найдите энтропию X.

b) Известно, что $X>K$, где K – заданное целое положительное число. Чему равна энтропия X?

3.11. Пусть X и Y обозначают две совместно распределённые дискретные случайные величины.

a) Покажите, что

$$H(X) = -\sum_{x,y} p(x,y) \log P(x),$$

$$H(Y) = -\sum_{x,y} p(x,y) \log P(y).$$

b) Используйте полученный выше результат, чтобы показать, что $H(X,Y) \le H(X) + H(Y)$. Когда наступает равенство?

c) Покажите, что $H(X|Y) \le H(X)$ и что равенство имеет место тогда, и только тогда, когда X и Y независимы.

3.12. Две двоичные случайные величины X и Y распределены согласно совместным вероятностям $P(X = Y = 0) = P(X = 0, Y = 1) = P(X = Y = 1) = 1/3$. Вычислите $H(X)$, $H(Y)$, $H(X|Y)$, $H(Y|X)$ и $H(X,Y)$.

3.13. Дан марковский процесс с одношаговой памятью, т.е. такой процесс, что $p(x_n|x_{n-1}, x_{n-2}, x_{n-3}, \ldots) = p(x_n|x_{n-1})$ для всех n. Покажите, что для стационарного марковского процесса энтропийная скорость определяется через $H(X_n|X_{n-1})$.

3.14. Пусть $Y = g(X)$, где g обозначает детерминированную функцию. Покажите, что в общем $H(Y) \le H(X)$. Когда наступает равенство?

3.15. Покажите, что $I(X;Y) = I(X) + I(Y) - I(XY)$.

3.16. Покажите, что для статистически независимых событий

$$H(X_1, X_2, \ldots, X_n) = \sum_{i=1}^{n} H(X_i).$$

3.17. Покажите, что в канале без шумов $H(X|Y) = 0$.

3.18. Покажите, что $I(X_3; X_2|X_1) = H(X_3|X_1) - H(X_3|X_1X_2)$ и что $H(X_3|X_1) \ge H(X_3|X_1X_2)$.

3.19. Пусть X является случайной величиной с ФПВ $p_x(x)$ и пусть $Y = aX + b$ – линейное преобразование X, где a и b – две константы. Определите дифференциальную энтропию $h(Y)$ через $h(X)$.

3.20. Выходы x_1, x_2 и x_3 от ДИБП с вероятностями $p_1 = 0{,}45$, $p_2 = 0{,}35$ и $p_3 = 0{,}2$ преобразуются линейным преобразованием $Y = aX + b$, где a и b – константы. Определите энтропию $H(Y)$ и поясните влияние преобразования на энтропию сигнала.

3.21. Оптимальный четырёхуровневый неравномерный квантователь для сигнала с гауссовским распределением амплитуд выдаёт четыре уровня a_1, a_2, a_3 и a_4 с вероятностями $p_1 = p_2 = 0{,}3365$ и $p_3 = p_4 = 0{,}1635$.

a) Определите код Хаффмена, который кодирует отдельные уровни, и определите среднюю битовую скорость.

b) Определите код Хаффмена, который кодирует два выходных уровня вместе, и определите среднюю битовую скорость.

c) Какую минимальную битовую скорость можно получить, кодируя J выходных уровней, когда $J \to \infty$.

3.22. Марковский источник первого порядка характеризуется вероятностями состояния $p(x_i)$, $i = 1, 2, \ldots, L$, и переходными вероятностями $P(x_k|x_i)$, $k = 1, 2, \ldots, L$ и $k \neq i$. Энтропия марковского источника $H(X) = \sum_{k=1}^{L} P(x_k) H(X|x_k)$, где $H(X|x_k)$ – энтропия источника при условии, что он находится в состоянии x_k.

Определите энтропию двоичного источника первого порядка, показанного на рис. 3.22, который имеет переходные вероятности $P(x_2|x_1) = 0{,}2$ и $P(x_1|x_2) = 0{,}3$ [заметим, что условные энтропии $H(X|X_1)$ и $H(X|X_2)$ определяются двоичными энтропийными функциями $H[P(x_2|x_1)]$ и $H[P(x_1|x_2)]$ соответственно]. Как соотносится энтропия марковского источника с энтропией двоичного ДИБП с теми же вероятностями выходных символов $P(x_1)$ и $P(x_2)$?

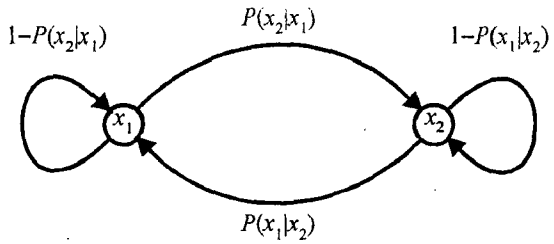


Рис. Р.3.22

3.23. Источник без памяти имеет алфавит $A = \{-5, -3, -1, 0, 1, 3, 5\}$ с соответствующими вероятностями $\{0{,}05;\ 0{,}1;\ 0{,}1;\ 0{,}15;\ 0{,}05;\ 0{,}25;\ 0{,}3\}$.

a) Найдите энтропию источника.

b) Предположив, что источник квантуется согласно правилу квантования

$$q(-5) = q(-3) = 4,$$
$$q(-1) = q(0) = q(1) = 0,$$
$$q(3) = q(5) = 4,$$

найдите энтропию квантованного источника.

3.24. Постройте троичный код Хаффмена, использующий выходные символы 0, 1 и 2 при кодировании источника с вероятностями выходных символов алфавита $\{0{,}05;\ 0{,}1;\ 0{,}15;\ 0{,}17;\ 0{,}18;\ 0{,}22;\ 0{,}13\}$. Какова результирующая средняя длина кодового слова? Сравните среднюю длину кодового слова с энтропией источника. (С каким основанием будете вычислять логарифмы в выражении для энтропии для полностью осмысленного сравнения?).

3.25. Найдите код Лемпела-Зива при кодировании двоичной последовательности источника 000100100000011000010000000100000010100001000000110100000001100.

Восстановите исходную последовательность по коду Лемпела-Зива. [Подсказка: Вам потребуются два прохода двоичной последовательности, чтобы принять решение о размере словаря.]

3.26. Найдите дифференциальную энтропию непрерывной случайной величины X в следующих случаях:

a) X–случайная величина с экспоненциальным распределением с параметром $\lambda > 0$, т.е.

$$f_X(x) = \begin{cases} \lambda^{-1} e^{-x/\lambda} & (x > 0), \\ 0 & (\text{для других } x). \end{cases}$$

b) X–случайная величина с распределением Лапласа с параметром $\lambda > 0$, т.е.

$$f_X(x) = \frac{1}{2\lambda} e^{-|x|/\lambda}.$$

c) X–случайная величина с треугольным законом распределения с параметром $\lambda > 0$, т.е.

$$f_X(x) = \begin{cases} (x+\lambda)/\lambda^2 & (-\lambda \le x \le 0), \\ (-x+\lambda)/\lambda^2 & (0 < x \le \lambda), \\ 0 & (\text{для других } x). \end{cases}$$

3.27. Можно показать, что для источника с рапределением Лапласа $f_X(x) = (2\lambda)^{-1} e^{-|x|/\lambda}$ функция скорость-искажение с абсолютной величиной меры ошибки искажений $d(x, \hat{x}) = |x - \hat{x}|$ определяется как

$$R(D) = \begin{cases} \log(\lambda / D) & (0 \le D \le \lambda), \\ 0 & (D > \lambda \quad). \end{cases}$$

(См. Бергер, 1971)

a) Сколько требуется бит/отсчёт для представления выходов источника со средним искажением, не превышающим $\lambda/2$?

b) Постройте график $R(D)$ для трёх различных значений λ и обсудите влияние изменения λ на этих кривых.

3.28. Можно показать, что если X – непрерывная случайная величина с нулевым средним и дисперсией σ^2, то её функция скорость-искажение при среднеквадратичной мере искажений удовлетворяет нижней и верхней границам, определяемым неравенствами

$$h(X) - \tfrac{1}{2}\log 2\pi e D \le R(D) \le \tfrac{1}{2}\log \tfrac{1}{2}\sigma^2,$$

где $h(X)$ означает дифференциальную энтропию случайной величины X (см. Ковер и Томас, 1991)

a) Покажите, что для гауссовской случайной величины верхней и нижней границ совпадают.

b) Постройте график для нижней и верхней границ для источника с лапласовским распределением при $\sigma^2=1$.

c) Постройте график для нижней и верхней границ для источника с треугольным распределением при $\sigma^2=1$.

3.29. Стационарный случайный процесс имеет автокорреляционную функцию $R_X(\tau) = \frac{1}{2} A^2 e^{-\alpha|\tau|} \cos 2\pi f_0 \tau$ и известно, что случайный процесс никогда не превышает по амплитуде величину 6. Сколько требуется уровней квантования амплитуды, чтобы гарантировать отношение сигнал/шум квантования не хуже 60 дБ?

3.30. Канал с аддитивным белым гауссовским шумом имеет выход $Y=X+N$, где X – вход канала, а N – шум с ФПВ:

$$p_N(n) = \frac{1}{\sqrt{2\pi}\sigma_n} e^{-n^2/2\sigma_n^2}.$$

Для случая, когда X – гауссовский белый шум с параметрами $E(X) = 0$ и $E(X^2) = \sigma_X^2$, определите:

a) условную дифференциальную энтропию $h(X|N)$;

b) среднюю взаимную информацию $I(X;Y)$.

3.31. ДИБП имеет алфавит из восьми символов x_i, $i = 1, 2, \ldots, 8$ с вероятностями из задачи 3.7. Используйте процедуру кодирования Хаффмена для нахождения троичного кода (с символами 0,1 и 2) для кодирования выхода источника. [Подсказка: прибавьте символ x_9 с вероятностью $p_9=0$ и группируйте по три символа на каждом шаге.]

3.32. Определите, существует ли двоичный код с кодовыми словами длиной $(n_1, n_2, n_3, n_4) = (1, 2, 2, 3)$, удовлетворяющий условию префиксности.

3.33. Рассмотрите двоичный блоковый код с 2^n кодовыми словами одинаковой длины n. Покажите, что неравенство Крафта выполняется для такого кода.

3.34. Покажите, что энтропия n-мерного гауссовского вектора $\mathbf{X}=[x_1\ x_2\ \ldots x_n]$ с нулевым средним и матрицей ковариаций $\mathbf{M}$ равна

$$H(\mathbf{X}) = \tfrac{1}{2}\log_2(2\pi e)^n\,|\mathbf{M}|.$$

3.35. Рассмотрите ДИБП с равновероятными двоичными выходными символами (0,1). Установите меру искажений как $D=P_e$, где P_e – вероятность ошибки при передаче двоичных символов пользователю через двоичный симметричный канал (ДСК). Тогда функция скорость-искажение равна (Бергер, 1971)

$$R(D) = 1 + D\log_2 D + (1-D)\log_2(1-D), \qquad 0 \le D = P_e \le \tfrac{1}{2}.$$

Постройте график $R(D)$ для $0 \le D \le 1/2$.

3.36. Вычислите функцию скорость-искажение для M-ичного симметричного канала

$$R(D) = \log_2 M + D\log_2 D + (1-D)\log_2\frac{1-D}{M-1}$$

для $M=2$, 4, 8 и 16. $D=P_M$ – вероятность ошибки.

3.37. Рассмотрите пользу от взвешенной СКО как меры искажений, определённой как

$$d_W(\mathbf{X},\tilde{\mathbf{X}}) = (\mathbf{X}-\tilde{\mathbf{X}})^{\mathrm{T}}\mathbf{W}(\mathbf{X}-\tilde{\mathbf{X}}),$$

где $\mathbf{W}$ – симметричная, положительно-определённая взвешивающая матрица. Путём факторизации $\mathbf{W}$ как $\mathbf{W}=\mathbf{P}^{\mathrm{T}}\mathbf{P}$ покажите, что $d_W(\mathbf{X},\tilde{\mathbf{X}})$ эквивалентно невзвешенной СКО как меры искажений $d_2(\mathbf{X}^{\mathrm{T}},\tilde{\mathbf{X}}^{\mathrm{T}})$, содержащей преобразованные векторы $\mathbf{X}^{\mathrm{T}}$ и $\tilde{\mathbf{X}}^{\mathrm{T}}$.

3.38. Рассмотрите стационарную случайную сигнальную последовательность $\{X(n)\}$ с нулевым средним и автокорреляционной функцией

$$\phi(n) = \begin{cases} 1 & (n=0), \\ \frac{1}{2} & (n=\pm 1), \\ 0 & (\text{для других } n). \end{cases}$$

a) Определите коэффициенты предсказания для предсказателя первого порядка с минимизацией СКО для $\{X(n)\}$, заданной посредством соотношения

$$\hat{x}(n) = a_1 x(n-1),$$

и соответствующее значение минимальной СКО ε_1.

b) Повторите (a) для предсказателя второго порядка

$$\hat{x}(n) = a_1 x(n-1) + a_2 x(n-2).$$

3.39. Рассмотрите кодирование случайных величин X_1 и X_2, которые характеризуются СФПВ $p(x_1,x_2)$, заданной как

$$p(x_1,x_2) = \begin{cases} 15/7 \\ 0 \end{cases}$$

как показано на рис. Р.3.39. Вычислите битовую скорость, требуемую при равномерном раздельном квантовании x_1 и x_2 (скалярное квантование) и комбинированном (векторном) квантовании (x_1, x_2). Определите разницу в битовой скорости при $a=4b$

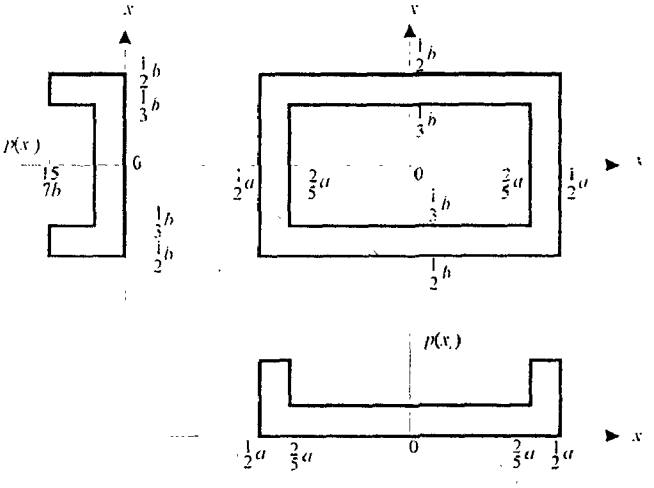


Рис. Р.3.39

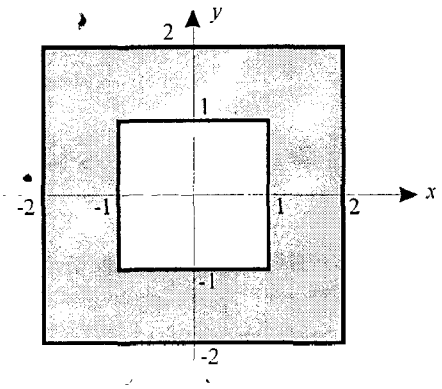


Рис.Р3.40

3.40. Рассмотрите кодирование двух случайных величин X и Y, которые имеют равномерное распределение в области между двумя квадратами, как показано на рис. Р3.40.

a) Найдите $f_X(x)$ и $f_Y(y)$.

b) Предположите, что каждая из случайных величин X и Y квантуется с использованием четырёхуровневого равномерного квантователя. Каково результирующее искажение? Каково результирующее число бит на пару (X, Y)?

c) Предположите, что вместо скалярного квантования X и Y мы используем векторный квантователь для достижения того же уровня искажений, как в (b). Каково результирующее число битов на выходную пару источника (X, Y)?

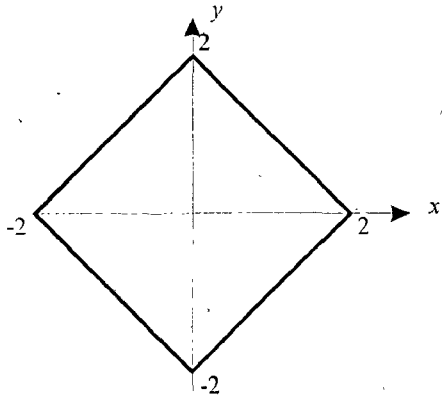


Рис. Р3.41

3.41. Две случайные величины X и Y распределены равномерно в квадрате, показанном на рис. Р3.41.

a) Найдите $f_X(x)$ и $f_Y(y)$.

b) Предположите, что каждая из случайных величин X и Y квантуется посредством четырёхуровневого равномерного квантователя. Каково результирующее искажение? Каково результирующее число бит на пару источника (X, Y)?

c) Предположите, что вместо скалярного квантования X и Y мы используем векторный квантователь с тем же числом бит на пару источника (X, Y), что в (*b*). Каково результирующее искажение для этого векторного квантователя?

4

ХАРАКТЕРИСТИКИ СИГНАЛОВ И СИСТЕМ СВЯЗИ

Сигналы можно характеризовать различными способами, как случайные или детерминированные, с дискретными либо непрерывными амплитудами низкочастотные или полосовые, с ограниченной или неограниченной энергией, с ограниченной или неограниченной мощностью и т.д. В этой главе мы рассмотрим характеристики сигналов и систем, которые обычно встречаются при передаче цифровой информации по каналам связи. В частности, мы введём представление различных форм сигналов при цифровой модуляции и опишем их спектральные характеристики.

Начнём с характеристики полосовых сигналов и систем, включая математические представления полосовых стационарных случайных процессов. Затем мы ознакомимся с векторным представлением сигналов. Завершим главу представлением сигналов цифровой модуляции и их спектральными характеристиками.

4.1. ПРЕДСТАВЛЕНИЕ ПОЛОСОВЫХ СИГНАЛОВ И СИСТЕМ

Многие сигналы, порождённые цифровыми сообщениями, передаются посредством какого-либо вида модуляции несущей. Канал, через который передается сигнал, ограничен по полосе интервалом частот, концентрируемых около частоты несущей, как при двухполосной модуляции, или в смежной от несущей полосе, как при однополосной модуляции. Сигналы и каналы, которые удовлетворяют условиям, что их полоса частот значительно меньше, чем их несущая, называют *узкополосными сигналами и каналами.* Модуляция, осуществляемая на передающей стороне системы связи для создания полосового сигнала, и демодуляция, осуществляемая на приёмной стороне, чтобы выделить цифровую информацию, предполагают преобразование частоты. Без потери общности и для математического удобства желательно представить все полосовые сигналы и каналы эквивалентными низкочастотными сигналами и каналами. Как следствие, качественные результаты различной техники модуляции и демодуляции, представленные в последующих главах, не зависят от частоты несущей и полосы частот канала. Представление полосовых сигналов и систем через эквивалентные низкочастотные формы и описания полосовых стационарных случайных процессов являются основными предметами этого раздела.

4.1.1. Представление полосовых сигналов

Предположим, что вещественный сигнал $s(t)$ имеет частоты, концентрированные в узкой полосе частот вблизи частоты f_c, как показано на рис. 4.1.1.

Наша цель – дать математическое представление таких сигналов. Сначала мы сконструируем сигнал, который содержит только положительные частоты из $S(f)$. Такой сигнал можно выразить как

$$S_+(f) = 2u(f)S(f), \qquad (4.1.1)$$

где $S_+(f)$ – преобразование Фурье $F[s(t)]$ от $s(t)$, а $u(f)$ – единичная ступенчатая функция. Эквивалентное представление (4.1.1) во временной области

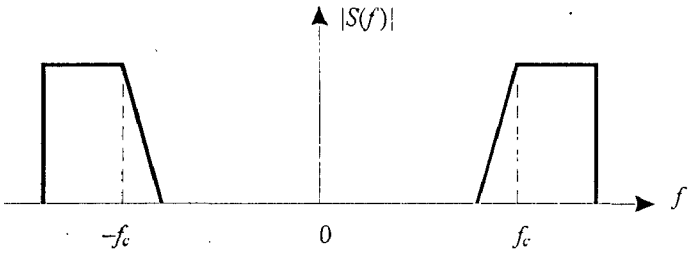


Рис. 4.1.1. Спектр полосового сигнала

$$s_+(t) = \int_{-\infty}^{\infty} S_+(f) e^{j2\pi ft} df = F^{-1}\left[2u(f)\right] * F^{-1}\left[S(f)\right]. \qquad (4.1.2)$$

Сигнал $s_+(t)$ называется *аналитическим сигналом* для $s(t)$. Заметим, что $F^{-1}\left[S(f)\right] = s(t)$ и

$$F^{-1}\left[2u(f)\right] = \delta(t) + \frac{j}{\pi t}. \qquad (4.1.3)$$

Следовательно,

$$s_+(t) = \left[\delta(t) + \frac{j}{\pi t}\right] * s(t) = s(t) + \frac{j}{\pi t} * s(t). \qquad (4.1.4)$$

Определим

$$\hat{s}(t) = \frac{1}{\pi t} * s(t) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{s(\tau)}{t-\tau} d\tau. \qquad (4.1.5)$$

Сигнал $\hat{s}(t)$ можно рассматривать как выход фильтра с импульсной характеристикой

$$h(t) = \frac{1}{\pi t}, \qquad -\infty < t < \infty, \qquad (4.1.6)$$

при подаче на вход сигнала $s(t)$. Такой фильтр называют преобразователем Гильберта. Частотная характеристика такого фильтра очень проста:

$$H(f) = \int_{-\infty}^{\infty} h(t) e^{j2\pi ft} dt = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{1}{t} e^{j2\pi ft} dt = \begin{cases} -j & (f > 0), \\ 0 & (f = 0), \\ j & (f < 0). \end{cases} \qquad (4.1.7)$$

Заметим, что $\left|H(f)\right| = 1$ при $f \neq 0$ и что фазовая характеристика

$$\Theta(f) = \begin{cases} -\pi/2 & \text{для } f > 0, \\ +\pi/2 & \text{для } f < 0. \end{cases}$$

Следовательно, этот фильтр по существу – фазовращатель на 90° для всех частот входного сигнала.

Аналитический сигнал $s_+(t)$ является полосовым сигналом. Мы можем получить эквивалентное низкочастотное представление, выполнив частотное преобразование $S_+(f)$. Определим $S_l(f)$ так:

$$S_l(f) = S_+(f + f_c). \qquad (4.1.8)$$

Эквивалентное соотношение во временной области

$$s_l(t) = s_+(t) e^{-j2\pi f_c t} = \left[s(t) + j\hat{s}(t)\right] e^{-j2\pi f_c t}, \qquad (4.1.9)$$

или, что эквивалентно,

$$s(t) + j\hat{s}(t) = s_l(t) e^{j2\pi f_c t}. \qquad (4.1.10)$$

В общем случае сигнал $S_l(f)$ комплексный (см. задачу 4.5), и его можно выразить так:

$$s_l(t) = x(t) + j\,y(t). \qquad (4.1.11)$$

Если мы подставим $s_l(t)$ в (4.1.10) и приравняем вещественные и мнимые части с каждой стороны, получим соотношения

$$s(t) = x(t)\cos 2\pi f_c t - y(t)\sin 2\pi f_c t, \qquad (4.1.12)$$

$$\hat{s}(t) = x(t)\sin 2\pi f_c t + y(t)\cos 2\pi f_c t. \qquad (4.1.13)$$

Выражение (4.1.12) – желательная форма представления полосового сигнала. Низкочастотные сигнальные компоненты $x(t)$ и $y(t)$ можно рассматривать как сигналы, модулирующие по амплитуде соответственно несущие $\cos 2\pi f_c t$ и $\sin 2\pi f_c t$. Поскольку эти несущие находятся в квадратуре (сдвинуты по фазе на 90°), $x(t)$ и $y(t)$ называют *квадратурными компонентами* полосового сигнала $s(t)$.

Другое представление для сигнала (4.1.12) такое:

$$s(t) = \mathrm{Re}\left\{\left[x(t) + j\,y(t)\right]e^{j2\pi f_c t}\right\} = \mathrm{Re}\left[s_l(t)\,e^{j2\pi f_c t}\right], \qquad (4.1.14)$$

где Re означает вещественную часть комплексной величины. Низкочастотный сигнал $s_l(t)$ обычно называют комплексной огибающей вещественного сигнала $s(t)$. Она является по существу эквивалентным низкочастотным сигналом. Наконец, третья возможная форма представления полосового сигнала получается, если представить

$$s_l(t) = a(t)\,e^{j\theta(t)}, \qquad (4.1.15)$$

где

$$a(t) = \sqrt{x^2(t) + y^2(t)}, \qquad (4.1.16)$$

$$\theta(t) = \mathrm{arc\,tg}\frac{y(t)}{x(t)}. \qquad (4.1.17)$$

Тогда

$$\begin{aligned} s_l(t) &= \mathrm{Re}\left[s_l(t)\,e^{j2\pi f_c t}\right] = \mathrm{Re}\left[a(t)\,e^{j\left[2\pi f_c t + \theta(t)\right]}\right] = \\ &= a(t)\cos\left[2\pi f_c t + \theta(t)\right]. \end{aligned} \qquad (4.1.18)$$

Сигнал $a(t)$ называют (вещественной) огибающей $s(t)$, а $\theta(t)$ называют фазой $s(t)$. Таким образом, (4.1.12), (4.1.14) и (4.1.18) являются эквивалентными представлениями полосовых сигналов. Преобразование Фурье $s(t)$

$$S(f) = \int_{-\infty}^{\infty} s(t)\,e^{-j2\pi ft}dt = \int_{-\infty}^{\infty}\left\{\mathrm{Re}\left[s_l(t)\,e^{j2\pi f_c t}\right]\right\}e^{-j2\pi ft}dt. \qquad (4.1.19)$$

Если использовать равенство

$$\mathrm{Re}(\xi) = \tfrac{1}{2}\left(\xi + \xi^*\right) \qquad (4.1.20)$$

в (4.1.19), то следует

$$\begin{aligned} S(f) &= \tfrac{1}{2}\int_{-\infty}^{\infty}\left[s_l(t)\,e^{j2\pi f_c t} + s_l^*(t)\,e^{-j2\pi f_c t}\right]e^{-j2\pi ft}dt = \\ &= \tfrac{1}{2}\left[S_l(f - f_c) + S_l^*(-f - f_c)\right], \end{aligned} \qquad (4.1.21)$$

где $S_l(f)$ – преобразование Фурье от $s_l(t)$. Это базовое соотношение между спектром действительного полосового сигнала $S(f)$ и спектром эквивалентного низкочастотного сигнала $S_l(f)$.

Энергия вещественного сигнала $s(t)$ определяется так:

$$\mathscr{E} = \int_{-\infty}^{\infty} s^2(t)\,dt = \int_{-\infty}^{\infty}\left\{\mathrm{Re}\left[s_l(t)\,e^{j2\pi f_c t}\right]\right\}^2 dt. \qquad (4.1.22)$$

Если равенство (4.1.20) использовать в (4.1.22), то следует результат

$$\mathscr{E} = \frac{1}{2}\int_{-\infty}^{\infty}|s_l(t)|^2\,dt + \frac{1}{2}\int_{-\infty}^{\infty}|s_l(t)|^2\cos\left[4\pi f_c t + 2\theta(t)\right]dt. \quad (4.1.23)$$

Рассмотрим второй интеграл в (4.1.23). Поскольку сигнал $s(t)$ узкополосный, то вещественная огибающая $a(t) = |s_l(t)|$ или, что эквивалентно, $a^2(t) = |s_l(t)|^2$ меняется медленно по сравнению с быстрыми изменениями функции косинуса. Графическая иллюстрация подынтегрального выражения во втором интеграле (4.1.21) дана на рис. 4.1.2. Величина этого интеграла равна площади под косинусной функцией, промодулированной сигналом $a^2(t)$.

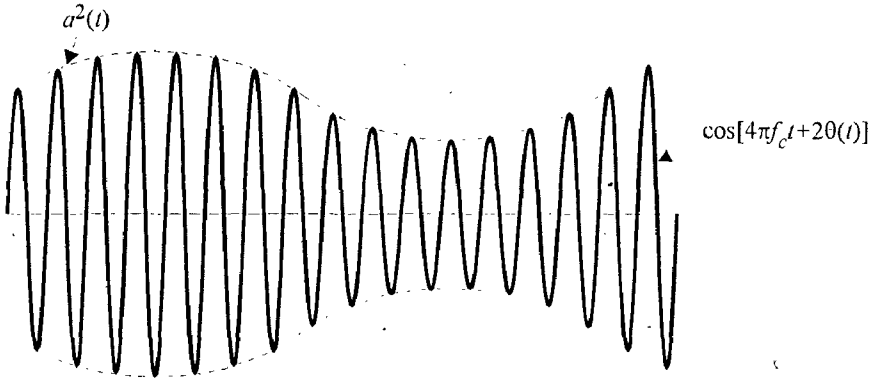


Рис. 4.1.2. Сигнал $a^2(t)\cos[4\pi f_c t+2\theta(t)]$

Поскольку модулирующий сигнал $a^2(t)$ меняется медленно по сравнению с косинусной функцией, площадь, определяемая вторым интегралом, очень мала по сравнению с величиной первого интеграла в 4.1.23, и, следовательно, вторым интегралом можно пренебречь. Таким образом, для всех практических приложений энергия полосового сигнала $s(t)$, выраженная через эквивалентный низкочастотный сигнал $s_l(t)$, равна

$$\mathscr{E} = \tfrac{1}{2}\int_{-\infty}^{\infty}|s_l(t)|^2\,dt, \quad (4.1.24)$$

где $|s(t)|$ является огибающей $a(t)$ для сигнала $s(t)$.

4.1.2. Представление линейных полосовых систем

Линейный фильтр (линейная система) может быть описан или своей импульсной характеристикой $h(t)$, или своей частотной характеристикой $H(f)$, которая является преобразованием Фурье от $h(t)$.

Поскольку $h(t)$ вещественно, то

$$H^*(-f) = H(f). \quad (4.1.25)$$

Определим

$$H_l(f - f_c) = \begin{cases} H(f) & (f > 0), \\ 0 & (f < 0). \end{cases} \quad (4.1.26)$$

Тогда

$$H_l^*(-f - f_c) = \begin{cases} 0 & (f > 0), \\ H^*(-f) & (f < 0). \end{cases} \quad (4.1.27)$$

Используя (4.1.25), получаем соотношение

$$H(f) = H_l(f - f_c) + H_l^*(-f - f_c), \quad (4.1.28)$$

которое похоже на (4.1.21), за исключением множителя $\frac{1}{2}$. Обратное преобразование Фурье $H(f)$ из (4.1.28) позволяет представить $h(t)$ в виде

$$h(t) = h_l(t)\,e^{j2\pi f_c t} + h_l^*(t)\,e^{-j2\pi f_c t} = \\ = 2\,\mathrm{Re}\left[h_l(t)\,e^{j2\pi f_c t}\right], \tag{4.1.29}$$

где $h_l(t)$ – обратное преобразование Фурье $H_l(f)$. В общем случае импульсная характеристика $h_l(t)$ эквивалентной низкочастотной системы принимает комплексные значения.

4.1.3. Отклик полосовой системы на полосовой сигнал

В разд. 4.1.1 и 4.1.2 мы показали, что узкополосные полосовые сигналы и системы могут быть представлены эквивалентными низкочастотными сигналами и системами. В этом разделе покажем, что выход полосовой системы на полосовой входной сигнал можно часто получить из эквивалентного низкочастотного входа сигнала и эквивалентной низкочастотной импульсной характеристики системы.

Предположим, что $s(t)$ – узкополосный полосовой сигнал, а $s_l(t)$ – эквивалентный низкочастотный сигнал. Этот сигнал поступает на узкополосную полосовую систему, определяемую своей полосовой импульсной характеристикой $h(t)$ или эквивалентной низкочастотной импульсной характеристикой (ИХ) $h_l(t)$. Выход полосовой системы также является полосовым сигналом, и, следовательно, его можно выразить в виде

$$r(t) = \mathrm{Re}\left[r_l(t)\,e^{j2\pi f_c t}\right], \tag{4.1.30}$$

где $r(t)$ связан со входным сигналом $s(t)$ и ИХ системы $h(t)$ интегралом свертки

$$r(t) = \int_{-\infty}^{\infty} s(\tau)\,h(t-\tau)\,d\tau. \tag{4.1.31}$$

Эквивалентно выход системы, представленной в частотной области, равен

$$R(f) = S(f)H(f). \tag{4.1.32}$$

С учетом (4.1.21) для $S(f)$ и (4.1.28) для $H(f)$ получаем результат

$$R(f) = \tfrac{1}{2}\left[S_l(f-f_c) + S_l^*(-f-f_c)\right]\times\left[H_l(f-f_c) + H_l^*(-f-f_c)\right]. \tag{4.1.33}$$

Когда $s(t)$ является узкополосным сигналом, а $h(t)$ – импульсной характеристикой узкополосной системы, то $S_l(f-f_c) = 0$ и $H_l(f-f_c) = 0$ для $f < 0$. Отсюда следует

$$S_l(f-f_c)H_l^*(-f-f_c) = 0, \quad S_l^*(-f-f_c)H_l(f-f_c) = 0.$$

Следовательно, (4.1.33) упрощается:

$$R(f) = \tfrac{1}{2}\left[S_l(f-f_c)H_l(f-f_c) + S_l^*(-f-f_c)H_l^*(-f-f_c)\right] = \\ = \tfrac{1}{2}\left[R_l(f-f_c) + R_l^*(-f-f_c)\right], \tag{4.1.34}$$

где

$$R_l(f) = S_l(f)H_l(f) \tag{4.1.35}$$

– спектр выхода эквивалентной низкочастотной системы, возбуждаемой эквивалентным низкочастотным сигналом. Ясно, что во временной области выход $r_l(t)$ определяется сверткой $s_l(t)$ и $h_l(t)$, т.е.

$$r_l(t) = \int_{-\infty}^{\infty} s_l(\tau)\,h_l(t-\tau)\,d\tau. \tag{4.1.36}$$

Комбинация (4.1.36) и (4.1.30) дает отношение между выходом полосовой системы $r(t)$ и эквивалентными низкочастотными функциями $s_l(t)$ и $h_l(t)$. Это простое отношение позволяет нам не учитывать произвольные линейные преобразования частот, которые встречаются при модуляции сигнала с целью смещения его спектра в частотной области конкретного канала.

Следовательно, для математического удобства будем иметь дело только с передачей эквивалентных низкочастотных сигналов через эквивалентные низкочастотные каналы.

4.1.4. Представление полосовых случайных процессов

Представление полосовых сигналов в разд. 4.1.1 касается детерминированных сигналов. В этом разделе рассмотрим представление полосовых стационарных случайных процессов. В частности, получим важные отношения между корреляционной функцией и спектральной плотностью мощности полосового сигнала и корреляционной функцией и спектральной плотностью мощности эквивалентного низкочастотного сигнала.

Предположим, что $n(t)$ является реализацией стационарного в широком смысле случайного процесса $N(t)$ с нулевым средним и спектральной плотностью мощности $\Phi_{nn}(f)$. Примем, что спектральная плотность мощности равна нулю вне интервала частот, группирующихся около частот $\pm f_c$, где f_c – *частота несущей*. Случайный процесс $N(t)$ называется *узкополосным полосовым случайным процессом*, если ширина его полосы частот Δf намного меньше f_c[1]. С учетом этого условия реализация процесса $n(t)$ может быть представлена в одной из трех форм, данных в разд. 4.1.1, а именно

$$n(t) = a(t)\cos\left[2\pi f_c t + \theta(t)\right], \tag{4.1.37}$$

$$n(t) = x(t)\cos 2\pi f_c t - y(t)\sin 2\pi f_c t, \tag{4.1.38}$$

$$n(t) = \mathrm{Re}\left[z(t)\,\mathrm{e}^{j2\pi f_c t}\right], \tag{4.1.39}$$

где $a(t)$ – огибающая, а $\theta(t)$ – фаза вещественного сигнала, $x(t)$ и $y(t)$ – квадратурные компоненты $n(t)$, а $z(t)$ – *комплексная огибающая* для $n(t)$.

Рассмотрим более подробно форму, определяемую (4.1.38). Сначала заметим, что если $N(t)$ имеет нулевое среднее, то случайные квадратурные компоненты $X(t)$ и $Y(t)$ должны также иметь нулевые средние. Далее, стационарность $N(t)$ подразумевает, что автокорреляционные и взаимокорреляционные функции $X(t)$ и $Y(t)$ обладают следующими свойствами:

$$\phi_{xx}(\tau) = \phi_{yy}(\tau), \tag{4.1.40}$$

$$\phi_{xy}(\tau) = -\phi_{yx}(\tau). \tag{4.1.41}$$

Покажем, что эти два свойства следуют из стационарности $N(t)$. Автокорреляционная функция $\phi_{nn}(\tau)$ для $N(t)$ равна

$$\begin{aligned}\phi_{nn}(\tau) &= E\left[N(t)N(t+\tau)\right] = E\left\{\left[X(t)\cos 2\pi f_c t - Y(t)\sin 2\pi f_c t\right]\times\right.\\ &\left.\times\left[X(t+\tau)\cos 2\pi f_c(t+\tau) - Y(t+\tau)\sin 2\pi f_c(t+\tau)\right]\right\} =\\ &= \phi_{xx}(\tau)\cdot\cos 2\pi f_c t\cdot\cos 2\pi f_c(t+\tau) + \phi_{yy}(\tau)\cdot\sin 2\pi f_c t\cdot\sin 2\pi f_c(t+\tau) -\\ &- \phi_{xy}(\tau)\cdot\sin 2\pi f_c t\cdot\cos 2\pi f_c(t+\tau) - \phi_{yx}(\tau)\cdot\cos 2\pi f_c t\cdot\sin 2\pi f_c(t+\tau).\end{aligned} \tag{4.1.42}$$

Используя соотношения

---

[1] В более общем случае достаточно потребовать, чтобы $\Delta f/2 < f_c$ (прп)

$$\cos A\cos B = \tfrac{1}{2}\left[\cos(A-B)+\cos(A+B)\right],$$
$$\sin A\sin B = \tfrac{1}{2}\left[\cos(A-B)-\cos(A+B)\right], \qquad (4.1.43)$$
$$\sin A\cos B = \tfrac{1}{2}\left[\sin(A-B)+\sin(A+B)\right]$$

в (4.1.42), получаем результат

$$\begin{aligned} E\left[N(t)N(t+\tau)\right] = {} & \tfrac{1}{2}\left[\phi_{xx}(\tau)+\phi_{yy}(\tau)\right]\cos 2\pi f_c\tau + \\ & + \tfrac{1}{2}\left[\phi_{xx}(\tau)-\phi_{yy}(\tau)\right]\cos 2\pi f_c(2t+\tau) - \\ & - \tfrac{1}{2}\left[\phi_{yx}(\tau)-\phi_{xy}(\tau)\right]\sin 2\pi f_c\tau - \\ & - \tfrac{1}{2}\left[\phi_{yx}(\tau)+\phi_{xy}(\tau)\right]\sin 2\pi f_c(2t+\tau). \end{aligned} \qquad (4.1.44)$$

Поскольку $N(t)$ – стационарный процесс, то правая часть (4.1.44) не должна зависеть от t. Но это условие может быть выполнено только при условии выполнения (4.1.40) и (4.1.41). Как следствие, (4.1.44) сводится к

$$\phi_{nn}(\tau) = \phi_{xx}(\tau)\cos 2\pi f_c\tau - \phi_{yx}(\tau)\sin 2\pi f_c\tau\,. \qquad (4.1.45)$$

Заметим, что соотношение между автокорреляционной функцией $\phi_{nn}(\tau)$ полосового процесса и корреляционной и взаимокорреляционной функциями $\phi_{xx}(\tau)$ и $\phi_{xy}(\tau)$ квадратурных компонент имеет форму (4.1.38), которая выражает полосовой процесс через квадратурные компоненты.

Автокорреляционная функция эквивалентного случайного низкочастотного процесса

$$Z(t) = X(t) + jY(t) \qquad (4.1.46)$$

определяется как

$$\phi_{zz}(\tau) = \tfrac{1}{2}E\left[Z^*(t)Z(t+\tau)\right]. \qquad (4.1.47)$$

Подставив (4.1.46) в (4.1.47) и выполнив соответствующие операции, получаем

$$\phi_{zz}(\tau) = \tfrac{1}{2}\left[\phi_{xx}(\tau)+\phi_{yy}(\tau)-j\phi_{xy}(\tau)+j\phi_{yx}(\tau)\right]. \qquad (4.1.48)$$

Теперь, если выполняются свойства (4.1.40) и (4.1.41), находим соотношение

$$\phi_{zz}(\tau) = \phi_{xx}(\tau) + j\phi_{yx}(\tau), \qquad (4.1.49)$$

которое выражает автокорреляционную функцию комплексной огибающей через автокорреляционную и взаимокорреляционную функцию квадратурных компонент. В заключение, используя результаты (4.1.49) и (4.1.45), имеем

$$\phi_{nn}(\tau) = \mathrm{Re}\left[\phi_{zz}(\tau)e^{j2\pi f_c\tau}\right]. \qquad (4.1.50)$$

Таким образом, автокорреляционная функция $\phi_{nn}(\tau)$ полосового случайного процесса $N(t)$ однозначно определяется автокорреляционной функцией $\phi_{zz}(\tau)$ эквивалентного низкочастотного случайного процесса $Z(t)$ и частоты несущей f_c. Спектральная плотность мощности $\Phi_{nn}(f)$ случайного процесса $N(t)$ определяется преобразованием Фурье $\phi_{zz}(\tau)$. Имеем

$$\Phi_{nn}(f) = \int_{-\infty}^{\infty}\left\{\mathrm{Re}\left[\phi_{zz}(\tau)e^{j2\pi f_c t}\right]\right\}e^{-j2\pi f\tau}\,d\tau = \tfrac{1}{2}\left[\Phi_{zz}(f-f_c)+\Phi_{zz}(-f-f_c)\right], \qquad (4.1.51)$$

где $\Phi_{zz}(f)$ – спектральная плотность мощности эквивалентного низкочастотного процесса $Z(t)$. Поскольку автокорреляционная функция $Z(t)$ удовлетворяет условию $\phi_{zz}(\tau) = \phi^*_{zz}(-\tau)$, то следует, что $\Phi_{zz}(f)$ является вещественной функцией частоты.

Свойства квадратурных компонент. Выше было показано, что взаимокорреляционная функция квадратурных компонент $X(t)$ и $Y(t)$ полосового стационарного случайного процесса $N(t)$ удовлетворяет условию симметрии (4.1.41). Далее, любая взаимокорреляционная функция удовлетворяет условию

$$\phi_{yx}(\tau) = \phi_{xy}(-\tau). \tag{4.1.52}$$

Из этих двух условий заключаем, что

$$\phi_{xy}(\tau) = -\phi_{xy}(-\tau). \tag{4.1.53}$$

Это означает, что $\phi_{xy}(\tau)$ является нечётной функцией τ. Следовательно, $\phi_{xy}(0) = 0$ и, значит, $X(t)$ и $Y(t)$ не коррелированы при $\tau = 0$. Конечно, это не означает, что процессы $X(t)$ и $Y(t+\tau)$ не коррелированы для всех τ, поскольку это бы означало, что $\phi_{xy}(\tau) = 0$ для всех τ. Если в самом деле $\phi_{xy}(\tau) = 0$ для всех τ, то $\phi_{zz}(\tau)$ является вещественной, и спектральная плотность мощности $\Phi_{zz}(f)$ удовлетворяет условию

$$\Phi_{zz}(f) = \Phi_{zz}(-f), \tag{4.1.54}$$

и наоборот. Это означает, что $\Phi_{zz}(f)$ симметрична относительно $f = 0$ (четная функция частоты). В частном случае, когда стационарный случайный процесс $N(t)$ гауссовский, квадратурные компоненты $X(t)$ и $Y(t+\tau)$ совместно гауссовские. Более того, при $\tau = 0$ они статистически независимы, и, следовательно, их совместная плотность вероятности

$$p(x,y) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}, \tag{4.1.55}$$

где дисперсия σ^2 определяется как $\sigma^2 = \phi_{xx}(0) = \phi_{yy}(0) = \phi_{nn}(0)$.

Представление белого шума. Белый шум является случайным процессом, который имеет постоянную спектральную плотность в неограниченном диапазоне частот. Этот вид шума не может быть выражен через узкополосные квадратурные компоненты вследствие широкополосности процесса.

В вопросах, связанных с демодуляцией узкополосных сигналов на фоне шумов, математически удобно представить аддитивный шум как белый и выразить его через квадратурные компоненты. Это можно выполнить, предполагая, что сигнал и шум на приёмной стороне прошли через идеальный полосовой фильтр, имеющий полосу пропускания более широкую, чем полоса сигнала. Такой фильтр может внести пренебрежимо малые искажения в сигнал, но он исключает частотные компоненты шума вне полосы пропускания фильтра.

Белый шум, прошедший через идеальный полосовой фильтр, называют полосовым белым шумом, и он имеет спектральную плотность вида, показанного на рис. 4.1.3. Полосовой белый шум можно представить в любой из форм, выражаемых формулами (4.1.37), (4.1.38) и (4.1.39). Спектральная плотность мощности и автокорреляционная функция эквивалентного белого низкочастотного шума равны соответственно

$$\Phi_{zz}(f) = \begin{cases} N_0 & \left(f \le \frac{1}{2}B\right), \\ 0 & \left(f > \frac{1}{2}B\right), \end{cases} \tag{4.1.56}$$

$$\phi_{zz}(\tau) = N_0 \frac{\sin \pi B\tau}{\pi\tau}. \tag{4.1.57}$$

Предельная форма $\phi_{zz}(\tau)$, когда полоса частот $B \to \infty$, выражается так:

$$\phi_{zz}(\tau) = N_0\delta(\tau). \tag{4.1.58}$$

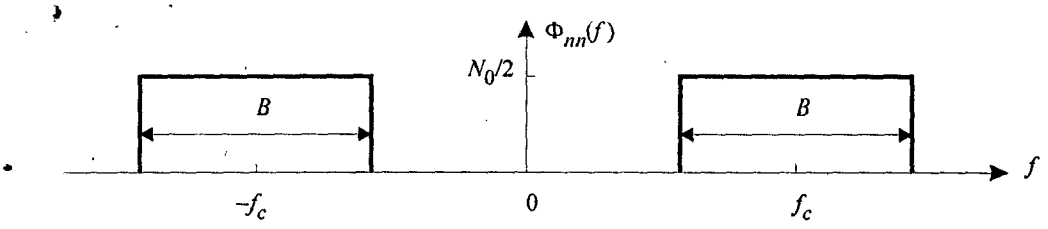


Рис. 4.1.3. Полосовой шум с равномерным спектром

Спектральная плотность мощности белого и полосового белого шума симметрична относительно $f = 0$, так что $\phi_{yx}(\tau) = 0$ для всех τ. Следовательно,

$$\phi_{zz}(\tau) = \phi_{xx}(\tau) = \phi_{yy}(\tau). \qquad (4.1.59)$$

Это означает, что квадратурные компоненты $X(t)$ и $Y(t)$ не коррелированы при всех временных сдвигах τ, а автокорреляционные функции $Z(t)$, $X(t)$ и $Y(t)$ одинаковы.

4.2. ГЕОМЕТРИЧЕСКОЕ ПРЕДСТАВЛЕНИЕ СИГНАЛОВ

В этом разделе мы продемонстрируем, что сигналы имеют характеристики, которые похожи на векторы, и приведем векторное представление сигналов. Начнём с некоторых базовых определений и концепций для векторов.

4.2.1. Концепция векторного пространства

Вектор $\mathbf{v}$ в n-мерном пространстве характеризуется своими n компонентами $[v_1\ v_2\ \ldots\ v_n]$. Его можно также представить как линейную комбинацию *единичных векторов* или *базисных векторов* $\mathbf{e}_i$, $1 \le i \le n$, т.е.

$$\mathbf{v} = \sum_{i=1}^{n} v_i\, \mathbf{e}_i, \qquad (4.2.1)$$

где, по определению, единичный вектор имеет единичную длину, а v_i является проекцией вектора $\mathbf{v}$ на единичный вектор $\mathbf{e}_i$.

Скалярное произведение двух n-мерных векторов $\mathbf{v}_1 = [v_{11}\ v_{12}\ \ldots\ v_{1n}]$ и $\mathbf{v}_2 = [v_{21}\ v_{22}\ \ldots\ v_{2n}]$ определяется как

$$\mathbf{v}_1 \cdot \mathbf{v}_2 = \sum_{i=1}^{n} v_{1i}\, v_{2i}. \qquad (4.2.2)$$

Два вектора $\mathbf{v}_1$ и $\mathbf{v}_2$ ортогональны, если $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$. В более общем виде совокупность m векторов $\mathbf{v}_k$, $1 \le k \le m$, ортогональна, если

$$\mathbf{v}_i \cdot \mathbf{v}_j = 0 \qquad (4.2.3)$$

для всех $1 \le i, j \le m$ и $i \ne j$.

Норма вектора $\mathbf{v}$ обозначается $\|\mathbf{v}\|$ и определяется

$$\|\mathbf{v}\| = (\mathbf{v} \cdot \mathbf{v})^{1/2} = \sqrt{\sum_{i=1}^{n} v_i^2}. \qquad (4.2.4)$$

Это просто длина вектора. Ансамбль m векторов называется *ортонормированным*, если все векторы ортогональны и каждый вектор имеет единичную норму. Совокупность m векторов называется линейно независимой, если ни один вектор не может быть представлен как линейная комбинация оставшихся векторов.

Два n-мерных вектора $\mathbf{v}_1$ и $\mathbf{v}_2$ удовлетворяют неравенству треугольника

$$\|\mathbf{v}_1 + \mathbf{v}_2\| \le \|\mathbf{v}_1\| + \|\mathbf{v}_2\|, \tag{4.2.5}$$

а равенство имеет место, если $\mathbf{v}_1$ и $\mathbf{v}_2$ имеют одинаковое направление, т.е. $\mathbf{v}_1 = a\mathbf{v}_2$, где a является положительным вещественным скаляром. Из неравенства треугольника следует неравенство Коши-Шварца

$$|\mathbf{v}_1 \cdot \mathbf{v}_2| \le \|\mathbf{v}_1\| \|\mathbf{v}_2\|, \tag{4.2.6}$$

с равенством, если $\mathbf{v}_1 = a\mathbf{v}_2$ [1]. Квадрат нормы суммы двух векторов можно выразить так:

$$\|\mathbf{v}_1 + \mathbf{v}_2\|^2 = \|\mathbf{v}_1\|^2 + \|\mathbf{v}_2\|^2 + 2\mathbf{v}_1 \cdot \mathbf{v}_2. \tag{4.2.7}$$

Если $\mathbf{v}_1$ и $\mathbf{v}_2$ ортогональны, тогда $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$ и, следовательно,

$$\|\mathbf{v}_1 + \mathbf{v}_2\|^2 = \|\mathbf{v}_1\|^2 + \|\mathbf{v}_2\|^2. \tag{4.2.8}$$

Это соотношение Пифагора для двух ортогональных n-мерных векторов. Напомним из алгебры матриц, что линейное преобразование в n-мерном векторном пространстве является матричным преобразованием вида

$$\mathbf{v}' = \mathbf{A}\mathbf{v}, \tag{4.2.9}$$

где матрица $\mathbf{A}$ преобразует вектор $\mathbf{v}$ в некоторый вектор $\mathbf{v}'$. В специальном случае, когда $\mathbf{v}' = \lambda\mathbf{v}$, т.е.

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}, \tag{4.2.10}$$

где λ – некоторый (положительный или отрицательный) скаляр, вектор $\mathbf{v}$ называется собственным вектором преобразования, а λ является соответствующим собственным значением.

В конце рассмотрим процедуру Грама–Шмидта для образования ансамбля ортонормированных векторов из ряда n-мерных векторов $\mathbf{v}_i$, $1 \le i \le m$. Мы начинаем выбором произвольного вектора ряда, скажем $\mathbf{v}_1$. Путем нормировки его длины получаем первый вектор ансамбля

$$\mathbf{u}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}. \tag{4.2.11}$$

Затем можем выбрать $\mathbf{v}_2$ и получить проекцию $\mathbf{v}_2$ на $\mathbf{u}_1$. Образуем вектор

$$\mathbf{u}_2' = \mathbf{v}_2 - (\mathbf{v}_2 \cdot \mathbf{u}_1)\mathbf{u}_1. \tag{4.2.12}$$

Далее нормируем вектор $\mathbf{u}_2'$ к единичной длине. Это даёт

$$\mathbf{u}_2 = \frac{\mathbf{u}_2'}{\|\mathbf{u}_2'\|}. \tag{4.2.13}$$

Процедура продолжается выбором вектора $\mathbf{v}_3$ и образованием проекции $\mathbf{v}_3$ на $\mathbf{u}_2$. Таким образом получаем

$$\mathbf{u}_3' = \mathbf{v}_3 - (\mathbf{v}_3 \cdot \mathbf{u}_1)\mathbf{u}_1 - (\mathbf{v}_3 \cdot \mathbf{u}_2)\mathbf{u}_2. \tag{4.2.14}$$

Затем образуется ортогональный вектор $\mathbf{u}_3$:

$$\mathbf{u}_3 = \frac{\mathbf{u}_3'}{\|\mathbf{u}_3'\|}. \tag{4.2.15}$$

Продолжая эту процедуру, можем образовать ансамбль из n_1 ортонормированных векторов, где в общем $n_1 \le n$. Если $m < n$, то $n_1 \le m$, а если $m \ge n$, то $n_1 \le n$.

---

[1] Можно показать, что неравенство (4.2.6) переходит в равенство не только при положительных, но и при отрицательных a (прп)

4.2.2. Концепция пространства сигналов

Как в случае векторов, мы можем провести параллельное рассмотрение ряда сигналов, определенных на некотором интервале $[a,b]$[1]. Скалярное произведение двух, в общем случае комплексных сигналов $x_1(t)$ и $x_2(t)$ обозначается $\langle x_1(t), x_2(t)\rangle$ и определяется как

$$\langle x_1(t), x_2(t)\rangle = \int_a^b x_1(t)\, x_2^*(t)\, dt. \qquad (4.2.16)$$

Сигналы ортогональны, если их скалярное произведение равно нулю. Норма сигнала определяется так:

$$\|x(t)\| = \left(\int_a^b |x(t)|^2\, dt \right)^{1/2}. \qquad (4.2.17)$$

Ансамбль m сигналов называется ортонормированным, если все сигналы попарно ортогональны, а их нормы равны 1. Сигналы линейно независимы, если ни один сигнал не выражается как линейная комбинация остальных сигналов. Неравенство треугольника для двух сигналов выражается подобно (4.2.5):

$$\|x_1(t) + x_2(t)\| \le \|x_1(t)\| + \|x_2(t)\|, \qquad (4.2.18)$$

а неравенство Коши-Шварца выражается подобно (4.2.6):

$$\left| \int_a^b x_1(t)\, x_2^*(t)\, dt \right| \le \left| \int_a^b |x_1(t)|^2\, dt \right|^{1/2} \left| \int_a^b |x_2(t)|^2\, dt \right|^{1/2}, \qquad (4.2.19)$$

причём равенство имеет место, если $x_2(t) = a x_1(t)$, где a – произвольное комплексное число.

4.2.3. Ортогональное разложение сигналов

В этом разделе мы ознакомимся с векторным представлением сигналов и таким образом продемонстрируем эквивалентность между сигналами и их векторными представлениями.

Предположим, что $s(t)$ является детерминированным вещественным сигналом с ограниченной энергией

$$\mathscr{E}_s = \int_{-\infty}^{\infty} [s(t)]^2 dt < \infty. \qquad (4.2.20)$$

Далее, предположим, что существует ансамбль функций $\{f_n(t), \quad n = 1,2,\ldots,N\}$, который ортонормирован в том смысле, что

$$\int_{-\infty}^{\infty} f_n(t)\, f_m(t)\, dt = \begin{cases} 0 & (m \ne n), \\ 1 & (m = n). \end{cases} \qquad (4.2.21)$$

Мы можем аппроксимировать сигнал $s(t)$ при помощи взвешенной линейной комбинации этих функций, т.е.

$$\hat{s}(t) = \sum_{k=1}^{K} s_k\, f_k(t), \qquad (4.2.22)$$

где $\{s_k, \quad 1 \le k \le K\}$ – коэффициенты в аппроксимации $s(t)$. Ошибка аппроксимации

$$e(t) = s(t) - \hat{s}(t). \qquad (4.2.23)$$

Выберем коэффициенты $\{s_k\}$ так, чтобы минимизировать энергию $\mathscr{E}_e$ ошибки аппроксимации. Имеем

---

[1] Автор в дальнейшем отождествляет интервал и отрезок (прп)

$$\mathscr{E}_e = \int_{-\infty}^{\infty} \left[s(t) - \hat{s}(t) \right]^2 dt = \int_{-\infty}^{\infty} \left[s(t) - \sum_{k=1}^{K} s_k \, f_k(t) \right]^2 dt . \qquad (4.2.24)$$

Оптимальные коэффициенты $\{s_k\}$ в представлении $s(t)$ рядом можно найти путём дифференцирования (4.2.24) по каждому из коэффициентов и приравнять первые производные нулю. В качестве альтернативы можем использовать хорошо известный результат из теории оценок, основанный на критерии минимума среднего квадрата ошибки оценивания, который гласит, что минимум $\mathscr{E}_e$ по s_k достигается тогда, когда ошибка ортогональна к каждой из функций ряда, т.е.

$$\int_{-\infty}^{\infty} \left[s(t) - \sum_{k=1}^{K} s_k \, f_k(t) \right] f_n(t) \, dt = 0, \qquad n = 1,2,\ldots,K . \qquad (4.2.25)$$

Поскольку функции $\{f_n(t)\}$ ортонормированы, из 4.2.25 следует

$$s_n = \int_{-\infty}^{\infty} s(t) \, f_n(t) \, dt, \qquad n = 1,2,\ldots,K . \qquad (4.2.26)$$

Таким образом, коэффициенты получаются как проекции сигнала $s(t)$ на каждую из функций $\{f_n(t)\}$. Как следствие, $\hat{s}(t)$ является проекцией $s(t)$ в K-мерном пространстве сигналов, заданном функциями $\{f_n(t)\}$. Иногда говорят, что пространство натянуто на функции $\{f_n(t)\}$. Минимальное значение среднего квадрата ошибки аппроксимации равно

$$\mathscr{E}_{\min} = \int_{-\infty}^{\infty} e(t) \, s(t) \, dt = \int_{-\infty}^{\infty} \left[s(t) \right]^2 dt - \int_{-\infty}^{\infty} \sum_{k=1}^{K} s_k \, f_k(t) \, s(t) \, dt = \mathscr{E}_s - \sum_{k=1}^{K} s_k^2 , \qquad (4.2.27)$$

и оно не отрицательно по определению.

Когда средний квадрат ошибки $\mathscr{E}_{\min} = 0$, то

$$\mathscr{E}_s = \sum_{k=1}^{K} s_k^2 = \int_{-\infty}^{\infty} \left[s(t) \right]^2 dt . \qquad (4.2.28)$$

При условии, что $\mathscr{E}_{\min} = 0$, сигнал $s(t)$ можно выразить так:

$$s(t) = \sum_{k=1}^{K} s_k \, f_k(t) . \qquad (4.2.29)$$

Равенство $s(t)$ правой части (4.2.20) понимается в том смысле, что ошибка представления имеет нулевую энергию.

Если каждый сигнал с ограниченной энергией можно представить рядом (4.2.29) при $\mathscr{E}_{\min} = 0$, совокупность ортонормированных функций $\{f_n(t)\}$ называют полной[1].

Пример 4.2.1. Тригонометрический ряд Фурье. Сигнал $s(t)$ с ограниченной энергией, который равен нулю везде, кроме области $0 \le t \le T$, и имеет ограниченное число разрывов на этом интервале, может быть представлен рядом Фурье:

$$s(t) = \sum_{k=1}^{\infty} \left(a_k \cos \frac{2\pi k t}{T} + b_k \sin \frac{2\pi k t}{T} \right), \qquad (4.2.30)$$

где коэффициенты $\{a_k, b_k\}$, которые минимизируют средний квадрат ошибки[2], определяются выражениями

---

[1] Для непрерывных сигналов (как в примере 4.3.1) это возможно, только если K не ограничено. Только тогда ортонормированный ансамбль является полным, а представление (4.2.29) называется обобщённым рядом Фурье (прп).

[2] При конечном числе членов ряда (прп).

$$a_k = \frac{1}{\sqrt{T}} \int_0^T s(t) \cos \frac{2\pi k t}{T} dt,$$
$$b_k = \frac{1}{\sqrt{T}} \int_0^T s(t) \sin \frac{2\pi k t}{T} dt. \qquad (4.2.31)$$

Ансамбль ортонормированных тригонометрических функций

$$\left\{ \sqrt{2/T} \cos 2\pi k\, t/T,\ \sqrt{2/T} \sin 2\pi k\, t/T \right\}$$

является полным, и, следовательно, ряд (4.2.30) обеспечивает нулевой средний квадрат ошибки. Эти свойства легко устанавливаются из проведённого выше рассмотрения.

Процедура Грама–Шмидта. Теперь предположим, что мы имеем ансамбль сигналов с ограниченной энергией $\{s_i(t),\ i = 1,2,\ldots M\}$, и хотим сконструировать ансамбль ортонормированных сигналов. Процедура ортонормирования Грама–Шмидта позволяет нам сконструировать такой ансамбль. Начнем с первого сигнала $s_1(t)$, причём предполагается, что он имеет энергию $\mathscr{E}_1$. Первый сигнал ортонормированного ансамбля конструируется легко:

$$f_1(t) = \frac{s_1(t)}{\sqrt{\mathscr{E}_1}}. \qquad (4.2.32)$$

Таким образом, сигнал $f_1(t)$ имеет форму $s_1(t)$, но нормирован к единичной энергии. Второй сигнал конструируется из $s_2(t)$, причём сначала вычисляется проекция $s_2(t)$ на $f_1(t)$:

$$c_{12} = \int_{-\infty}^{\infty} s_2(t) f_1(t) dt. \qquad (4.2.33)$$

Затем $c_{12} f_1(t)$ вычитается из $s_2(t)$ для получения

$$f_2'(t) = s_2(t) - c_{12} f_1(t). \qquad (4.2.34)$$

Этот сигнал ортогонален $f_1(t)$, но не имеет единичной энергии. Если $\mathscr{E}_2$ означает энергию для $f_2'(t)$, то нормированный сигнал, который ортогонален к $f_1(t)$, равен

$$f_2(t) = \frac{f_2'(t)}{\sqrt{\mathscr{E}_2}}. \qquad (4.2.35)$$

В общем, ортогонализация k-й функции ведёт к

$$f_k(t) = \frac{f_k'(t)}{\sqrt{\mathscr{E}_k}}, \qquad (4.2.36)$$

где

$$f_k'(t) = s_k(t) - \sum_{i=1}^{k-1} c_{ik} f_i(t) \qquad (4.2.37)$$

и

$$c_{ik} = \int_{-\infty}^{\infty} s_k(t) f_i(t) dt, \qquad i = 1,2,\ldots,k-1. \qquad (4.2.38)$$

Таким образом, процесс ортогонализации продолжается, пока все M сигналов не исчерпаны и не образованы $N \le M$ ортонормированных сигналов. Размерность N-сигнального пространства равна M, если исходные сигналы ансамбля линейно независимы, т.е. ни один из сигналов не является линейной комбинацией других сигналов.

Пример 4.2.2. Применим процедуру Грама–Шмидта к ансамблю четырёх сигналов, показанных на рис. 4.2.1(a). Сигнал $s_1(t)$ имеет энергию $\mathscr{E}_1 = 2$, так что $f_1(t) = \sqrt{\frac{1}{2}} s_1(t)$. Далее мы видим, что $c_{12} = 0$; следовательно, $s_2(t)$ и $f_1(t)$ ортогональны. Как следствие, $f_2(t) = s_2(t)/\sqrt{\mathscr{E}_k} = \sqrt{\frac{1}{2}} s_2(t)$. Чтобы получить $f_3(t)$, вычислим c_{13} и c_{23}, которые равны $c_{13} = \sqrt{2}$ и $c_{23} = 0$.

Таким образом,

$$f_3'(t) = s_3(t) - \sqrt{2}\, f_1(t) = \begin{cases} -1 & (2 \le t \le 3), \\ 0 & (\text{для других } t). \end{cases}$$

Поскольку $f_3'(t)$ имеет единичную энергию, то следует, что $f_3(t) = f_3'(t)$. Для определения $f_4(t)$ находим, что $c_{14} = -\sqrt{2}$, $c_{24} = 0$ и $c_{34} = 1$. Поэтому

$$f_4'(t) = s_4(t) + \sqrt{2}\, f_1(t) - f_3(t) = 0.$$

Как следствие, $s_4(t)$ является линейной комбинацией $f_1(t)$ и $f_3(t)$ и поэтому $f_4(t) = 0$.

Три ортонормированные функции показаны на рис. 4.2.1(b).

Поскольку мы сконструировали ансамбль ортонормированных сигналов $\{f_n(t)\}$, можем выразить M сигналов $\{s_n(t)\}$ как линейную комбинацию от $\{f_n(t)\}$.

Таким образом, можно написать

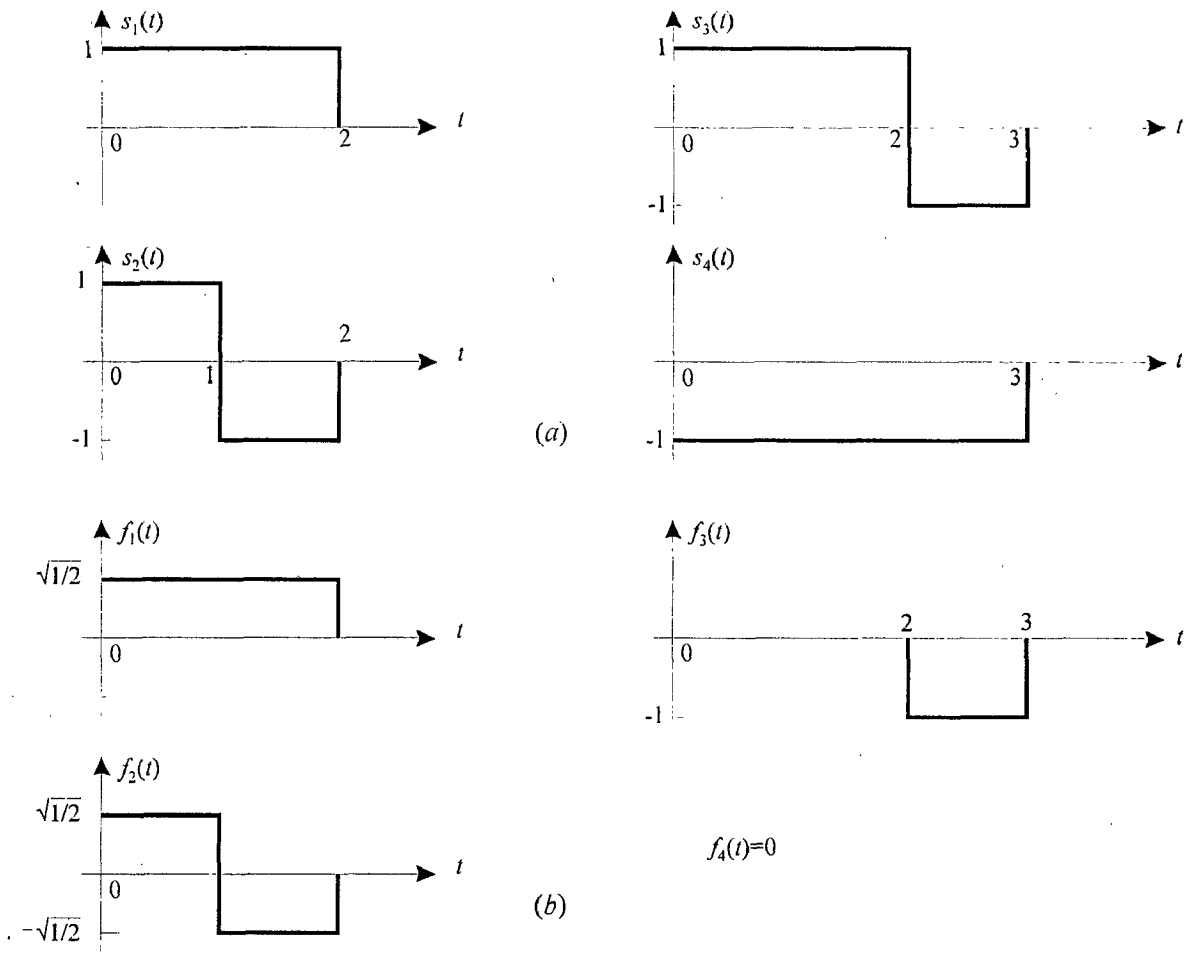


Рис. 4.2.1. Ортогонализация Грама–Шмидта для сигналов $\{s_i(t),\ i=1, 2, 3\ (a)$ и соответствующие ортогональные сигналы (b)

$$s_k(t) = \sum_{k=1}^{N} s_{kn} f_n(t), \qquad k = 1,2,\ldots, M, \tag{4.2.39}$$

и

$$\mathscr{E}_k = \int_{-\infty}^{\infty} \left[s_k(t)\right]^2 dt = \sum_{n=1}^{N} s_{kn}^2 = \left\|\mathbf{s}_k\right\|^2. \tag{4.2.40}$$

Основываясь на выражении (4.2.39), каждый сигнал можно представить вектором

$$\mathbf{s}_k = \left[s_{k1}\ s_{k1}\ \ldots\ s_{kN}\right]. \tag{4.2.41}$$

или, что эквивалентно, точкой в N-мерном пространстве сигналов с координатами $\left\{s_{ki},\ i = 1,2,\ldots N\right\}$. Энергия k-го сигнала равна квадрату длины вектора или, что эквивалентно, квадрату евклидова расстояния от начала координат к точке N-мерного пространства. Таким образом, любой сигнал можно представить геометрически как точку в пространстве сигналов, заданном ортонормированными функциями.

Пример 4.2.3. Получим векторное представление четырех сигналов, показанных на рис. 4.2.1(а), используя ортонормальный ансамбль функций из рис. 4.2.1(*b*).

Поскольку размерность пространства сигналов $N = 3$, каждый сигнал описывается тремя компонентами. Сигнал $s_1(t)$ характеризуется вектором $\mathbf{s}_1 = \left(\sqrt{2},0,0\right)$. Аналогично сигналы $s_2(t)$, $s_3(t)$, $s_4(t)$ характеризуются соответственно векторами $\mathbf{s}_2 = \left(0,\sqrt{2},0\right)$, $\mathbf{s}_3 = \left(\sqrt{2},0,1\right)$, $\mathbf{s}_4 = \left(-\sqrt{2},0,1\right)$. Эти векторы показаны на рис. 4.2.2. Их длины равны $\left|\mathbf{s}_1\right| = \sqrt{2}$, $\left|\mathbf{s}_2\right| = \sqrt{2}$, $\left|\mathbf{s}_3\right| = \sqrt{3}$, $\left|\mathbf{s}_4\right| = \sqrt{3}$, а соответствующие энергии сигналов $\mathscr{E}_k = \left|\mathbf{s}_k\right|^2$, $k = 1,2,3,4$.

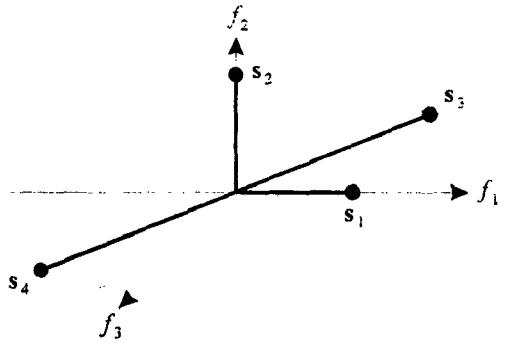


Рис. 4.2.2. Четыре сигнальных вектора, представленных в виде точек в трехмерном функциональном пространстве

Мы показали, что ансамбль M сигналов с ограниченной энергией можно представить взвешенной линейной комбинацией ортонормированных функций $\left\{f_n(t)\right\}$ размерностью $N \le M$. Функции $\left\{f_n(t)\right\}$ получены применением процедуры ортонормализации Грама–Шмидта из $\left\{s_n(t)\right\}$. Следует подчеркнуть, что функции $\left\{f_n(t)\right\}$, полученные преобразованием Грама–Шмидта, не являются уникальными (единственными). Если мы изменим порядок формирования ортонормированных сигналов из $\left\{s_n(t)\right\}$, получим другой ортонормированный ансамбль и соответствующее векторное представление сигналов $\left\{s_n(t)\right\}$ будет зависеть от выбора ортонормальных функций $\left\{f_n(t)\right\}$. Все же, вектора $\left\{\mathbf{s}_n\right\}$

будут сохранять геометрическую конфигурацию и их длины будут инвариантны по отношению к выбору ортонормированных функций $\{f_n(t)\}$.

Пример 4.2.4. Альтернативный ансамбль ортонормированных функций для четырёх сигналов из рис. 4.2.1 показан на рис. 4.2.3(*a*).

Используя эти функции для представления $\{s_n(t)\}$, получаем соответствующие векторы $\mathbf{s}_1 = (1,1,0)$, $\mathbf{s}_2 = (1,-1,0)$, $\mathbf{s}_3 = (1,1,-1)$, $\mathbf{s}_4 = (-1,-1,-1)$, которые показаны на рис. 4.2.3(*b*). Заметим, что длины векторов идентичны тем, которые получены из прежних ортонормированных функций $\{f_n(t)\}$.

Ортогональные представления, описанные выше, были разработаны для вещественных сигналов. Рассмотрение комплексных сигналов оставлено как упражнение для читателей (см. задачи 4.6 и 4.7).

В заключение рассмотрим случай, когда сигнал является полосовым и представлен в виде

$$s_m(t) = \mathrm{Re}\left[s_{lm}(t)\mathrm{e}^{j2\pi f_c t}\right], \qquad m = 1,2,\ldots,M, \tag{4.2.42}$$

где $s_{lm}(t)$ – эквивалентные низкочастотные сигналы. Напомним, что энергии сигналов можно выразить через $s_m(t)$ или $s_{lm}(t)$ так:

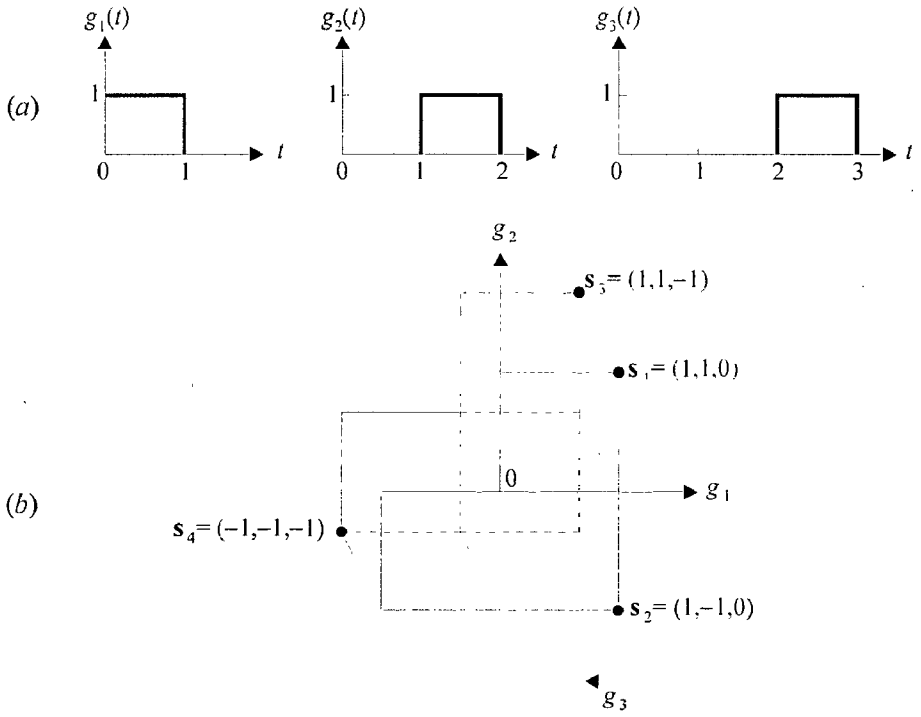


Рис. 4.2.3. Альтернативный ансамбль ортонормированных функций для четырех сигналов рис. 4.2.1 (*a*) и соответствующие сигнальные точки (*b*)

$$\mathscr{E}_m = \int_{-\infty}^{\infty} s_m^2(t)\,dt = \int_{-\infty}^{\infty} |s_{lm}(t)|^2\,dt. \tag{4.2.43}$$

Похожесть между сигналами любой пары, например $s_m(t)$ и $s_k(t)$, измеряется коэффициентом взаимной корреляции

$$\frac{1}{\sqrt{\mathscr{E}_m \mathscr{E}_k}} \int_{-\infty}^{\infty} s_m(t) s_k(t)\,dt = \mathrm{Re}\left\{\frac{1}{2\sqrt{\mathscr{E}_m \mathscr{E}_k}} \int_{-\infty}^{\infty} s_{lm}(t) s_{lk}^*(t)\,dt\right\}. \tag{4.2.44}$$

Определим комплексный коэффициент взаимной корреляции ρ_{km} так:

$$\rho_{km} = \frac{1}{2\sqrt{\mathscr{E}_m \mathscr{E}_k}} \int_{-\infty}^{\infty} s_{lm}^*(t) s_{lk}(t)\, dt. \qquad (4.2.45)$$

Тогда

$$\mathrm{Re}(\rho_{km}) = \frac{1}{\sqrt{\mathscr{E}_m \mathscr{E}_k}} \int_{-\infty}^{\infty} s_m(t) s_k(t)\, dt, \qquad (4.2.46)$$

или, что эквивалентно,

$$\mathrm{Re}(\rho_{km}) = \frac{\mathbf{s}_m \cdot \mathbf{s}_k}{\|\mathbf{s}_m\| \, \|\mathbf{s}_k\|} = \frac{\mathbf{s}_m \cdot \mathbf{s}_k}{\sqrt{\mathscr{E}_m \mathscr{E}_k}}. \qquad (4.2.47)$$

Коэффициенты взаимной корреляции между парами сигналов или сигнальных векторов определяют совокупность параметров, характеризующих похожесть ансамбля сигналов. Другой родственный параметр – расстояние Евклида $d_{km}^{(e)}$ между парой сигналов – определяется так:

$$d_{km}^{(e)} = \|\mathbf{s}_m - \mathbf{s}_k\| = \left\{ \int_{-\infty}^{\infty} [s_m(t) - s_k(t)]^2 \, dt \right\}^{1/2} = \left\{ \mathscr{E}_m + \mathscr{E}_k - 2\sqrt{\mathscr{E}_m \mathscr{E}_k}\, \mathrm{Re}(\rho_{km}) \right\}^{1/2}. \qquad (4.2.48)$$

Когда $\mathscr{E}_m = \mathscr{E}_k = \mathscr{E}$ для всех m и k, это выражение упрощается:

$$d_{km}^{(e)} = \left\{ 2\mathscr{E}\left[1 - \mathrm{Re}(\rho_{km})\right] \right\}^{1/2}. \qquad (4.2.49)$$

Итак, расстояние Евклида является альтернативной мерой похожести (или несходства) совокупности сигналов или соответствующих сигнальных векторов.

В следующем разделе мы опишем сигналы цифровой модуляции и используем пространство сигналов для их представления. Можно заметить, что сигналы цифровой модуляции удобно представить через две ортонормированные базисные функции вида

$$\begin{aligned} f_1(t) &= \sqrt{\tfrac{2}{T}} \cos 2\pi f_c t, \\ f_2(t) &= -\sqrt{\tfrac{2}{T}} \sin 2\pi f_c t. \end{aligned} \qquad (4.2.50)$$

Если $s_{lm}(t)$ выразить как $s_{lm}(t) = x_l(t) + jy_l(t)$, то следует, что $s_m(t)$ в (4.2.42) можно выразить так:

$$s_m(t) = x_l(t) f_1(t) + y_l(t) f_2(t), \qquad (4.2.51)$$

где $x_l(t)$ и $y_l(t)$ представляют модулирующие сигналы.

4.3. ПРЕДСТАВЛЕНИЕ СИГНАЛОВ ЦИФРОВОЙ МОДУЛЯЦИИ

При передаче цифровой информации по каналам связи модулятор является устройством отображения цифровой информации в форму аналоговых сигналов, которые согласованы с характеристиками каналов. Отображение обычно осуществляется посредством выбора блоков из $k = \log_2 M$ двоичных символов из символов информационной последовательности $\{a_n\}$ и выбора одного из $M = 2^k$ детерминированных сигналов с ограниченной энергией $\{s_m(t),\ m = 1,2,\ldots,M\}$, для передачи его по каналу за время передачи k информационных символов.

Когда отображение цифровой последовательности $\{a_n\}$ в сигнал осуществляется так, что сигнал, передаваемый на данном временном интервале, зависит от одного или более сигналов, переданных раньше, то говорят, что модулятор имеет *память*. С другой стороны, если отображение информационной последовательности $\{a_n\}$ в сигналы $\{s_m(t)\}$ осуществляется так, что передаваемые сигналы не зависят от ранее переданных, модулятор называют *без памяти*.

В дополнение к классификации модуляторов на модуляторы с памятью или без памяти мы их еще классифицируем как линейные или нелинейные. Линейность требует выполнения принципа суперпозиции (наложения) при отображении цифровой информационной последовательности в последовательные сигналы. При нелинейной модуляции принцип суперпозиций не применим для сигналов, передаваемых в последовательные временные интервалы.

Начнем с описания методов модуляции без памяти.

4.3.1. Методы модуляции без памяти

Как сказано выше, модулятор в цифровой системе связи отображает последовательность информационных символов в соответствующую последовательность сигналов. Эти сигналы могут отличаться по амплитуде, по фазе или по частоте или могут зависеть от двух или более сигнальных параметров. Мы рассмотрим каждый из этих видов сигналов отдельно, а начнём с линейной цифровой амплитудно-импульсной модуляции (АИМ), которую проще называют амплитудной модуляцией (АМ). Во всех случаях предполагаем, что информационная последовательность символов на входе модулятора является двоичной и появляется со скоростью R бит/с.

Амплитудно-импульсная модуляция. Цифровой АМ сигнал можно представить так:

$$\begin{aligned} s_m(t) &= \mathrm{Re}\left[A_m g(t) e^{j2\pi f_c t}\right] = \\ &= A_m g(t)\cos 2\pi f_c t, \qquad m = 1,2,\ldots,M, \qquad 0 \le t \le T, \end{aligned} \tag{4.3.1}$$

где $\{A_m,\ 1 \le m \le M\}$ означает ряд из M возможных амплитуд, соответствующих $M = 2^k$ возможным k-битовым блокам или *символам*. Амплитуда сигнала A_m принимает дискретные значения (уровни)

$$A_m = (2m - 1 - M)d, \qquad m = 1,2,\ldots,M, \tag{4.3.2}$$

где $2d$ – расстояние между соседними амплитудами сигналов. Сигнал $g(t)$ является вещественным сигнальным импульсом, форма которого определяет спектр передаваемого сигнала, как мы увидим позже. Скорость передачи канальных символов при АМ равна R/k. Это скорость, с которой происходят изменения амплитуды гармонической несущей для того, чтобы отразить передачу новой информации. Временной интервал $T_b = 1/R$ называется информационным (*битовым*) *интервалом*, а временной интервал $T = k/R = kT_b$ называется *символьным интервалом* или *интервалом канального символа*.

Сигналы АМ имеют энергию

$$\mathscr{E}_m = \int_0^T s_m^2(t)\,dt = \tfrac{1}{2} A_m^2 \int_0^T g^2(t)\,dt = \tfrac{1}{2} A_m^2 \mathscr{E}_g, \tag{4.3.3}$$

где $\mathscr{E}_g$ означает энергию импульса $g(t)$.

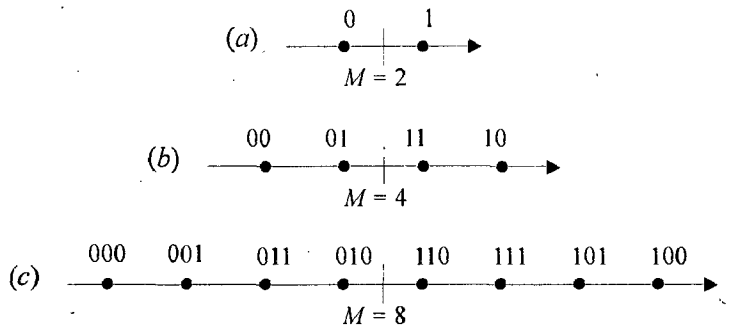


Рис. 4.3.1. Пространственная диаграмма сигналов цифровой АМ

Ясно, что сигналы АМ являются одномерными ($N=1$), и, следовательно, их можно представить в общем виде так:

$$s_m(t) = s_m f(t), \qquad (4.3.4)$$

где $f(t)$ определен как полосовой сигнал с единичной энергией:

$$f(t) = \sqrt{\frac{2}{\mathscr{E}_g}} g(t)\cos 2\pi f_c t, \qquad (4.3.5)$$

и

$$s_m = A_m \sqrt{\tfrac{1}{2}\mathscr{E}_g}, \qquad m = 1,2,\ldots,M. \qquad (4.3.6)$$

На рис. 4.3.1 даны соответствующие пространственные диаграммы сигналов для $M=2$, $M=4$, $M=8$. Цифровая АМ называется также *модуляцией с амплитудным сдвигом* (MAC, ASK).

Отображение или задание k информационных бит $M=2^k$ возможными амплитудами сигнала можно сделать различными способами. Наилучшее задание – это такое, при котором соседние амплитуды сигналов соответствуют информационным двоичным блокам, различающимся в одном разряде, как показано на рис. 4.3.1. Это отображение называется кодом Грея. Он важен при демодуляции сигнала, поскольку наиболее вероятные ошибки вызывает ошибочный выбор амплитуды, соседней по отношению к той, которая действительно передана. В этом случае, в k-битовой информационной последовательности возникает ошибка только в одном бите.

Заметим, что евклидово расстояние между какой-либо парой сигнальных точек равно

$$d_{mn}^{(e)} = \sqrt{\left(s_m - s_n\right)^2} = \sqrt{\tfrac{1}{2}\mathscr{E}_g}\,|A_m - A_n| = d\sqrt{2\mathscr{E}_g}\,|m-n|. \qquad (4.3.7)$$

Следовательно, расстояние между парой соседних сигнальных точек, т.е. минимальное значение евклидова расстояния, равно

$$d_{\min}^{(e)} = d\sqrt{2\mathscr{E}_g}. \qquad (4.3.8)$$

Модулированные сигналы АМ, представленные (4.3.1), являются двухполосными (ДП) сигналами и требуют в два раза большую полосу частот, чем низкочастотный передаваемый сигнал. В качестве альтернативы можем использовать однополосную (одной боковой полосы, ОБП) АМ, которую можно представить (нижнюю или верхнюю полосу) так:

$$s_m(t) = \mathrm{Re}\left\{A_m\left[g(t) \pm j\hat{g}(t)\right]e^{j2\pi f_c t}\right\}, \qquad m = 1,2,\ldots,M, \qquad (4.3.9)$$

где $\hat{g}(t)$ – преобразование Гильберта от $g(t)$. Таким образом, полоса частот ОБП равна половине полосы частот, занимаемой сигналом ДП.

Рассмотренный сигнал цифровой АМ можно интерпретировать как передачу по эквивалентному каналу без несущей. В этом случае сигнал АМ можно представить в виде

$$s_m(t) = A_m g(t), \qquad m = 1,2,\ldots,M. \qquad (4.3.10)$$

Его называют *базовым* (низкочастотным) или *видеосигналом*. Для примера на рис. 4.3.2(*a*) показан четырехуровневый базовый сигнал АМ.

Модулированная по несущей версия этого сигнала дана на рис. 4.3.2 (*b*).

В частном случае $M=2$ рассматриваемая двоичная АМ имеет специальное свойство:

$$s_1(t) = -s_2(t).$$

Следовательно, эти два сигнала имеют одинаковую энергию и коэффициент их взаимной корреляции равен –1. Такие сигналы называют *противоположными*.

Сигналы фазовой модуляции. При цифровой фазовой (нелинейной) модуляции M сигналов можно представить в виде

$$s_m(t) = \mathrm{Re}\, g(t)\, e^{j2\pi(m-1)/M}\, e^{j2\pi f_c t} = g(t)\cos\left[2\pi f_c t + \frac{2\pi(m-1)}{M}\right] =$$

$$g(t)\cos\frac{2\pi(m-1)}{M}\cos 2\pi f_c t - g(t)\sin\frac{2\pi(m-1)}{M}\sin 2\pi f_c t, \qquad (4.3.11)$$

$$m = 1,2,\ldots,M, \qquad 0 \le t \le T,$$

где $g(t)$ определяет огибающую сигнала, а $\theta_m = 2\pi(m-1)/M$, $m = 1,2,\ldots,M$, определяет M возможных значений фазы несущей, которая переносит передаваемую информацию. Цифровую фазовую модуляцию (ФМ) называют также модуляцией с фазовым сдвигом (МФС, PSK).

Заметим, что рассматриваемые формы сигналов имеют одинаковую энергию, т.е.

$$\mathscr{E} = \int_0^T s_m^2(t)\,dt = \tfrac{1}{2}\int_0^T g^2(t)\,dt = \tfrac{1}{2}\mathscr{E}_g. \qquad (4.3.12)$$

Далее, ФМ сигналы можно представить как линейную комбинацию двух ортонормированных сигналов $f_1(t)$ и $f_2(t)$, т.е.

$$s_m(t) = s_{m1} f_1(t) + s_{m2} f_2(t), \qquad (4.3.13)$$

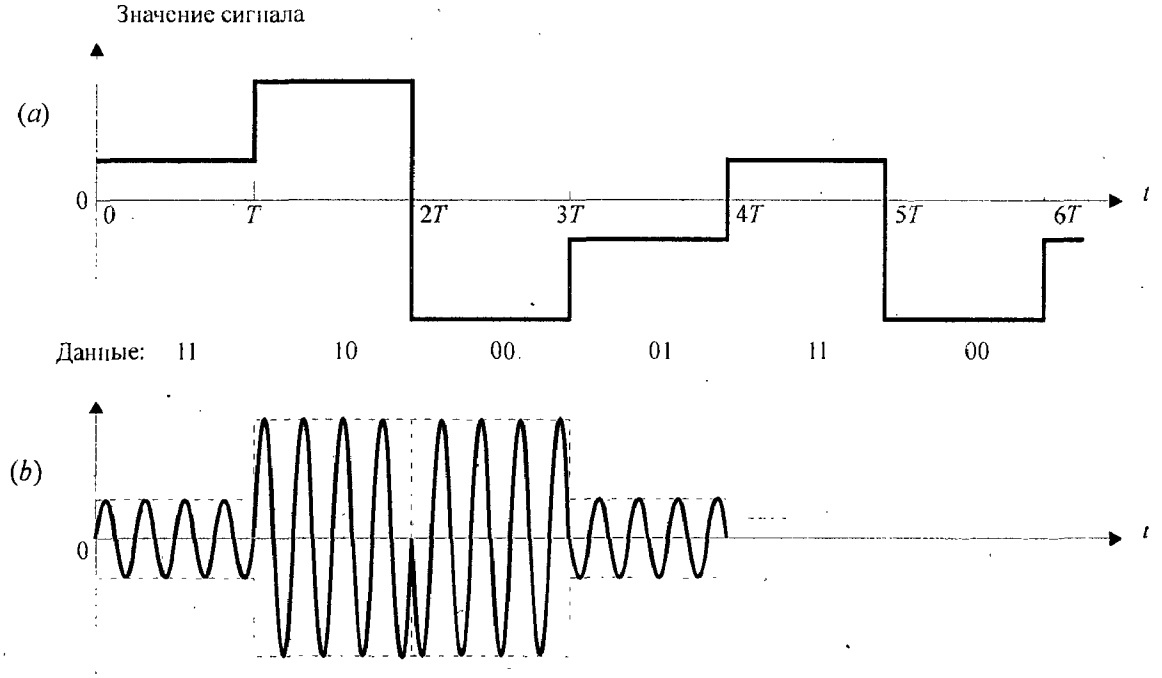


Рис. 4.3.2. Базовый АМ сигнал (видеосигнал) (*a*) и полосовой АМ сигнал (*b*)

где

$$f_1(t) = \sqrt{\frac{2}{\mathscr{E}_g}}\, g(t)\cos 2\pi f_c t, \qquad (4.3.14)$$

$$f_2(t) = -\sqrt{\frac{2}{\mathscr{E}_g}}\, g(t)\sin 2\pi f_c t. \qquad (4.3.15)$$

а двухмерные векторы $\mathbf{s}_m = \left[s_{m1}\ s_{m2}\right]$ определяются так:

$$\mathbf{s}_m = \left[\sqrt{\frac{\mathscr{E}_g}{2}}\cos\frac{2\pi(m-1)}{M}\quad \sqrt{\frac{\mathscr{E}_g}{2}}\sin\frac{2\pi(m-1)}{M}\right], \qquad m = 1,2,\ldots,M. \qquad (4.3.16)$$

Пространственные диаграммы ФМ сигналов для $M = 2, 4$ и 8 даны на рис. 4.3.3.

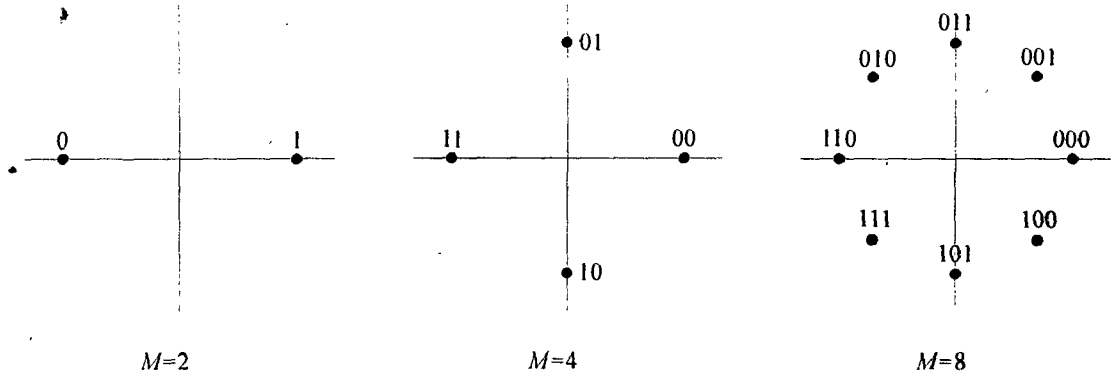


Рис. 4.3.3. Пространственная диаграмма для ФМ сигналов

Видим, что случаю $M=2$ соответствуют одномерные противоположные сигналы, которые идентичны рассмотренным двоичным сигналам АМ.

Как и в случае АМ, отображение или задание k информационных бит в $M=2^k$ возможных значений фаз можно сделать различными путями. Предпочтительное отображение – коды Грея, так что наиболее вероятные ошибки, вызываемые шумами, будут возникать в одном бите k-битового символа.

Евклидово расстояние между точками ФМ сигналов равно

$$d_{mn}^{(e)} = |\mathbf{s}_m - \mathbf{s}_n| = \left\{ \mathscr{E}_g \left[1 - \cos\frac{2\pi}{M}(m-n) \right] \right\}^{1/2} \qquad (4.3.17)$$

Минимальное расстояние по Евклиду соответствует случаю, когда $|m-n|=1$, т.е. соседним значениям фаз. При этом

$$d_{\min}^{(e)} = \sqrt{\mathscr{E}_g \left(1 - \cos\frac{2\pi}{M} \right)}. \qquad (4.3.18)$$

Квадратурная амплитудная модуляция. Хорошую частотную эффективность можно получить не только при АМ/ОБП, но и путём одновременной передачи двух отдельных k-битовых информационных блоков на двух несущих, находящихся в квадратуре ($\cos 2\pi f_c t$ и $\sin 2\pi f_c t$). Такая техника модуляции названа квадратурной АМ или КАМ (QAM), и соответствующие сигналы можно выразить так:

$$s_m(t) = \mathrm{Re}\left[(A_{mc} + jA_{ms}) g(t) e^{j2\pi f_c t} \right] = A_{mc} g(t) \cos 2\pi f_c t - A_{ms} g(t) \sin 2\pi f_c t, \qquad (4.3.19)$$
$$m = 1,2,\ldots,M, \qquad 0 \le t \le T,$$

где A_{mc} и A_{ms} – информационные амплитуды сигнала для квадратурных несущих, а $g(t)$ – форма импульса.

Альтернативно сигнал КАМ можно выразить так:

$$s_m(t) = \mathrm{Re}\left[V_m e^{j\theta_m} g(t) e^{j2\pi f_c t} \right] = V_m g(t) \cos\left(2\pi f_c t + \theta_m \right), \qquad (4.3.20)$$

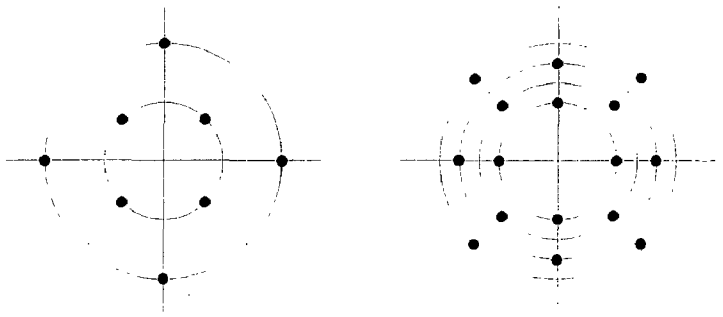
где $V_m = \sqrt{A_{mc}^2 + A_{ms}^2}$ и $\theta_m = \mathrm{arc\,tg}\left(A_{ms} / A_{mc} \right)$. Из этой формы представления видно, что сигнал КАМ можно рассматривать как комбинацию амплитудной и фазовой модуляции.

Действительно, мы можем образовать определенную комбинацию M_1-уровневой АМ и M_2-позиционной ФМ, чтобы сконструировать комбинированное АМ-МФ сигнальное созвездие, содержащее $M = M_1 M_2$ точек пространства сигналов. Если $M_1 = 2^n$ и $M_2 = 2^m$, то сигнальное созвездие комбинированной АМ-ФМ сводится к мгновенной передаче

$m+n=\log M_1 M_2$ двоичных символов, возникающих со скоростью $R/(m+n)$. Примеры сигнальных пространственных диаграмм для комбинированной АМ-МФ показаны на рис. 4.3.4 для M=8 и M=16.

Как в случае АМ сигналов, КАМ сигналы можно представить как линейную комбинацию двух ортонормированных сигналов $f_1(t)$ и $f_2(t)$, т.е.

$$s_m(t) = s_{m1} f_1(t) + s_{m2} f_2(t), \tag{4.3.21}$$



M=8 M=16

Рис. 4.3.4. Примеры пространственных диаграмм для комбинированной АМ-ФМ

где

$$f_1(t) = \sqrt{\frac{2}{\mathscr{E}_g}} g(t)\cos 2\pi f_c t, \quad f_2(t) = -\sqrt{\frac{2}{\mathscr{E}_g}} g(t)\sin 2\pi f_c t \tag{4.3.22}$$

и

$$\mathbf{s}_m = \left[s_{m1} \quad s_{m2}\right] = \left[A_{mc}\sqrt{\tfrac{1}{2}\mathscr{E}_g} \quad A_{ms}\sqrt{\tfrac{1}{2}\mathscr{E}_g}\right]. \tag{4.3.23}$$

Расстояние Евклида между произвольной парой сигнальных векторов равно

$$d_{mn}^{(e)} = \left|\mathbf{s}_m - \mathbf{s}_n\right| = \sqrt{\tfrac{1}{2}\mathscr{E}_g\left[\left(A_{mc} - A_{nc}\right)^2 + \left(A_{ms} - A_{ns}\right)^2\right]}. \tag{4.3.24}$$

Для частного случая, когда амплитуда сигналов принимает ряд дискретных значений $\{(2m-1-M)d, \quad m=1,2,\ldots,M\}$, пространственная диаграмма сигналов является прямоугольной, как показано на рис. 4.3.5.

В этом случае минимальное расстояние Евклида (между смежными точками) равно

$$d_{\min}^{(e)} = d\sqrt{2\mathscr{E}_g}, \tag{4.3.25}$$

что является тем же результатом, что для АМ.

Многомерные сигналы. Из вышесказанного очевидно, что цифровая модуляция несущей по амплитуде и фазе позволяет конструировать сигналы, которые соответствуют двухмерным векторам и пространственным диаграммам сигналов. Если мы хотим сконструировать сигнал, соответствующий вектору большей размерности, можем использовать или временную, или частотную, или обе области для того, чтобы увеличить размерность пространства.

Предположим, что мы имеем N-мерные сигнальные векторы. Для любого N можем разделить интервал времени длины $T_1 = NT$ на N подынтервалов длиной $T = T_1/N$. В каждом интервале длины T можем использовать двоичную АМ (одномерный сигнал), чтобы передать элемент N-мерного сигнального вектора. Таким образом, N временных отрезков используется для передачи N-мерного сигнального вектора.

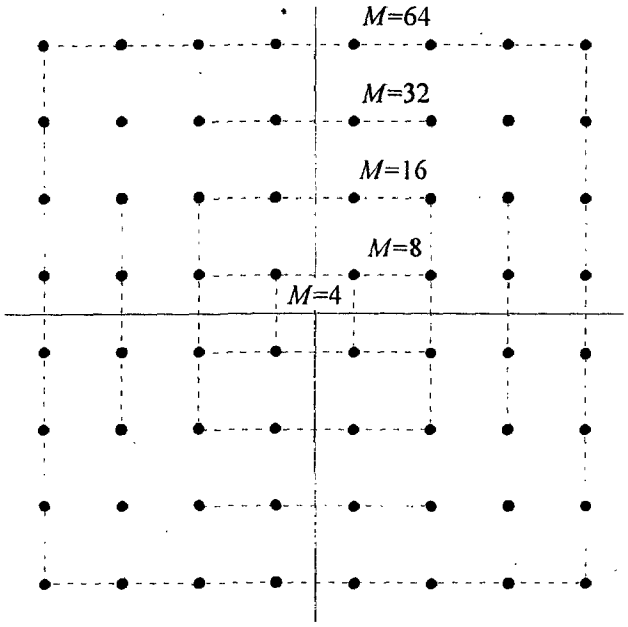


Рис. 4.3.5. Несколько пространственных диаграмм для прямоугольной КАМ

Если N четно, отрезок длиной T можно использовать для мгновенной передачи двух компонент N-мерного вектора путем независимой модуляции амплитуды квадратурных несущих соответствующими компонентами. Таким путем N-мерный сигнальный вектор передается за $\frac{1}{2}NT$ секунд ($\frac{1}{2}N$ временных отрезков).

Альтернативно полоса частот $N\Delta f$ может быть подразделена на N частотных отрезков, каждый шириной Δf. N-мерный сигнальный вектор можно передать через канал путем одновременной (параллельной) модуляции амплитуды N несущих, одна на каждый из N частотных отрезков. Надо позаботиться о достаточном частотном разносе Δf между смежными несущими с тем, чтобы не возникала взаимная интерференция между сигналами на N несущих. Если используются квадратурные несущие на каждом частотном отрезке, то N-мерный вектор (N – чётно) можно передать на $\frac{1}{2}N$ частотных отрезках, что сокращает используемую каналом полосу частот вдвое.

В более общем виде мы можем использовать совместно временную и частотную области для передачи N-мерного сигнального вектора. Например, рис. 4.3.6 иллюстрирует разделение частотно-временной области на 12 ячеек. Таким образом, можно передать N=12-мерный сигнальный вектор при АМ или N=24-мерный сигнальный вектор с использованием двух квадратурных несущих (КАМ) на каждом отрезке.

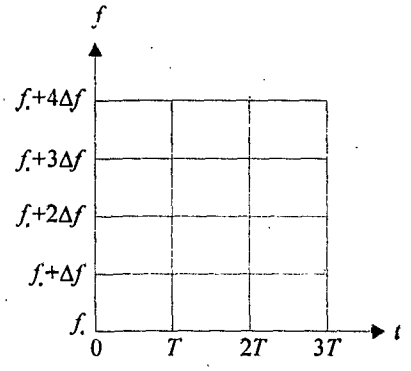


Рис. 4.3.6. Разделение осей времени и частоты на индивидуальные отрезки

Ортогональные многомерные сигналы. Как специальный случай конструирования многомерных сигналов с нелинейной модуляцией рассмотрим случай конструирования M ортогональных сигналов равной энергии, которые различаются по частоте и представлены как

$$s_m(t) = \mathrm{Re}\left[s_{lm}(t)\,e^{j2\pi f_c t}\right] = \sqrt{\frac{2\mathscr{E}}{T}}\cos\left[2\pi f_c t + 2\pi m \Delta f\, t\right], \quad m = 1,2,\ldots,M, \qquad 0 \le t \le T, \tag{4.3.26}$$

где эквивалентный низкочастотный сигнал определяется так:

$$s_{lm}(t) = \sqrt{\frac{2\mathscr{E}}{T}}\,e^{j2\pi m \Delta f t}, \qquad m = 1,2,\ldots,M, \qquad 0 \le t \le T. \tag{4.3.27}$$

Этот вид частотной модуляции (ЧМ) называется *модуляцией частотным сдвигом* (МЧС, FSK).

Эти формы сигналов характеризуются равной энергией и коэффициентами взаимной корреляции

$$\rho_{km} = \frac{2\mathscr{E}/T}{2\mathscr{E}}\int_0^T e^{j2\pi(m-k)\Delta f t}\,dt = \frac{\sin\pi T(m-k)\Delta f}{\pi T(m-k)\Delta f}\,e^{j\pi T(m-k)\Delta f}. \tag{4.3.28}$$

Вещественная часть ρ_{km} равна

$$\rho_r \equiv \mathrm{Re}(\rho_{km}) = \frac{\sin\left[\pi T(m-k)\Delta f\right]}{\pi T(m-k)\Delta f}\cos\left[\pi T(m-k)\Delta f\right] = \frac{\sin\left[2\pi T(m-k)\Delta f\right]}{2\pi T(m-k)\Delta f}. \tag{4.3.29}$$

Сначала заметим, что $\mathrm{Re}(\rho_{km}) = 0$, когда $\Delta f = 1/(2T)$ и $m \ne k$. Поскольку случай $|m-k| = 1$ соответствует соседним частотным интервалам, то $\Delta f = 1/(2T)$ представляет минимальную величину частотного разноса между смежными сигналами для ортогональности M сигналов. Кривые зависимости $\mathrm{Re}(\rho_{km})$ от Δf и $|\rho_{km}|$ от Δf показаны на рис. 4.3.7. Заметим также, что $|\rho_{km}| = 0$, если Δf кратно $1/T$, в то время как $\mathrm{Re}(\rho_{km}) = 0$, когда Δf кратно $1/(2T)$.

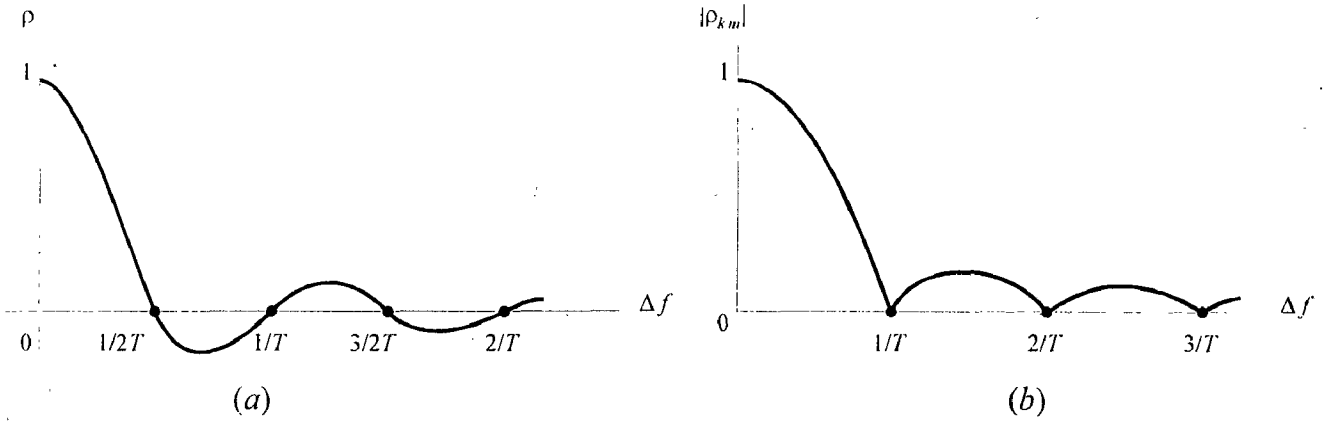


Рис. 4.3.7. Коэффициент взаимной корреляции как функция от частотного разноса для сигналов МЧС

Для случая, когда $\Delta f = 1/(2T)$, ансамбль из M сигналов МЧС эквивалентен N-мерным векторам

$$\begin{aligned}
\mathbf{s}_1 &= \left[\sqrt{\mathcal{E}} \quad 0 \quad 0 \quad \ldots \quad 0 \quad 0\right], \\
\mathbf{s}_2 &= \left[0 \quad \sqrt{\mathcal{E}} \quad 0 \quad \ldots \quad 0 \quad 0\right], \\
&\vdots \; \vdots \; \vdots \; \vdots \; \vdots \; \vdots \\
\mathbf{s}_N &= \left[0 \quad 0 \quad 0 \quad \ldots \quad 0 \quad \sqrt{\mathcal{E}}\right],
\end{aligned} \tag{4.3.30}$$

где N=M. Расстояние между парами сигналов

$$d_{km}^{(e)} = \sqrt{2\mathcal{E}} \qquad \text{для всех } m,k, \tag{4.3.31}$$

что является также минимальным расстоянием. Рисунок 4.3.8 показывает диаграмму пространства сигналов для $M = N = 3$ и $M = N = 2$.

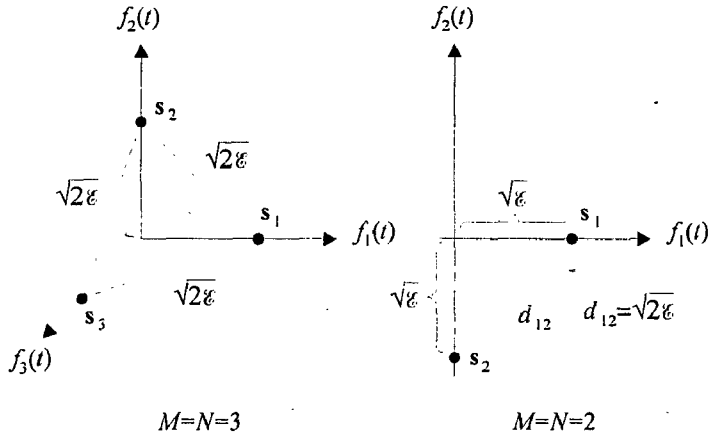


Рис. 4.3.8. Ортогональные сигналы для M=N=3 и для M=N=2

Биортогональные сигналы. Ансамбль из M биортогональных сигналов можно сконструировать из $\frac{1}{2}M$ ортогональных сигналов добавлением к каждому сигналу противоположного сигнала. Таким образом, требуется $N = \frac{1}{2}M$ измерений для конструирования ансамбля из M биортогональных сигналов. Рисунок 4.3.9 иллюстрирует ансамбль биортогональных сигналов для $M = 4$ и $M = 6$.

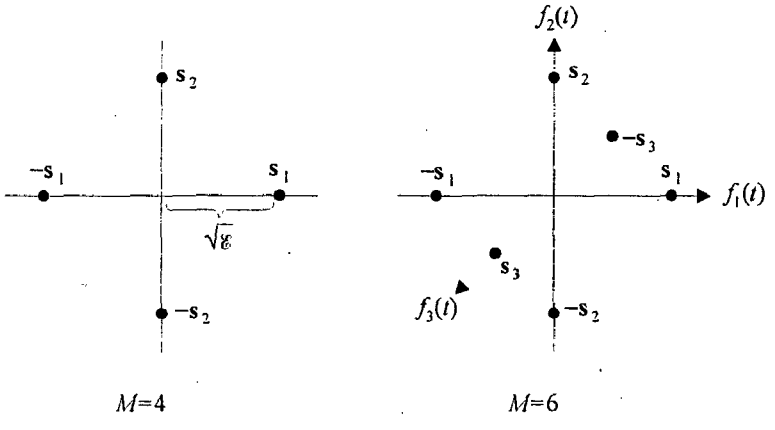


Рис. 4.3.9. Пространственная диаграмма биортогональных сигналов для M=4 и M=6

Заметим, что корреляция между парами сигналов $\rho_r = -1$ или 0. Соответствующие расстояния $d = 2\sqrt{\mathcal{E}}$ или $\sqrt{2\mathcal{E}}$, причём последнее определяет минимальное расстояние.

Симплексные сигналы. Предположим, что имеется ансамбль из M ортогональных сигналов $\{s_m(t)\}$ или, что эквивалентно, векторов $\{\mathbf{s}_m\}$. Их среднее значение

$$\bar{\mathbf{s}} = \frac{1}{M}\sum_{m=1}^{M}\mathbf{s}_m . \qquad (4.3.32)$$

Теперь сконструируем другой ансамбль из M сигналов путём вычитания из каждого ортогонального сигнала среднего значения

$$\mathbf{s}'_m = \mathbf{s}_m - \bar{\mathbf{s}}, \qquad m = 1,2,\ldots,M, \qquad 0 \le t \le T . \qquad (4.3.33)$$

Смысл вычитания сводится к переносу начала координат ансамбля M ортогональных сигналов в точку $\bar{\mathbf{s}}$. Результирующие сигналы называют *симплексными сигналами*, и они имеют следующие свойства. Первое: энергия сигналов равна

$$\left|\mathbf{s}'_m\right|^2 = \left|\mathbf{s}_m - \bar{\mathbf{s}}\right|^2 = \mathscr{E} - \frac{2}{M}\mathscr{E} + \frac{1}{M}\mathscr{E} = \mathscr{E}\left(1 - \frac{1}{M}\right). \qquad (4.3.34)$$

Второе: взаимная корреляция для любой пары сигналов одинакова и равна

$$\mathrm{Re}\left(\rho_{mn}\right) = \frac{\mathbf{s}'_m \cdot \mathbf{s}'_n}{\left|\mathbf{s}'_m\right|\left|\mathbf{s}'_n\right|} = \frac{-1/M}{1-1/M} = -\frac{1}{M-1}. \qquad (4.3.35)$$

Следовательно, для всех m,n ансамбль сигналов *одинаково коррелирован* и требует меньшей энергии по сравнению с ортогональным ансамблем (коэффициент ослабления $1-1/M$). Так как была перемещена только точка начала координат сигналов, расстояние между любой парой сигналов сохраняется равным $d = \sqrt{2\mathscr{E}}$, что равно расстоянию для пары сигналов ортогонального ансамбля. Рисунок 4.3.10 иллюстрирует симплексные сигналы при $M = 2$, 3 и 4. Заметим что размерность пространства сигналов $N = M - 1$.

Формы сигналов для двоичных кодов. Ансамбль из M сигналов может быть создан ансамблем M двоичных кодовых слов вида

$$\mathbf{C}_m = \left[c_{m1} \quad c_{m2} \ldots c_{mN}\right], \qquad m = 1,2,\ldots,M, \qquad (4.3.36)$$

где $c_{mj} = 0$ или 1 для всех m и j.

Каждая компонента кодового слова отображается в элементарный двоичный сигнал ФМ:

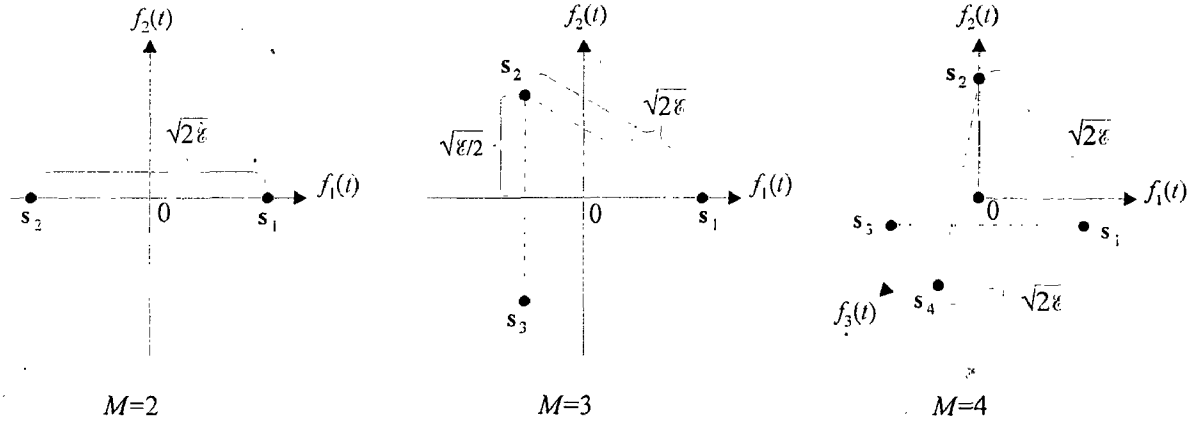


Рис. 4.3.10. Пространственные диаграммы сигналов для M-мерного симплекса

$$\begin{aligned} c_{mj} = 1 &\Rightarrow s_{mj}(t) = \sqrt{\tfrac{2\mathscr{E}_c}{T_c}}\cos 2\pi f_c t \quad (0 \le t \le T), \\ c_{mj} = 0 &\Rightarrow s_{mj}(t) = -\sqrt{\tfrac{2\mathscr{E}_c}{T_c}}\cos 2\pi f_c t \quad (0 \le t \le T), \end{aligned} \qquad (4.3.37)$$

где $T_c = T/N$ и $\mathscr{E}_c = \mathscr{E}/N$.

Таким образом, M кодовых слов $\{C_m\}$ отображаются ансамблем из M сигналов $\{s_m(t)\}$.

Сигналы можно представить в векторной форме так:

$$\mathbf{s}_m = \left[s_{m1} \quad s_{m2} \ldots s_{mN}\right], \quad m = 1,2,\ldots,M, \tag{4.3.38}$$

где $s_{mj} = \pm\sqrt{\mathscr{E}/N}$ для всех m и j. N называют блоковой длиной кода, оно также определяет размерность M сигналов.

Отметим, что имеются 2^N возможных сигналов, которые можно сконструировать посредством 2^N возможных двоичных кодовых слов. Мы можем выбрать $M < 2^N$ сигналов для передачи информации. Мы также отметим, что 2^N возможных сигнальных точек соответствуют вершинам N-мерного гиперкуба с центром в начале координат. Рисунок 4.3.11 иллюстрирует сигнальные точки для случая размерности $N = 2$ и 3.

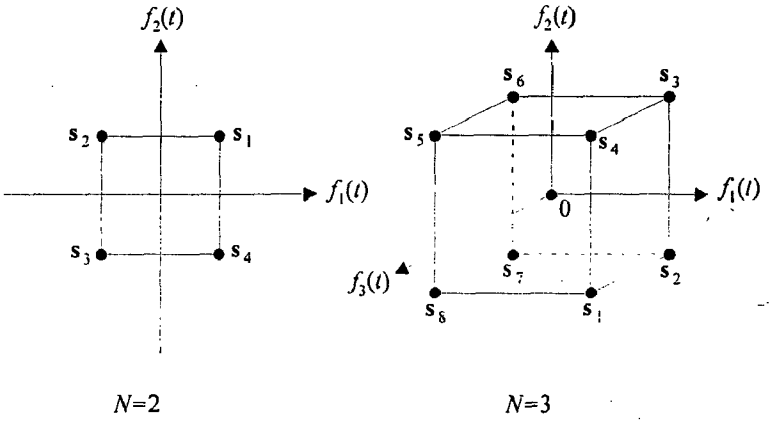


Рис. 4.3.11. Пространственные диаграммы сигналов, создаваемых двоичными кодами

Каждый из M сигналов имеет энергию $\mathscr{E}$. Взаимная корреляция между каждой парой сигналов зависит от того, как мы выбрали M сигналов из 2^N возможных. Эта тема обсуждается в гл. 7. Ясно, что соседние сигнальные точки имеют коэффициент взаимной корреляции

$$\rho_r = \frac{\mathscr{E}(1-2/N)}{\mathscr{E}} = \frac{N-2}{N} \tag{4.3.39}$$

и соответствующее расстояние

$$d^{(e)} = \sqrt{2\mathscr{E}\left(1-\rho_r\right)} = \sqrt{4\mathscr{E}/N}\,. \tag{4.3.40}$$

Этим заканчиваем наше обсуждение сигналов цифровой модуляции без памяти.

4.3.2. Линейная модуляция с памятью

Сигналы модуляции, которые обсуждались в предыдущем разделе, были классифицированы как сигналы без памяти, поскольку не было зависимости между сигналами, которые передаются на неперекрывающихся символьных интервалах. В этом разделе мы представим некоторые сигналы модуляции, в которых имеется зависимость между сигналами, которые передаются в последовательных символьных интервалах. Эта зависимость сигналов обычно вводится с целью такого формирования спектра передаваемых символов, который был бы согласован со спектральными характеристиками канала. Зависимость между сигналами, передаваемыми в различных символьных интервалах, обычно используется при кодировании данных источника на входе модулятора при помощи *кодированной модуляции*, как это описывается в гл. 9.

В этом разделе мы хотим представить примеры модулированных сигналов с памятью и характеризовать их память в терминах цепей Маркова. Мы ограничим рассмотрение базовыми (низкочастотными) сигналами. Обобщение на полосовые сигналы относительно несложно.

Рисунок 4.3.12 иллюстрирует три различных базовых сигнала и соответствующие им данные источника информации. Первый сигнал, называемый NRZ[1] (non return to zero – двоичный сигнал без возвращения к нулевому уровню – ДБН), – простейший. Двоичный информационный символ 1 представлен прямоугольным импульсом положительной полярности A, а двоичный информационный символ 0 – прямоугольным импульсом отрицательной полярности $-A$.

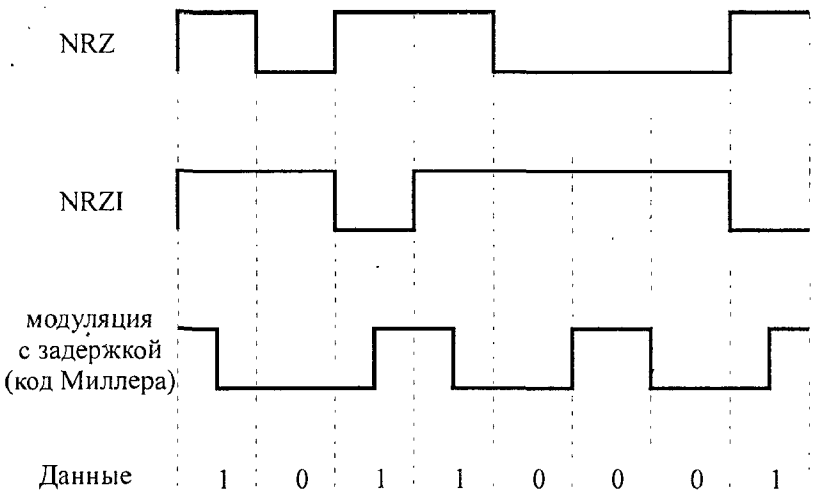


Рис. 4.3.12. Базовые сигналы

Следовательно, NRZ отображает модуляцию без памяти и она эквивалентна двоичной АМ или двоичной ФМ в системе с модулированной несущей. Сигнал NRZI[2] отличается от NRZ тем, что переход от одного уровня амплитуды к другому имеет место только при передаче 1. Уровень амплитуды не меняется, когда передаётся символ 0. Этот тип преобразования сигнала назван *дифференциальным кодированием*. Операция кодирования математически описывается соотношением

$$b_k = a_k \oplus b_{k-1}, \qquad (4.3.41)$$

где $\{a_k\}$ – двоичная информационная последовательность источника на входе кодера, $\{b_k\}$ – выходная последовательность кодера, а $\oplus$ означает суммирование по модулю 2. Если $b_k = 1$, передаваемый сигнал – это прямоугольный импульс с амплитудой A, а если $b_k = 0$, передаваемый сигнал – прямоугольный импульс с амплитудой $-A$. Следовательно, выход источника отображается одним из двух сигналов, точно таких, как в NRZ-сигнале.

Операция дифференциального кодирования вводит память в сигнал. Комбинацию операций кодера и модулятора можно представить диаграммой состояний (марковская цепь), как показано на рис. 4.3.13. В состоянии S_1 передаётся сигнал $s(t)$, в состоянии S_0 – сигнал $-s(t)$. Указание $a_k / \pm s(t)$ означает передачу сигнала $s(t)$ или $-s(t)$ на следующем шаге при поступлении очередного информационного символа a_k. Диаграмму можно описать с помощью двух матриц перехода, соответствующих двум возможным входным

[1] Его можно назвать двоичным (двухполярным) сигналом без памяти (прп).
[2] Его можно назвать двоичным (двухполярным) сигналом с памятью (прп).

символам $\{0,1\}$. Заметим, что если $a_k = 0$, состояние кодера не меняется. Следовательно, матрица перехода для $a_k = 0$ простейшая:

$$\mathbf{T}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \tag{4.3.42}$$

где $t_{ij} = 1$, если a_k появляется при переходе от состояния i к состоянию j, $i = 1, 2$ и $j = 1, 2$; в других случаях $t_{ij} = 0$. Аналогично, матрица перехода состояний для $a_k = 1$ равна:

$$\mathbf{T}_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \tag{4.3.43}$$

Таким образом, эти две матрицы перехода состояний характеризуют NRZI сигнал.

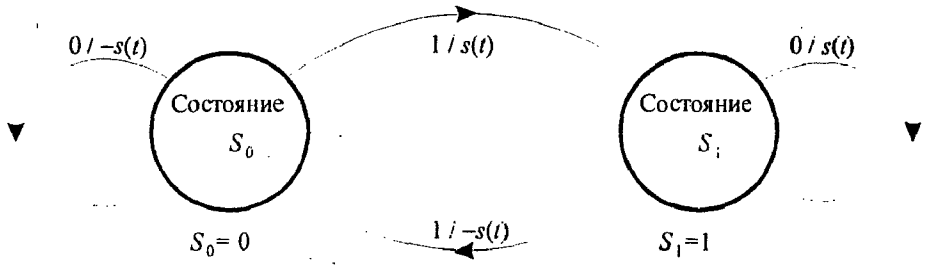


Рис. 4.3.13. Диаграмма состояний для сигнала NRZI

Другой путь появления памяти, вводимый докодерной обработкой, иллюстрируется посредством решётчатой диаграммы. Решётчатая диаграмма NRZI-сигнала показана на рис. 4.3.14. Решётка обеспечивает такую же информацию относительно зависимости сигнала, как диаграмма состояний, но она также отображает эволюцию во времени переходов состояний. Состояния системы отмечены на решётке точками (узлами), а на переходах между ними (называемых ветвями) отмечены поступающие информационные символы и передаваемые сигналы.

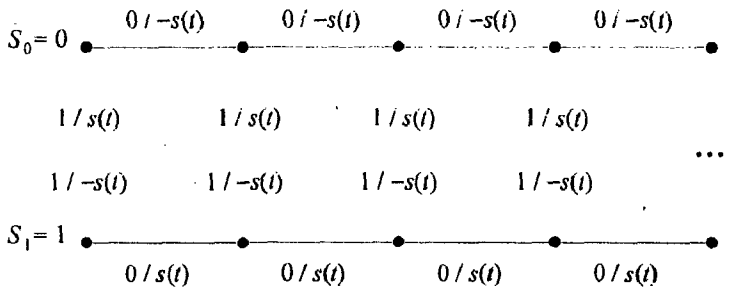


Рис. 4.3.14. Решетчатая диаграмма для сигнала NRZI

Сигнал, образованный модуляцией с задержкой, также имеет память. Как будет показано в гл. 9, модуляция с задержкой эквивалентна кодированию источника данных посредством кода Миллера, использующего NRZI для передачи кодированных данных. Этот вид цифровой модуляции широко используется для цифровой магнитной записи и в системах модуляции несущей с ФМ. Сигнал может быть описан диаграммой состояний, которая имеет четыре состояния, как показано на рис. 4.3.15 (*a*). Имеются два элементарных сигнала $s_1(t)$ и $s_2(t)$ и их негативы $-s_1(t)$ и $-s_2(t)$, которые используются для передачи двоичной информации. Эти сигналы иллюстрируются на рис. 4.3.15 (*b*).

Отображение символов соответствующими сигналами иллюстрируется диаграммой состояний. Матрицу перехода состояний, которая характеризует память этих методов кодирования и модуляции, легко получить из диаграммы состояний рис. 4.3.15.

Когда $a_k = 0$, мы имеем для матрицы перехода

$$\mathbf{T}_1 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}. \qquad (4.3.44)$$

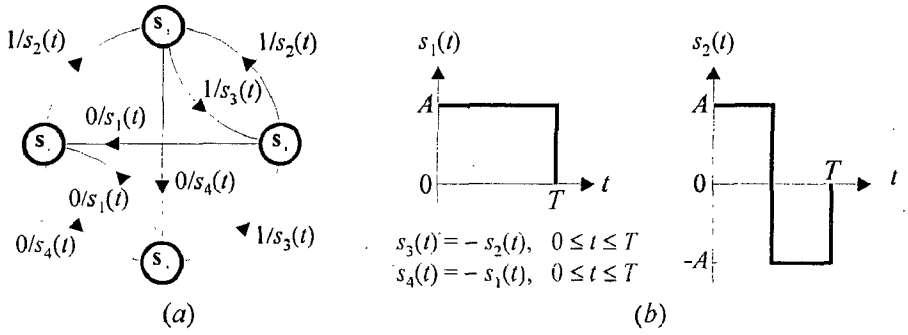


Рис. 4.3.15. Диаграмма состояний (*a*) и базовые формы сигналов (*b*) для модуляции с задержкой (код Миллера)

Когда $a_k = 1$, матрица перехода равна

$$\mathbf{T}_2 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \qquad (4.3.45)$$

Таким образом, эти две матрицы перехода состояний размером 4×4 характеризуют диаграмму состояний сигналов, кодированных по Миллеру.

Разновидности модуляции с памятью, такие как NRZI и кодирование по Миллеру, в общем характеризуются K-мерной марковской цепью со стационарными вероятностями состояний $\{p_i, i = 1,2,\ldots,K\}$ и вероятностями перехода $\{p_{ij}, i,j = 1,2,\ldots,K\}$. С каждым переходом связан сигнал $\{s_j(t), j = 1,2,\ldots,K\}$. Таким образом, p_{ij} означает вероятность того, что передаётся сигнал $s_j(t)$ на данном сигнальном интервале после передачи сигнала $s_i(t)$ на предыдущем сигнальном интервале. Вероятности переходов могут быть упорядочены в форме матрицы

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1K} \\ p_{21} & p_{22} & \cdots & p_{2K} \\ \vdots & \vdots & \cdots & \vdots \\ p_{K1} & p_{K2} & \cdots & p_{KK} \end{bmatrix}. \qquad (4.3.46)$$

Её называют матрицей переходных вероятностей. Матрица переходных вероятностей $\mathbf{P}$ легко получается из матриц переходов $\{\mathbf{T}_i\}$ и соответствующих вероятностей появления входных символов (или, что эквивалентно, стационарных вероятностей состояний $\{p_i\}$). Общее соотношение можно выразить так:

$$\mathbf{P} = \sum_{i=1}^{2} q_i \mathbf{T}_i , \qquad (4.3.47)$$

где $q_1 = P(a_k = 0)$ и $q_2 = P(a_k = 1)$.

Для сигнала NRZI с равновероятными состояниями $p_1 = p_2 = \frac{1}{2}$ и матриц перехода, определяемых (4.3.42) и (4.3.42), матрица переходных вероятностей равна

$$\mathbf{P} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}. \qquad (4.3.48)$$

Аналогично матрица переходных вероятностей для сигналов, кодированных по Миллеру, при равновероятных символах ($q_1 = q_2 = \frac{1}{2}$, или, что эквивалентно, $p_1 = p_2 = p_3 = p_4 = \frac{1}{4}$) равна

$$\mathbf{P} = \begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{bmatrix}. \qquad (4.3.49)$$

Матрица переходных вероятностей используется при определении спектральных характеристик техники цифровой модуляции с памятью, как мы покажем в разд. 4.4.

4.3.3. Нелинейные методы модуляции с памятью

В этом разделе мы рассмотрим класс методов цифровой модуляции, в которых фаза сигнала поддерживается непрерывной. Такая поддержка приводит к модуляции по фазе или по частоте с памятью. Метод модуляции нелинеен.

Частотная модуляция с непрерывной фазой (ЧМНФ). Обычный сигнал ЧМ или модуляции с частотным сдвигом (МЧС или FSK) генерируется путём сдвига частоты несущей на величину $f_n = \frac{1}{2}\Delta f I_n$, $I_n = \pm 1, \pm 3, \ldots, \pm(M-1)$, чтобы отразить цифровую информацию, которую надо передать. Этот вид модуляции сигналов был описан в разделе 4.3.1 и он без памяти. Переход от одной частоты к другой может быть выполнен посредством $M = 2^k$ отдельных генераторов, настроенных на необходимые частоты, и выбора одной из M частот согласно частному значению k-битового символа (блока), который должен быть передан на сигнальном интервале длиной $T = k/R$ секунд.

Однако такое резкое переключение с выхода одного генератора на выход другого в смежных сигнальных интервалах приводит к относительно большим долям боковых частотных составляющих вне основной спектральной полосы сигнала, и, следовательно, этот метод требует большую полосу частот для передачи сигнала. Чтобы избежать использования сигналов с большими долями боковых полос, информационный сигнал может модулировать одну несущую, частота которая меняется непрерывно. Результирующий частотно-модулированный сигнал имеет в этом случае непрерывную фазу и поэтому назван ЧМ с непрерывной фазой (ЧМНФ, CPFSK). Этот вид ЧМ сигнала имеет память, обусловленную тем, что фазу несущей заставляют быть непрерывной.

Чтобы представить сигнал ЧМНФ, мы начнём с сигнала АМ

$$d(t) = \sum_n I_n g(t - nT), \qquad (4.3.50)$$

где $\{I_n\}$ означает последовательность амплитуд, полученную путём отображения k-битовых блоков двоичных символов от информационных последовательностей $\{a_n\}$ в уровни амплитуды $\pm 1, \pm 3, \ldots, \pm(M-1)$, $g(t)$ – прямоугольный импульс с амплитудой $1/2T$ и

длительностью T секунд. Сигнал $d(t)$ используется для частотной модуляции несущей. Следовательно, эквивалентный низкочастотный сигнал $\upsilon(t)$ можно в этом случае выразить так:

$$\upsilon(t)=\sqrt{\frac{2\mathscr{E}}{T}}\exp\left\{j\left[4\pi Tf_d\int_{-\infty}^{t}d(\tau)\,d\tau+\phi_0\right]\right\}, \qquad (4.3.51)$$

где f_d – максимальная девиация частоты, а ϕ_0 – начальная фаза несущей. Частотно-модулированный сигнал, соответствующий (4.3.51), можно выразить как

$$s(t)=\sqrt{\frac{2\mathscr{E}}{T}}\cos\left[2\pi f_c t+\phi(t;\mathbf{I})+\phi_0\right], \qquad (4.3.52)$$

где $\phi(t;\mathbf{I})$ представляет меняющуюся во времени фазу несущей:

$$\phi(t;\mathbf{I})=4\pi Tf_d\int_{-\infty}^{t}d(\tau)\,d\tau=4\pi Tf_d\int_{-\infty}^{t}\left[\sum_n I_n g(\tau-nT)\right]d\tau. \qquad (4.3.53)$$

Заметим, что $d(t)$ содержит разрывы, интеграл же от $d(t)$ непрерывен. Следовательно, мы имеем сигнал с непрерывной фазой. Фаза несущей на интервале $nT\le t\le(n+1)T$ определяется интегрированием (4.3.53) таким образом:

$$\phi(t;\mathbf{I})=2\pi f_d T\sum_{k=-\infty}^{n-1}I_k+2\pi f_d q(t-nT)I_n=\theta_n+2\pi hI_n q(t-nT), \qquad (4.3.54)$$

где h, θ_n и $q(t)$ определяются так:

$$h=2f_dT, \qquad (4.3.55)$$

$$\theta_n=\pi h\sum_{k=-\infty}^{n-1}I_k, \qquad (4.3.56)$$

$$q(t)=\begin{cases}0 & (t<0),\\ t/2T & (0\le t<T),\\ \frac{1}{2} & (t>T).\end{cases} \qquad (4.3.57)$$

Видим, что θ_n представляет накопление (память) от всех информационных символов, переданных до момента $(n-1)T$. Параметр h называется индексом модуляции.

Модуляция с непрерывной фазой (МНФ). Выраженный в виде (4.3.54) сигнал ЧМНФ является специальным случаем общего класса сигналов МНФ (модуляции с непрерывной фазой), в которой фаза несущей определяется так:

$$\phi(t;\mathbf{I})=2\pi\sum_{k=-\infty}^{n-1}I_k h_k q(t-kT),\qquad nT\le t\le(n+1)T, \qquad (4.3.58)$$

где $\{I_k\}$ – последовательность информационных символов, выбранных из алфавита $\pm1,\pm3,\ldots,\pm(M-1)$, $\{h_k\}$ – последовательность индексов модуляции, а $q(t)$ – нормированная огибающая сигнала.

Когда $h_k=h$ для всех k, индекс модуляции фиксирован для всех символов. Когда индекс модуляции меняется от одного символа к другому, сигнал МНФ называется *многоиндексным* (*multi-h*). В этом случае $\{h_k\}$ меняется циклически, принимая значения ряда индексов. Форму сигнала $q(t)$ можно представить в общем виде как интеграл от импульса $g(t)$:

$$q(t)=\int_0^t g(\tau)\,d\tau. \qquad (4.3.59)$$

Если $g(t)=0$ для $t>T$, сигнал МНФ называют МНФ с полным откликом. Если $g(t)\neq 0$ для $t>T$, модулированный сигнал называют МНФ с частичным (парциальным) откликом. Рисунок 4.3.16 иллюстрирует несколько форм огибающих импульсов $g(t)$ и соответствующих форм $q(t)$. Очевидно, что неограниченное число разновидностей сигналов МНФ можно генерировать выбором различных огибающих импульсов $g(t)$ и изменением индекса модуляции h и размера алфавита M.

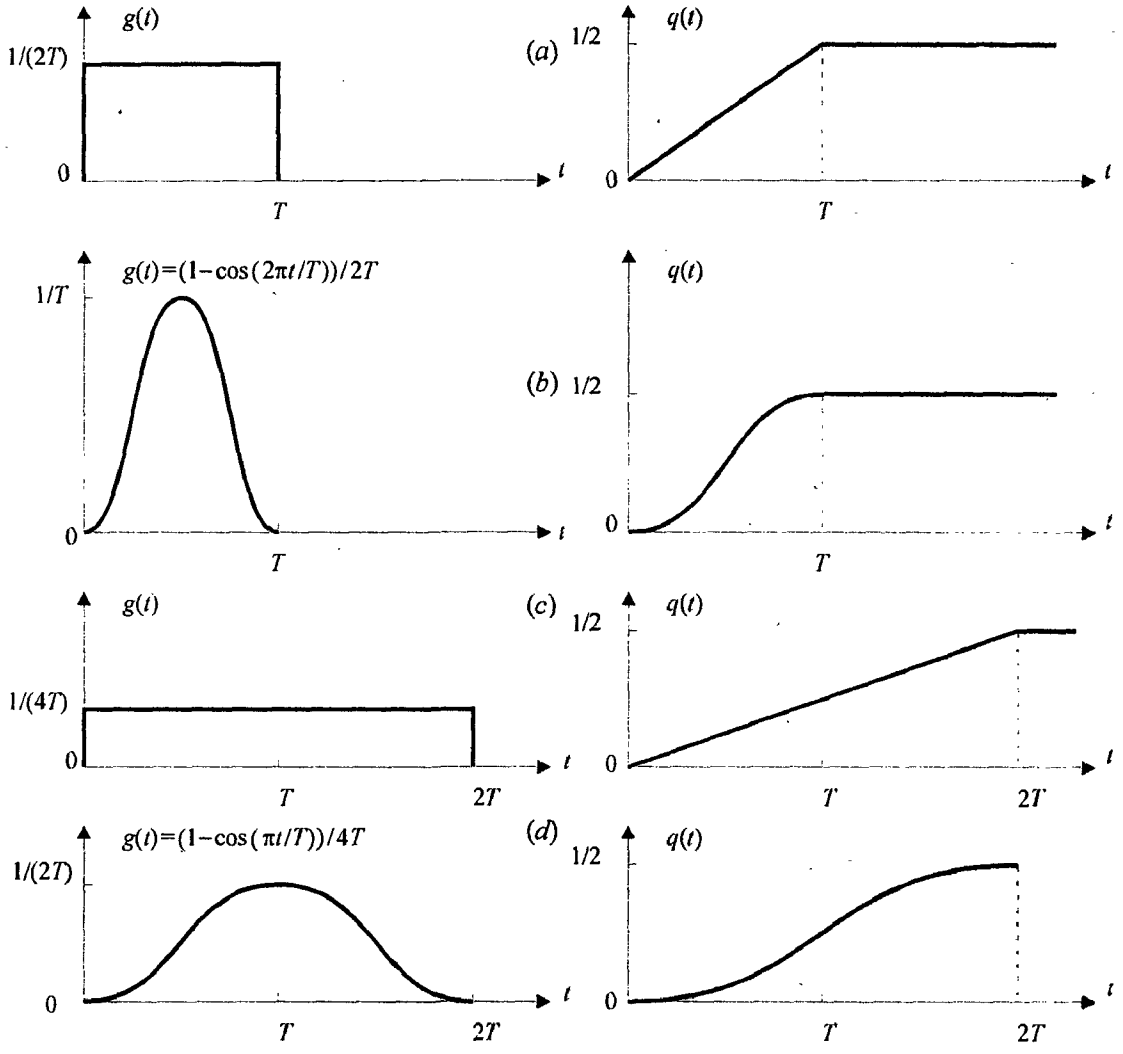


Рис. 4.3.16. Формы импульсов для полного отклика МНФ (*a*, *b*) и для парциального отклика МНФ (*c*, *d*)

Поучительно нарисовать ряд фазовых траекторий $\phi(t;\mathbf{I})$, генерируемых возможными значениями информационных последовательностей $\{I_n\}$. Например, для случая ЧМНФ с двоичными символами $I_n=\pm 1$ и прямоугольным импульсом ряд фазовых траекторий, начинающихся при времени $t=0$, показан на рис. 4.3.17. Для сравнения фазовые траектории для четырёхпозиционной ЧМНФ иллюстрируются на рис. 4.3.18. Эти фазовые диаграммы называют *фазовым деревом*. Видим, что фазовые деревья для ЧМНФ являются кусочно-линейными, как следствие того факта, что импульс $g(t)$ прямоугольный. Гладкие

фазовые траектории и фазовые деревья получены также при использовании импульсов, которые не имеют разрывов, таких как класс импульсов приподнятого косинуса. Для примера фазовая траектория, генерируемая последовательностью $(1,-1,-1,-1,\ 1,\ 1,-1,\ 1)$, при парциальном отклике импульса приподнятого косинуса длины $3T$ иллюстрируется на рис. 4.3.19. Для сравнения показаны соответствующие фазовые траектории, генерируемые при ЧМНФ.

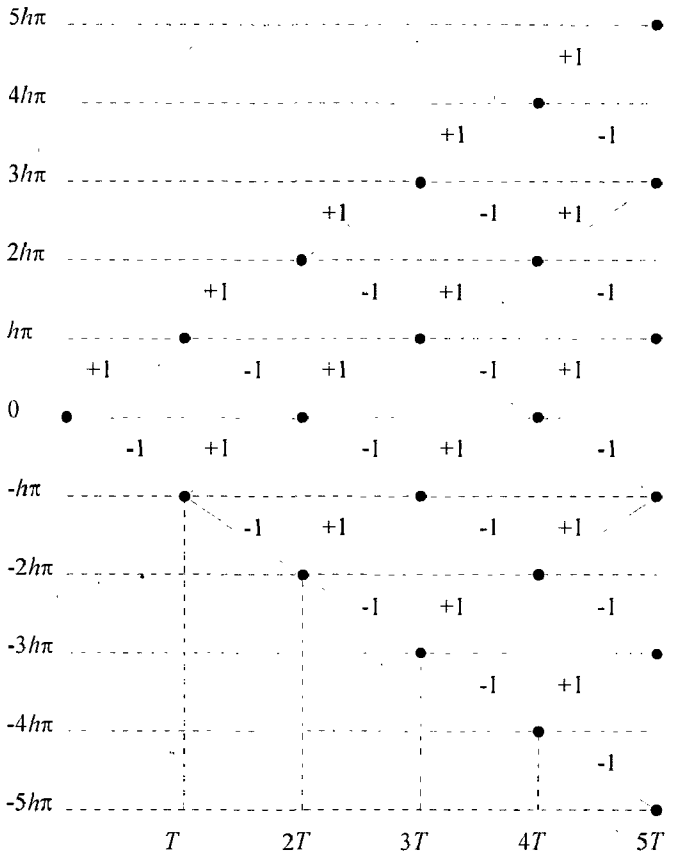


Рис. 4.3.17. Фазовые траектории для двоичной ЧМНФ

Фазовые деревья, показанные на этих рисунках, растут со временем. Однако фаза несущей однозначна только в области от $\phi = 0$ до $\phi = 2\pi$ или, что эквивалентно, от $\phi = -\pi$ до $\phi = \pi$. Если фазовые траектории определить по модулю 2π, скажем в области $(-\pi, \pi)$, фазовое дерево превратится в структуру, называемую фазовой решёткой. Для надлежащего обозрения диаграмм фазовых решёток можем строить две квадратурные компоненты $x_c(t;\mathbf{I}) = \cos\phi(t;\mathbf{I})$ и $x_s(t;\mathbf{I}) = \sin\phi(t;\mathbf{I})$ как функции времени. Таким образом, мы генерируем трёхмерный график, в котором квадратурные компоненты x_c и x_s возникают на поверхности цилиндра единичного радиуса. Например, рис. 4.3.20 иллюстрирует фазовую решётку или фазовый цилиндр, получающийся при двоичной модуляции с индексом модуляции $h = \frac{1}{2}$ и использовании импульса принятого косинуса длиной $3T$.

Простое представление фазовых траекторий можно получить, показывая только финальные значения фаз сигнала в моменты времени $t = nT$. В этом случае мы ограничиваем индекс модуляции сигнала МНФ рациональными значениями. В частности, предположим, что $h = m/p$, где m и p – взаимно простые целые числа. Тогда МНФ-сигнал с полным откликом в моменты времени $t = nT$ будет иметь финальные состояния

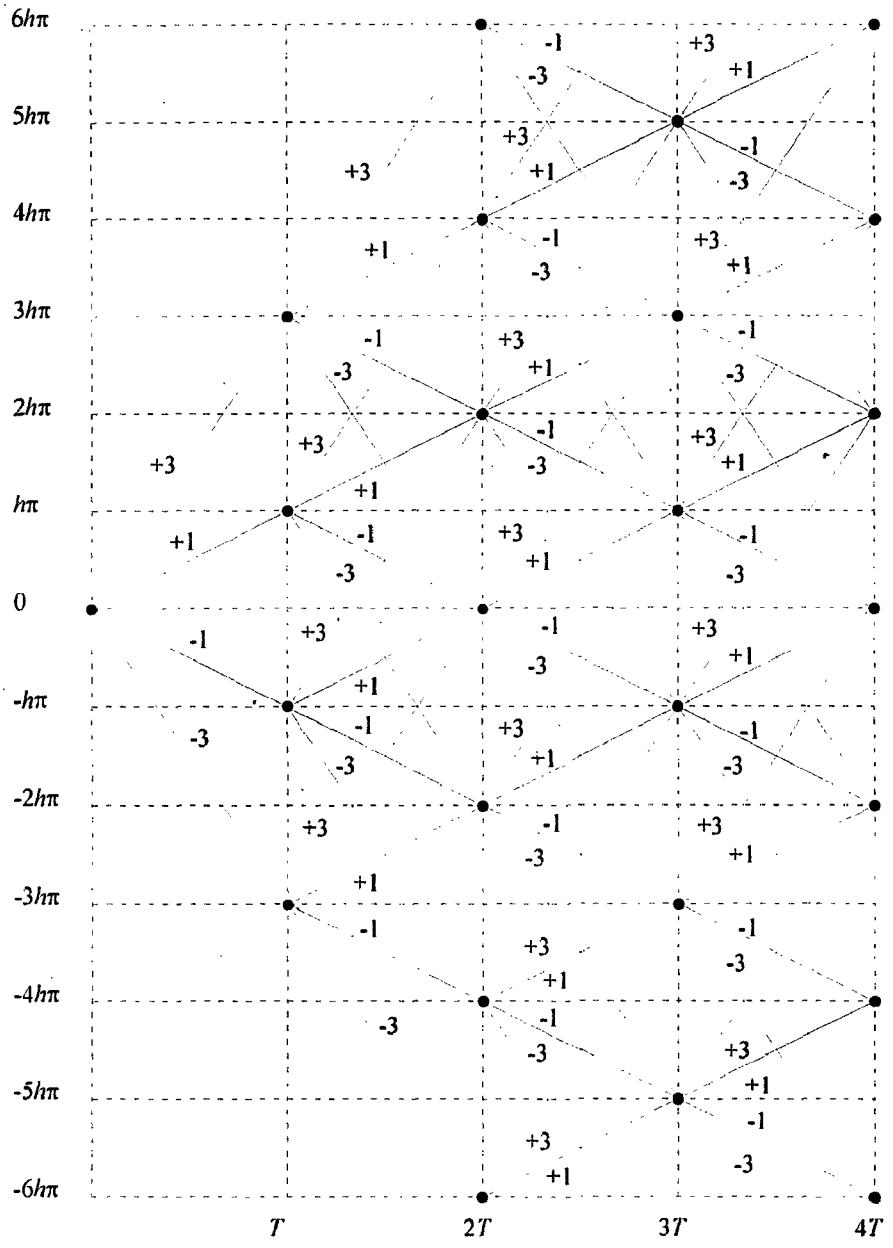


Рис. 4.3.18. Фазовые траектории для четырёхпозиционной ЧМНФ

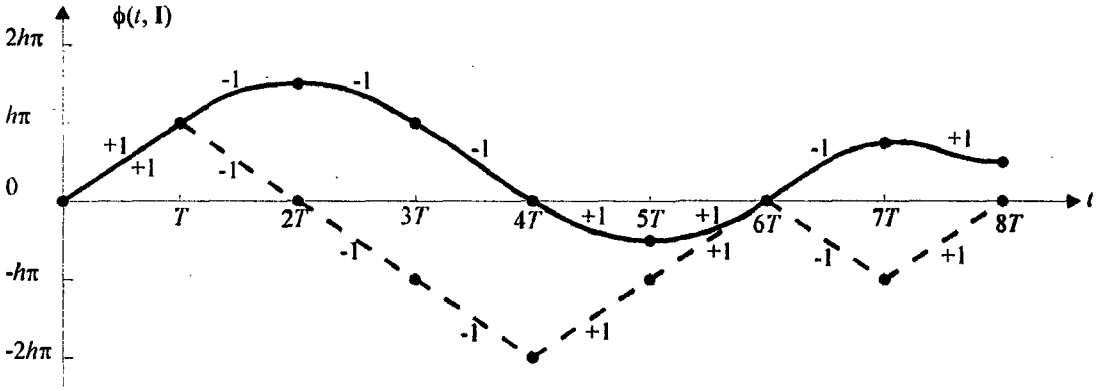


Рис. 4.3.19. Фазовые траектории для двоичной ЧМНФ (штриховой линией)) и двоичной МНФ с парциальным откликом, основанным на импульсе приподнятого косинуса длительностью $3T$ [*Sundberg* (1986), © 1986 *IEEE*]

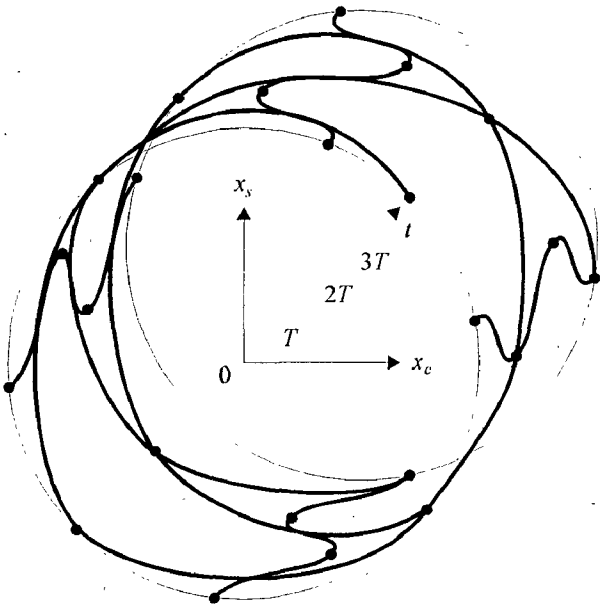


Рис. 4.3.20. Фазовый цилиндр для двоичной МНФ с h = 1/2 и с импульсом приподнятого косинуса длительностью $3T$
[*Sundberg* (1986), © 1986 *IEEE*]

$$\Theta_s = \left\{0, \tfrac{\pi m}{p}, \tfrac{2\pi m}{p}, \ldots, \tfrac{(p-1)\pi m}{p}\right\}, \tag{4.3.60}$$

когда m – чётно, и

$$\Theta_s = \left\{0, \tfrac{\pi m}{p}, \tfrac{2\pi m}{p}, \ldots, \tfrac{(2p-1)\pi m}{p}\right\}, \tag{4.3.61}$$

когда m – нечётно. Следовательно, имеется p финальных состояний фазы, когда m – чётно, и $2p$ состояний, когда m – нечётно. С другой стороны, когда огибающая импульса простирается на L символьных интервалах (МНФ с парциальным откликом), число состояний фазы может увеличиваться до максимального значения S_t, где

$$S_t = \begin{cases} pM^{L-1} & (m \text{ чётное}), \\ 2pM^{L-1} & (m \text{ нечётное}), \end{cases} \tag{4.3.62}$$

где M – объём алфавита. Например, двоичная ЧМНФ (полный отклик, прямоугольный импульс) с $h = \frac{1}{2}$ имеет $S_t = 4$ финальных фазовых состояний. *Решётка состояний* для этого сигнала показана на рис. 4.3.21. Подчеркнём, что переход фазы из одного состояния в другое не затрагивает промежуточные фазовые траектории. Они здесь представляют фазовые переходы для состояний в моменты времени $t = nT$.

Альтернативной по отношению к решётке состояний является диаграмма состояний, которая иллюстрирует переходы состояний в моменты времени $t = nT$. Она даже является более компактным представлением сигнальных характеристик МНФ. Только возможные финальные состояния фазы и их переходы отражены на диаграмме состояний. Время здесь не выступает как переменная. Для примера на рис. 4.3.22 показана диаграмма состояний для сигнала ЧМНФ с h=1/2.

Модуляция с минимальным сдвигом (ММС, MSK). ММС – специальная форма двоичной ЧМНФ (и, следовательно, МНФ), в которой индекс модуляции $h = \frac{1}{2}$. Фаза несущей на интервале $nT \le t \le (n+1)T$ равна

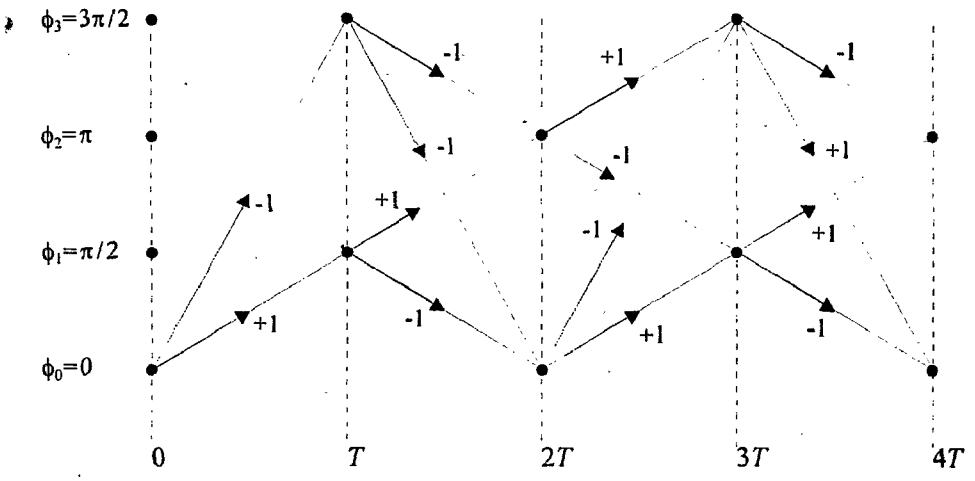


Рис. 4.3.21. Решетка состояний для двоичной ЧМНФ с h=1/2

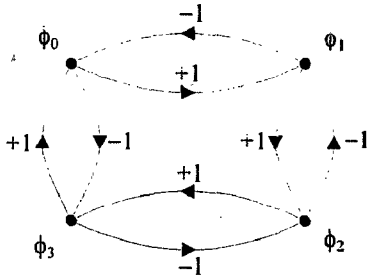


Рис. 4.3.22. Диаграмма состояний для двоичной ЧМНФ с h=1/2

$$\phi(t;\mathbf{I}) = \tfrac{1}{2}\pi \sum_{k=-\infty}^{n-1} I_k + \pi I_n q(t-kT) = \\ = \theta_n + \tfrac{1}{2}\pi I_n \left(\frac{t-nT}{T}\right), \qquad nT \le t \le (n+1)T, \tag{4.3.63}$$

а сигнал модулированной несущей равен

$$s(t) = A\cos\left[2\pi f_c t + \theta_n + \tfrac{1}{2}\pi I_n\left(\frac{t-nT}{T}\right)\right] = \\ = A\cos\left[2\pi\left(f_c + \frac{I_n}{4T}\right)t - \tfrac{1}{2}n\pi I_n + \theta_n\right], \qquad nT \le t \le (n+1)T. \tag{4.3.64}$$

Формула (4.3.64) указывает на то, что сигнал двоичной ЧМНФ может быть выражен как синусоида, имеющая одно из двух возможных значений частоты на интервале $nT \le t \le (n+1)T$.. Если мы определим эти частоты так:

$$f_1 = f_c - \frac{1}{4T}, \\ f_2 = f_c + \frac{1}{4T}, \tag{4.3.65}$$

тогда сигнал двоичной ЧМНФ, определяемый (4.3.64), можно записать в виде

$$s_i(t) = A\cos\left[2\pi f_i t + \theta_n + \tfrac{1}{2}n\pi(-1)^{i-1}\right], \quad i = 1, 2. \tag{4.3.66}$$

Разность частот $\Delta f = f_2 - f_1 = 1/2T$. Напомним, что $\Delta f = 1/2T$ – это минимальная разность частот, необходимая для обеспечения ортогональности сигналов $s_1(t)$ и $s_2(t)$ на

сигнальном интервале длиной T. Это объясняет, почему двоичную МНФ с $h=\frac{1}{2}$ называют модуляцией с минимальным сдвигом (ММС). Фаза на n-м сигнальном интервале определяется состоянием фазы сигнала, которая образуется для непрерывности фазы между соседними интервалами.

ММС можно также представить как разновидность четырёхфазного ФМ. Конкретно мы можем выразить эквивалентный низкочастотный модулирующий сигнал в виде

$$\upsilon(t)=\sum_{n=-\infty}^{\infty}\left[I_{2n}g(t-2nT)-jI_{2n+1}g(t-2nT-T)\right], \qquad (4.3.67)$$

где $g(t)$ – сигнальный импульс, определяемый так:

$$g(t)=\begin{cases}\sin\frac{\pi t}{2T} & (0\le t\le 2T),\\ 0 & (\text{для других } t).\end{cases} \qquad (4.3.68)$$

Таким образом, этот тип сигнала можно рассматривать, как четырёхпозиционный сигнал ФМ, в котором огибающая импульса является полупериодом синусоиды. Чётные двоичные (± 1) символы $\{I_{2n}\}$ от информационной последовательности $\{I_n\}$ передаются при помощи косинусоиды несущей, в то время как нечётные двоичные (±1) символы $\{I_{2n+1}\}$ передаются при помощи синусоиды несущей. Скорость передачи двух ортогональных несущих равна $1/2T$ бит/с, так что суммарная скорость передачи равна $1/T$ бит/с. Заметим, что битовые переходы на синусной и косинусной несущей смещены во времени на T секунд. Из этих соображений сигнал

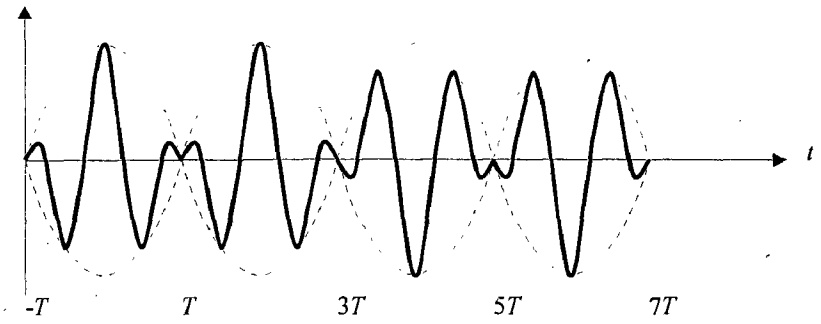
$$s(t)=A\left\{\left[\sum_{n=-\infty}^{\infty}I_{2n}g(t-2nT)\right]\cos 2\pi f_c t+\left[\sum_{n=-\infty}^{\infty}I_{2n+1}g(t-2nT-T)\right]\sin 2\pi f_c t\right\} \qquad (4.3.69)$$

называют офсетной квадратурной ФМ (ОКФМ, OQPSK) или квадратурной ФМ со сдвигом (КФМС, SQPSK).

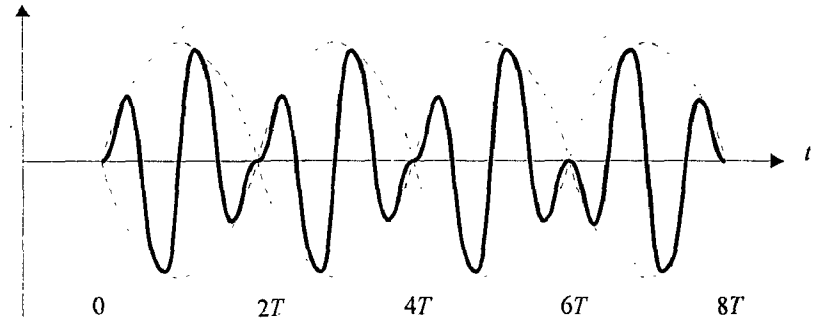
Рисунок 4.3.23 иллюстрирует представление ММС-сигналов как двух смещённых квадратурно-модулированных двоичных ФМ-сигналов. Сумма двух квадратурных сигналов является частотно-модулированным сигналом с постоянной амплитудой.

Интересно сравнить форму сигнала ММС с ОКФМ, в котором импульс $g(t)$ является прямоугольным на интервале $(0\le t\le 2T)$, с обычной квадратурной ФМ (КФМ), в которой импульс $g(t)$ также прямоугольный на интервале $(0\le t\le 2T)$. Ясно, что все три метода модуляции работают при одинаковой скорости передачи данных. ММС сигнал имеет непрерывную фазу. ОКФМ-сигнал с прямоугольным импульсом принципиально является суммой двух двоичных ФМ-сигналов, в которых переходы фазы возникают через T секунд. Таким образом, сигнал имеет скачки фазы на $\pm 90^\circ$, которые могут возникнуть не чаще, чем через T секунд. С другой стороны, обычная четырёхпозиционная ФМ с постоянной амплитудой может иметь скачки фазы $\pm 180^\circ$ или $\pm 90^\circ$ каждые $2T$ секунд. Иллюстрация этих трёх типов сигналов дана на рис. 4.3.24.

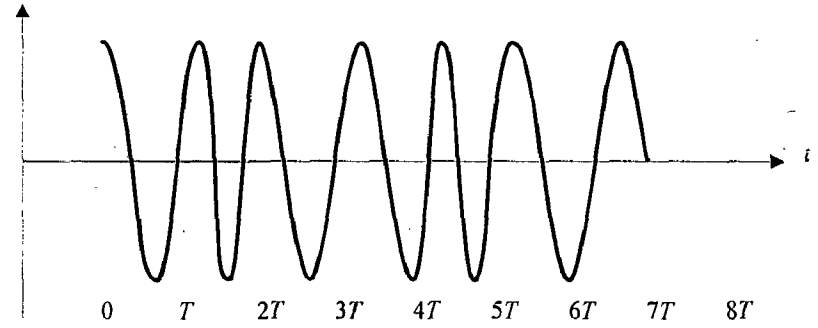
Пространственные диаграммы для сигналов МНФ. В общем, сигналы с непрерывной фазой не могут быть представлены в дискретных точках в пространстве сигналов, как в случае АМ, ФМ и КАМ, поскольку фаза несущей меняется во времени. Вместо этого сигнал с непрерывной фазой описывается переменными фазами или траекториями перехода от одного состояния фазы к другому.



(*a*) Синфазная сигнальная компонента



(*b*) Квадратурная сигнальная компонента



(*c*) Сигнал ММС [сумма (*a*) и (*b*)]

Рис. 4.3.23. Представление сигнала ММС суммой двух взаимно сдвинутых сигналов ФМ, каждый с синусоидальной огибающей

Для сигналов МНФ с постоянной амплитудой переменные траектории образуют окружность. Для примера на рис. 4.3.25 иллюстрируется диаграмма пространства сигналов (фазовые траектории) для МНФ с $h=\frac{1}{4}$, $h=\frac{1}{3}$, $h=\frac{1}{2}$, $h=\frac{2}{3}$.

Места начала и конца этих фазовых траекторий отмечены на рисунке точкой. Заметим, что длина фазовой траектории увеличивается с ростом h. Рост h ведёт также к расширению полосы частот, как будет показано в следующем разделе.

Многоуровневая МНФ. Многоуровневая МНФ является обобщением обычной МНФ, в которой амплитуда сигнала может принять ряд значений, в то время как фаза поддерживается непрерывной. Для примера рассмотрим двухуровневый сигнал ЧМНФ, который можно представить так:

$$s(t)=2A\cos\left[2\pi f_c t+\phi_2(t;\mathbf{I})\right]+A\cos\left[2\pi f_c t+\phi_1(t;\mathbf{J})\right], \qquad (4.3.70)$$

где

$$\phi_2(t;\mathbf{I})=\pi h\sum_{k=-\infty}^{n-1} I_k+\frac{\pi h I_n(t-nT)}{T}, \quad nT\le t\le(n+1)T, \qquad (4.3.71)$$

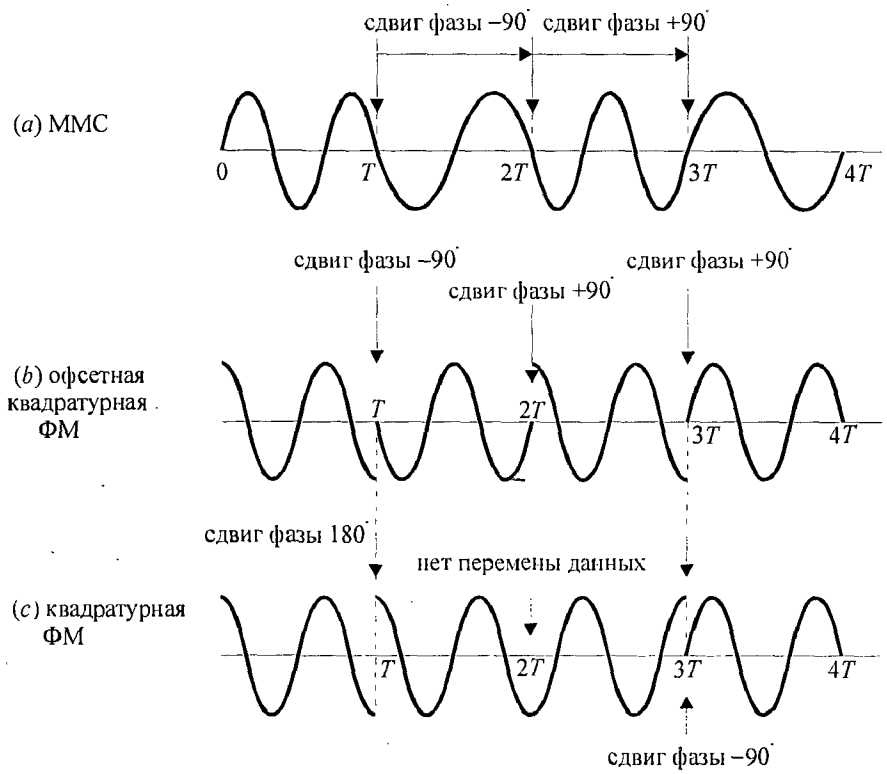


Рис. 4.3.24. Сигнал для (*a*) ММС, (*b*) офсетной квадратурной ФМ (прямоугольный импульс) и (*c*) обычной квадратурной ФМ (прямоугольный импульс) [*Gronemeyer* и *McBride* (1976); © 1976 *IEEE*]

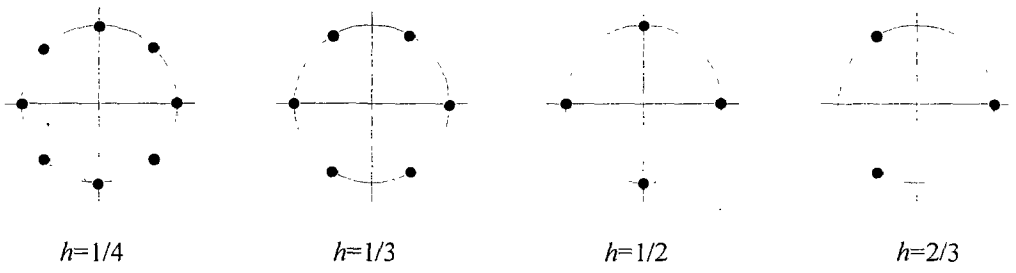


Рис. 4.3.25. Пространственная диаграмма сигнала ЧМНФ

$$\phi_1(t;\mathbf{J}) = \pi h \sum_{k=-\infty}^{n-1} J_k + \frac{\pi h J_n (t - nT)}{T}, \quad nT \le t \le (n+1)T. \tag{4.3.72}$$

Информация передаётся последовательностями символов $\{I_n\}$ и $\{J_n\}$, которые связаны с двумя независимыми двоичными информационными последовательностями $\{a_n\}$ и $\{b_n\}$, принимающими значение $\{0, 1\}$. Видим, что сигнал в (4.3.70) является суперпозицией двух сигналов ЧМНФ с различными амплитудами.

Для детальной проработки рассмотрим случай, когда $h = \frac{1}{2}$, так что мы имеем суперпозицию двух ММС-сигналов. В точке передачи компоненты с различными амплитудами находятся либо в фазе, либо в противофазе. Изменение фазы сигнала определяется фазой компоненты с большой амплитудой, в то время как изменение амплитуды определяется компонентой с меньшей амплитудой. Поэтому меньшая компонента управляется так, чтобы в начале и в конце символьного интервала, она находилась в фазе или была сдвинута на 180° относительно компоненты с большой

амплитудой, независимо от фазы последней. При таком управлении последовательности символов $\{I_n\}$ и $\{J_n\}$ можно выразить так:

$$I_n = 2a_n - 1,$$
$$J_n = I_n(1-2b_n) = I_n\left(1-\frac{b_n}{h}\right). \qquad (4.3.73)$$

Эти соотношения отражены в табл. 4.3.1.

Таблица 4.3.1

a_n	b_n	I_n	J_n	Амплитудно-фазовые отношения
0	0	−1	−1	Амплитуда постоянна; фаза уменьшается
0	1	−1	1	Амплитуда меняется; фаза уменьшается
1	0	1	1	Амплитуда постоянна; фаза растёт
1	1	1	−1	Амплитуда меняется; фаза растёт

Как обобщение, сигнал многоуровневой ЧМНФ с n компонентами можно выразить так:

$$s(t) = 2^{N-1}\cos\left[2\pi f_c t + \phi_N(t;\mathbf{I})\right] + \sum_{m=1}^{N-1} 2^{m-1}\cos\left[2\pi f_c t + \phi_m(t;\mathbf{J}_m)\right], \qquad (4.3.74)$$

где

$$\phi_N(t;\mathbf{I}) = \pi h I_n \frac{t-nT}{T} + \pi h \sum_{k=-\infty}^{n-1} I_k, \quad nT \le t \le (n+1)T, \qquad (4.3.75)$$

$$\phi_m(t;\mathbf{J}_m) = I_n\pi\left[h+\tfrac{1}{2}(J_{mn}+1)\right]\frac{t-nT}{T} + \sum_{k=-\infty}^{n-1}\pi I_k\left[h+\tfrac{1}{2}(J_{mk}+1)\right], \quad nT \le t \le (n+1)T. \qquad (4.3.76)$$

Последовательности $\{I_n\}$ и $\{J_{mn}\}$ статистически независимы, они двоичные, а символы принимают значения из ряда $\{1, -1\}$.

Из (4.3.75) и (4.3.76) видим, что каждая компонента в сумме будет или в фазе, или со сдвигом 180° относительно фазы наибольшей компоненты в концах интервала n-го символа, т.е. при $t = (n+1)T$. Таким образом, состояния сигналов определяются уровнями амплитуд из ряда значений $\{1, 3, 5, \ldots, 2^N - 1\}$ и значениями фаз из ряда $\{0, \pi\theta, 2\pi\theta, \ldots, 2\pi - \pi h\}$. Управление фазой требуется для того, чтобы поддерживать непрерывной фазу сигнала МНФ.

Рисунок 4.3.26 иллюстрирует диаграмму состояний сигнала для двухамплитудной $(N=2)$ ЧМНФ с $h = \frac{1}{4}, \frac{1}{3}, \frac{1}{2}$ и $\frac{2}{3}$. Диаграммы состояний *трёхкомпонентной* $(N=3)$ ЧМНФ показаны на рис. 4.3.27. В этом случае имеются *четыре* уровня амплитуд. Число состояний зависит как от индекса модуляции h, как и от N. Дополнительные многоуровневые формы сигналов ЧМНФ можно получить с использованием как огибающих импульсов, отличных от прямоугольных, так и сигнальных импульсов, которые тянутся более чем на интервал одного символа (парциальный отклик).

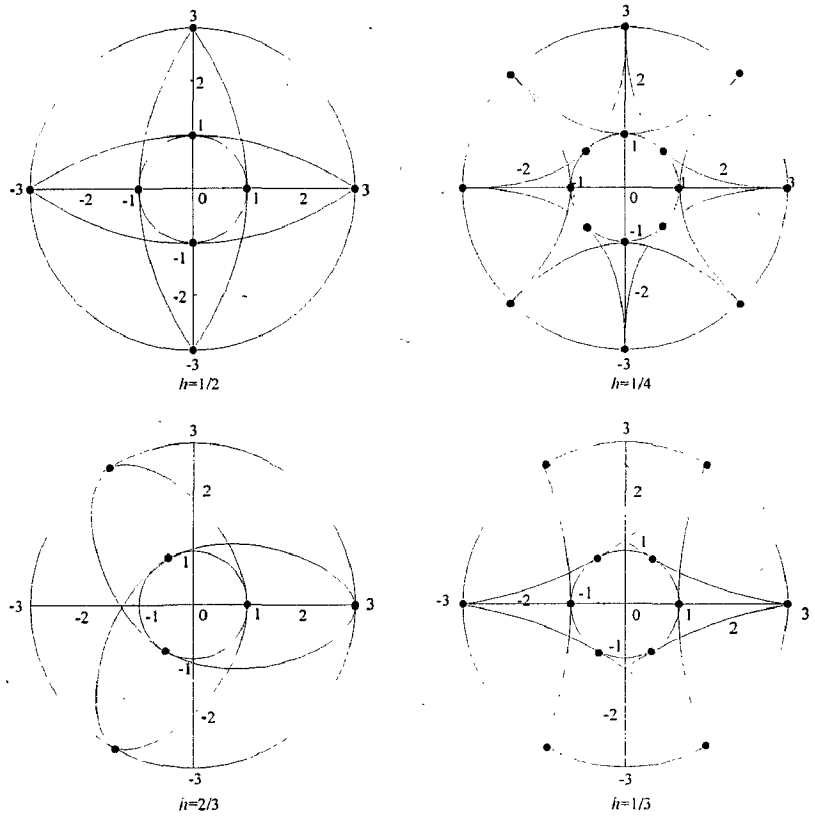


Рис. 4.3.26. Пространственная диаграмма сигнала двухкомпонентной ЧМНФ

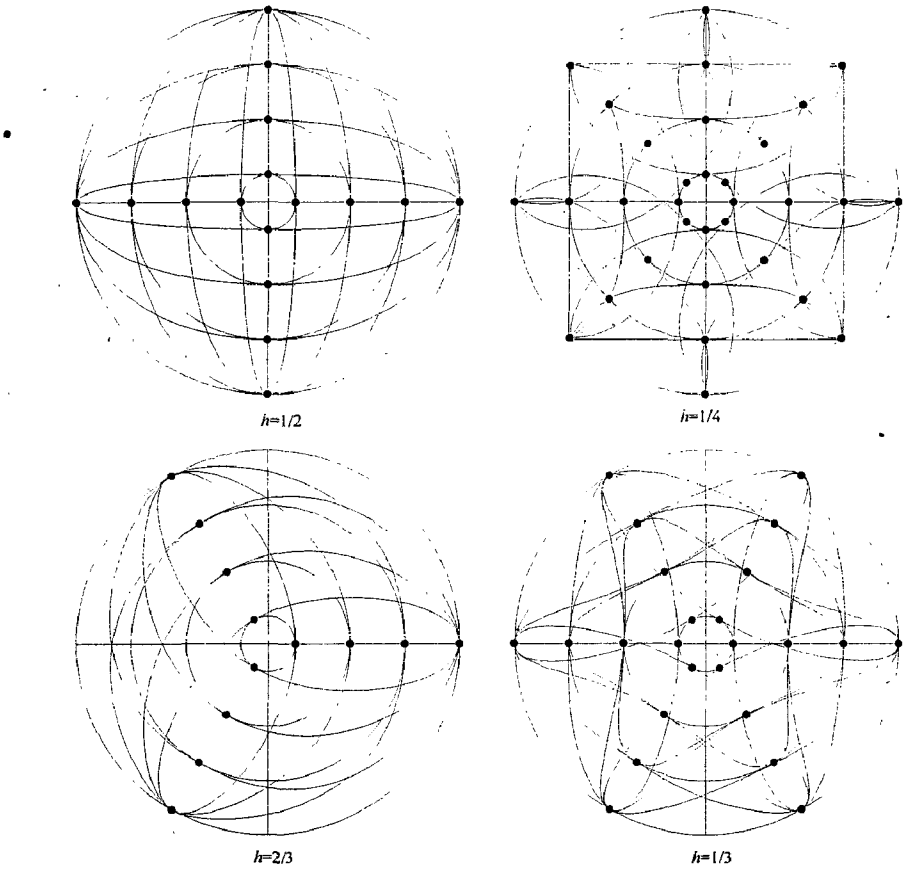


Рис. 4.3.27. Пространственная диаграмма сигнала трёхкомпонентной ЧМНФ

4.4. СПЕКТРАЛЬНЫЕ ХАРАКТЕРИСТИКИ СИГНАЛОВ ЦИФРОВОЙ МОДУЛЯЦИИ

В большинстве систем цифровой связи имеющаяся в распоряжении полоса частот ограничена. Как следствие, проектировщик системы должен рассмотреть, как ограничение полосы частот канала влияет на выбор техники модуляции, используемой при передаче информации. Из этих соображений для нас важно определить спектральные характеристики сигналов цифровой модуляции, рассмотренных в разд. 4.3.

Поскольку информационная последовательность случайная, сигнал цифровой модуляции является случайным процессом. Мы интересуемся спектральной плотностью мощности такого процесса. Исходя из спектральной плотности мощности, мы можем определить полосу частот канала, необходимую для передачи информационного сигнала. Сначала определим спектральные характеристики класса сигналов с линейной модуляцией. Затем рассмотрим нелинейные ЧМНФ, МНФ и базовый модулирующий сигнал с памятью.

4.4.1. Спектр мощности сигналов линейной модуляции

Начиная с формы сигнала

$$s(t) = \mathrm{Re}\left[\upsilon(t)\,e^{j2\pi f_c t}\right],$$

которая отображает полосовой сигнал $s(t)$ через эквивалентный низкочастотный $\upsilon(t)$, можем выразить автокорреляционную функцию $s(t)$ так:

$$\phi_{ss}(\tau) = \mathrm{Re}\left[\phi_{\upsilon\upsilon}(\tau)e^{j2\pi f_c \tau}\right], \qquad (4.4.1)$$

где $\phi_{\upsilon\upsilon}(\tau)$ – автокорреляционная функция низкочастотного эквивалента $\upsilon(t)$.

Преобразование Фурье (4.4.1) даёт желаемое выражение для спектральной плотности мощности $\Phi_{ss}(f)$ в виде

$$\Phi_{ss}(f) = \tfrac{1}{2}\left[\Phi_{\upsilon\upsilon}(f - f_c) + \Phi_{\upsilon\upsilon}(-f - f_c)\right], \qquad (4.4.2)$$

где $\Phi_{\upsilon\upsilon}(f)$ – спектральная плотность мощности $\upsilon(t)$. Достаточно определить автокорреляционную функцию и спектральную плотность мощности для эквивалентного низкочастотного сигнала $\upsilon(t)$.

Сначала рассмотрим методы линейной цифровой модуляции, для которых $\upsilon(t)$ можно представить в общем виде так:

$$\upsilon(t) = \sum_{n=-\infty}^{\infty} I_n g(t - nT), \qquad (4.4.3)$$

где скорость передачи канальных символов равна $1/T = R/k$ симв./с, а $\{I_n\}$ представляет последовательность символов, которая возникает при отображении k-битовых блоков в соответствующие сигнальные точки, выбираемые по соответствующей диаграмме пространства сигналов. Отметим, что в АМ последовательность $\{I_n\}$ вещественна и соответствует значениям амплитуд передаваемого сигнала, но в ФМ, КАМ и комбинированной АМ-ФМ последовательность $\{I_n\}$ комплексная, так как точка сигналов имеет двухмерное представление. Автокорреляционная функция случайного процесса $\upsilon(t)$ равна

$$\begin{aligned}\phi_{\upsilon\upsilon}(t+\tau;\, t) &= \tfrac{1}{2}E\left[\upsilon^*(t)\,\upsilon(t+\tau)\right] = \\ &= \tfrac{1}{2}\sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty} E\left[I_n^* I_m\right] g^*(t - nT)\, g(t + \tau - mT).\end{aligned} \qquad (4.4.4)$$

Предположим, что последовательность информационных символов $\{I_n\}$ стационарна в широком смысле со средним μ_I и автокорреляционной функцией

$$\phi_{ii}(m) = \tfrac{1}{2} E\left[I_n^* I_{n+m}\right]. \tag{4.4.5}$$

Тогда (4.4.4) можно выразить так:

$$\begin{aligned}\phi_{\upsilon\upsilon}(t+\tau;\, t) &= \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \phi_{ii}(m-n) g^*(t-nT) g(t+\tau-mT) = \\ &= \sum_{m=-\infty}^{\infty} \phi_{ii}(m) \sum_{n=-\infty}^{\infty} g^*(t-nT) g(t+\tau-nT-mT).\end{aligned} \tag{4.4.6}$$

Вторая сумма в (4.4.6), именно

$$\sum_{n=-\infty}^{\infty} g^*(t-nT) g(t+\tau-nT-mT),$$

– периодическая функция по переменной t с периодом T. Следовательно, $\phi_{\upsilon\upsilon}(t+\tau;\, t)$ – также периодическая функция по переменной t с периодом T. Это означает, что

$$\phi_{\upsilon\upsilon}(t+T+\tau;\, t+T) = \phi_{\upsilon\upsilon}(t+\tau;\, t). \tag{4.4.7}$$

Кроме того, среднее значение $\upsilon(t)$, которое равно

$$E[\upsilon(t)] = \mu_I \sum_{n=-\infty}^{\infty} g(t-nT), \tag{4.4.8}$$

– периодическая функция по переменной t с периодом T. Следовательно, $\upsilon(t)$ является случайным процессом, имеющим периодические средние значения и автокорреляционную функцию. Такой процесс называется *циклостационарным процессом* или *периодически стационарным процессом в широком смысле*, как описано в разд. 2.2.6.

Чтобы рассчитать спектральную плотность мощности циклостационарного процесса зависимость $\phi_{\upsilon\upsilon}(t+\tau;\, t)$ от переменной t должна быть исключена. Это можно сделать просто путём усреднения $\phi_{\upsilon\upsilon}(t+\tau;\, t)$ по t по одному периоду T. Таким образом,

$$\begin{aligned}\overline{\phi}_{\upsilon\upsilon}(\tau) &= \tfrac{1}{T} \int_{-T/2}^{T/2} \phi_{\upsilon\upsilon}(t+\tau;t)\, dt = \\ &= \sum_{m=-\infty}^{\infty} \phi_{ii}(m) \sum_{n=-\infty}^{\infty} \tfrac{1}{T} \int_{-T/2}^{T/2} g^*(t-nT) g(t+\tau-nT-mT)\, dt = \\ &= \sum_{m=-\infty}^{\infty} \phi_{ii}(m) \sum_{n=-\infty}^{\infty} \tfrac{1}{T} \int_{-T/2-nT}^{T/2-nT} g^*(t) g(t+\tau-mT)\, dt.\end{aligned} \tag{4.4.9}$$

Мы интерпретируем интеграл в (4.4.9) вместе с суммой по n как временную автокорреляционную функцию $g(t)$ и определим её как

$$\phi_{gg}(\tau) = \int_{-\infty}^{\infty} g^*(t) g(t+\tau)\, dt. \tag{4.4.10}$$

Следовательно, (4.4.9) можно выразить так:

$$\overline{\phi}_{\upsilon\upsilon}(\tau) = \frac{1}{T} \sum_{m=-\infty}^{\infty} \phi_{ii}(m) \phi_{gg}(\tau-mT). \tag{4.4.11}$$

Преобразование Фурье (4.4.11) даёт (среднюю) спектральную плотность мощности $\upsilon(t)$ в виде

$$\Phi_{\upsilon\upsilon}(f) = \frac{1}{T} |G(f)|^2 \Phi_{ii}(f), \tag{4.4.12}$$

где $G(f)$ – преобразованное Фурье для $g(t)$, а $\Phi_{ii}(f)$ определяет спектральную плотность

мощности информационной последовательности, определяемую как

$$\Phi_{ii}(f)=\sum_{m=-\infty}^{\infty}\phi_{ii}(m)\,e^{-j2\pi fmT}. \qquad (4.4.13)$$

Результат (4.4.12) иллюстрирует зависимость спектральной плотности мощности $\upsilon(t)$ от спектральных характеристик импульса $g(t)$ и информационной последовательности $\{I_n\}$. Это означает, что спектральными характеристиками $\upsilon(t)$ можно управлять через огибающую импульса $g(t)$ и корреляционные характеристики информационной последовательности.

В то время как зависимость $\Phi_{\upsilon\upsilon}(f)$ от $G(f)$ легко понять в (4.4.12), влияние корреляционных свойств информационной последовательности более тонкое. Прежде всего заметим, что для произвольной автокорреляционной функции $\phi_{ii}(m)$ соответствующая спектральная плотность мощности $\Phi_{ii}(f)$ – периодическая функция по частоте с периодом $1/T$. Действительно выражение (4.4.13), определяющее спектр $\Phi_{ii}(f)$ по $\phi_{ii}(m)$, является комплексным рядом Фурье с коэффициентами Фурье $\{\phi_{ii}(m)\}$. Как следствие, автокорреляционная последовательность $\phi_{ii}(m)$ определяется так:

$$\phi_{ii}(m)=T\int_{-T/2}^{T/2}\Phi_{ii}(f)e^{j2\pi f_m T}df. \qquad (4.4.14)$$

Во-вторых, рассмотрим случай, когда информационные символы в последовательности вещественные и взаимно некоррелированные. В этом случае автокорреляционную функцию $\phi_{ii}(m)$ можно выразить так:

$$\phi_{ii}(m)=\begin{cases}\sigma_i^2+\mu_i^2, & m=0,\\ \mu_i^2, & m\neq 0,\end{cases} \qquad (4.4.15)$$

где σ_i^2 означает дисперсию информационных символов. Если (4.4.15) подставить в (4.4.13), получим

$$\Phi_{ii}(f)=\sigma_i^2+\mu_i^2\sum_{m=-\infty}^{\infty}e^{-j2\pi fmT}. \qquad (4.4.16)$$

Сумма в (4.4.16) – периодическая функция частоты с периодом $1/T$. Её можно рассмотреть как комплексный ряд Фурье для периодической последовательности δ-импульсов с периодом $1/T$. Следовательно, (4.4.16) можно также выразить в виде

$$\Phi_{ii}(f)=\sigma_i^2+\frac{\mu_i^2}{T}\sum_{m=-\infty}^{\infty}\delta\left(f-\frac{m}{T}\right). \qquad (4.4.17)$$

Подстановка (4.4.17) в (4.4.12) определяет спектральную плотность мощности $\upsilon(t)$ для случая, когда информационные символы не коррелированы. Получаем

$$\Phi_{\upsilon\upsilon}(f)=\frac{\sigma_i^2}{T}\left|G(f)\right|^2+\frac{\mu_i^2}{T^2}\sum_{m=-\infty}^{\infty}\left|G\left(\frac{m}{T}\right)\right|^2\delta\left(f-\frac{m}{T}\right). \qquad (4.4.18)$$

Выражение (4.4.18) для спектральной плотности мощности специально разделено на два слагаемых, чтобы подчеркнуть два различных вида спектральных компонент. Первое определяет непрерывный спектр, и его огибающая зависит только от спектральной характеристики сигнального импульса $g(t)$. Второе слагаемое состоит из дискретных частотных компонент, появляющихся через интервал $1/T$. Каждая такая компонента имеет мощность, пропорциональную $\left|G(f)\right|^2$ при $f=m/T$. Заметим, что дискретные частотные

компоненты исчезают, когда информационные символы имеют нулевое среднее, т.е. $\mu_i = 0$. Это условие обычно желательно для техники цифровой модуляции. Оно выполняется, когда информационные символы равновероятны и симметрично расположены на комплексной плоскости. Таким образом, проектировщик системы может управлять спектральными характеристиками сигналов цифровой модуляции путём специального подбора характеристик информационной последовательности, которую нужно передать.

Пример 4.4.1. Чтобы проиллюстрировать влияние $g(t)$ на огибающую спектра, рассмотрим прямоугольный импульс, показанный на рис. 4.4.1(а). Преобразование Фурье от $g(t)$ равно

$$G(f) = AT\frac{\sin \pi f T}{\pi f T} e^{-j\pi f T}.$$

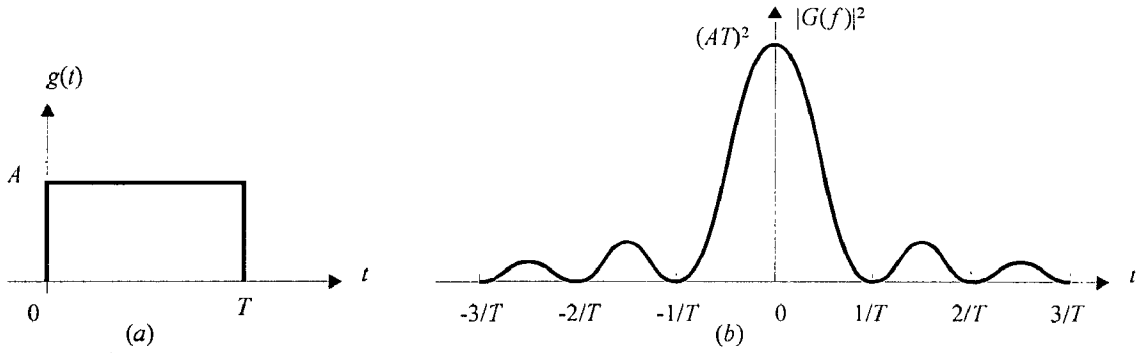


Рис. 4.4.1. Прямоугольный импульс и его спектральная плотность энергии $|G(f)|^2$

Следовательно,

$$|G(f)|^2 = (AT)^2\left(\frac{\sin \pi f T}{\pi f T}\right)^2. \qquad (4.4.19)$$

Этот спектр показан на рис. 4.4.1(b). Заметим, что спектральная плотность принимает нулевые значения в точках оси частот, кратных $1/T$, и убывает обратно квадрату частоты. Как следствие наличия нулей в $G(f)$, все дискретные спектральные компоненты в (4.4.18), кроме одной, исчезают. Подставляя (4.4.19) в (4.4.18), имеем

$$\Phi_{vv}(f) = \sigma_i^2 A^2 T\left(\frac{\sin \pi f T}{\pi f T}\right)^2 + A^2\mu_i^2\delta(f). \qquad (4.4.20)$$

Пример 4.4.2. В качестве второй иллюстрации влияния на огибающую спектра импульса $g(t)$ рассмотрим импульс приподнятого косинуса

$$g(t) = \frac{A}{2}\left[1 + \cos\frac{2\pi}{T}\left(t - \frac{T}{2}\right)\right], \qquad 0 \le t \le T. \qquad (4.4.21)$$

График этой функции дан на рис. 4.4.2(a). Его преобразование Фурье легко получить, и его можно выразить в виде

$$G(f) = \frac{AT}{2}\frac{\sin \pi f T}{\pi f T(1 - f^2T^2)} e^{-j\pi f T}. \qquad (4.4.22)$$

Квадрат амплитуды $G(f)$ показан на рис. 4.4.2(b). Интересно отметить, что спектр имеет нули в точках $f = n/T$, $n = \pm 2, \pm 3, \pm 4, \ldots$. Следовательно, все дискретные

спектральные компоненты в (4.4.18), кроме тех, которые на частотах $f = 0$ и $f = \pm 1/T$, исчезают. По сравнению со спектром при прямоугольном импульсе спектр приподнятого косинуса имеет более широкий главный лепесток, но хвосты уменьшаются обратно f^6.

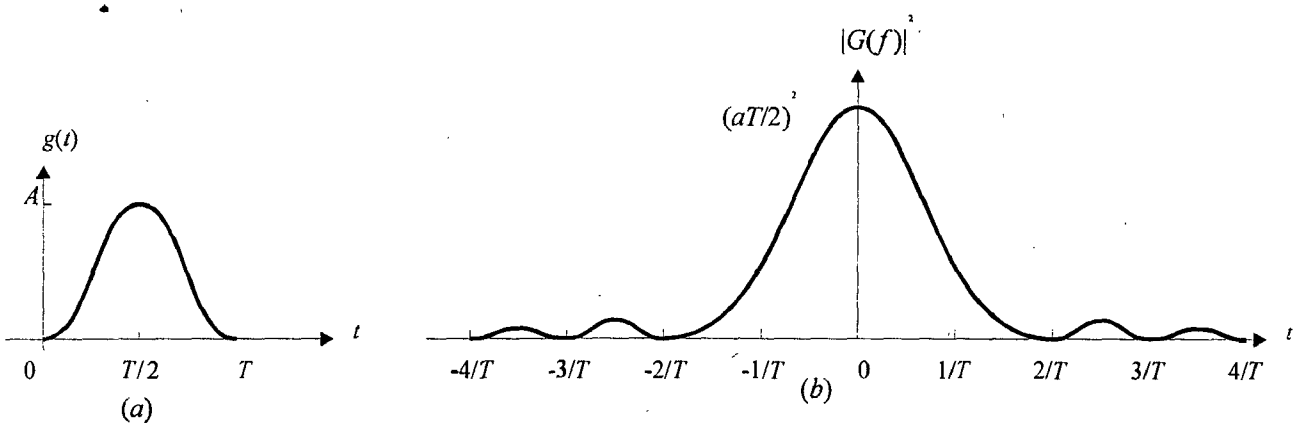


Рис. 4.4.2. Импульс приподнятого косинуса и его спектральная плотность энергии $|G(f)|^2$

Пример 4.4.3. Чтобы проиллюстрировать влияние на огибающую спектра операций, выполняемых по отношению к информационной последовательности, рассмотрим двоичную последовательность $\{b_n\}$, по которой формируем символы информационной последовательности

$$I_n = b_n + b_{n-1}. \tag{4.4.23}$$

Предполагается, что последовательность $\{b_n\}$ содержит некоррелированные случайные величины, каждое с нулевым средним и единичной дисперсией. Тогда автокорреляционная функция последовательности $\{I_n\}$ равна

$$\phi_{ii}(m) = E\left(I_n I_{n+m}\right) = \begin{cases} 2 & m = 0, \\ 1 & m = \pm 1, \\ 0 & \text{другие } m. \end{cases} \tag{4.4.24}$$

Следовательно спектральная плотность мощности входной последовательности равна

$$\Phi_{ii}(f) = 2\left(1 + \cos 2\pi f T\right) = 4\cos^2 \pi f T, \tag{4.4.25}$$

и соответствующая спектральная плотность мощности для (низкочастотного) модулирующего сигнала

$$\Phi_{\upsilon\upsilon}(f) = \frac{4}{T}\left|G(f)\right|^2 \cos^2 \pi f T. \tag{4.4.26}$$

4.4.2. Спектр мощности для сигналов ЧМНФ и МНФ

В этом разделе мы получим спектральную плотность мощности для класса сигналов МНФ с постоянной амплитудой, которые были описаны в разд. 4.3.3. Начнём с расчёта автокорреляционной функции и её преобразования Фурье, как мы это сделали в случае линейной модуляции. Сигнал МНФ с постоянной амплитудой выражается так:

$$s(t;\mathbf{I}) = A\cos\left[2\pi f_c t + \phi(t;\mathbf{I})\right], \tag{4.4.27}$$

где

$$\phi(t;\mathbf{I}) = 2\pi h \sum_{k=-\infty}^{\infty} I_k q(t - kT). \tag{4.4.28}$$

Каждый символ последовательности $\{I_n\}$ может принять одно из M значений $\{\pm 1, \pm 3, \ldots, \pm(M-1)\}$. Эти символы статистически независимы и одинаково распределены с априорными вероятностями

$$P_n = P(I_k = n), \quad n = \pm 1, \pm 2, \ldots, \pm(M-1). \tag{4.4.29}$$

Импульс $g(t) = dq(t)/dt$ равен нулю вне интервала $[0, LT]$, $q(t) = 0$, $t < 0$ и $q(t) = \frac{1}{2}$ для $t > LT$. Автокорреляционная функция эквивалентного низкочастотного сигнала $\upsilon(t) = e^{j\phi(t;I)}$ равна

$$\phi_{\upsilon\upsilon}(t+\tau;t) = \tfrac{1}{2} E\left[\exp\left(j2\pi h \sum_{k=-\infty}^{\infty} I_k\left[q(t+\tau-kT) - q(t-kT)\right]\right)\right]. \tag{4.4.30}$$

Сначала выразим сумму в показателе экспоненты как произведение экспонент. Результат равен

$$\phi_{\upsilon\upsilon}(t+\tau;t) = \tfrac{1}{2} E\left[\prod_{k=-\infty}^{\infty} \exp\left\{j2\pi h I_k\left[q(t+\tau-kT) - q(t-kT)\right]\right\}\right]. \tag{4.4.31}$$

Далее найдём математическое ожидание по символам $\{I_k\}$. Поскольку эти символы статистически независимы, получаем

$$\phi_{\upsilon\upsilon}(t+\tau;t) = \prod_{k=-\infty}^{\infty}\left(\sum_{\substack{n=-(M-1)\\ n \text{ нечетные}}}^{M-1} P_n \exp\left\{j2\pi hn\left[q(t+\tau-kT) - q(t-kT)\right]\right\}\right). \tag{4.4.32}$$

Наконец, усреднённая во времени автокорреляционная функция равна

$$\bar{\phi}_{\upsilon\upsilon}(\tau) = \tfrac{1}{T}\int_0^T \phi_{\upsilon\upsilon}(t+\tau;t)dt. \tag{4.4.33}$$

Хотя (4.4.32) подразумевает, что имеется неограниченное число множителей, импульс $g(t) = dq(t)/dt = 0$ для $t < 0$ и $t > LT$, а $q(t) = 0$ для $t < 0$. Как следствие, только ограниченное число слагаемых в произведении имеет ненулевые значения показателя экспоненты. Таким образом, (4.4.32) можно существенно упростить. Если принять $\tau = \xi + mT$, где $0 \le \xi < T$ и $m = 0, 1, \ldots$, то усреднённая автокорреляционная функция (4.4.33) приводится к результату

$$\bar{\phi}_{\upsilon\upsilon}(\xi + mT) = \\ = \frac{1}{2T}\int_0^T \prod_{k=1-L}^{m+1}\left(\sum_{\substack{n=-(M-1)\\ n \text{ нечетные}}}^{M-1} P_n \exp\left\{j2\pi hn\left[q(t+\xi-(k-m)T) - q(t-kT)\right]\right\}\right). \tag{4.4.34}$$

Рассмотрим $\bar{\phi}_{\upsilon\upsilon}(\xi + mT)$ для $\xi + mT \ge LT$. В этом случае (4.4.34) можно выразить так:

$$\bar{\phi}_{\upsilon\upsilon}(\xi + mT) = \left[\psi(jh)\right]^{m-L}\lambda(\xi), \qquad m \ge L, \quad 0 \le \xi < T, \tag{4.4.35}$$

где $\psi(jh)$ – характеристическая функция для случайной последовательности $\{I_n\}$, определяется так:

$$\psi(jh) = E\left(e^{j\pi hI_n}\right) = \sum_{\substack{n=-(M-1)\\ n \text{ нечетные}}}^{M-1} P_n e^{j\pi hn}, \tag{4.4.36}$$

а $\lambda(\xi)$ – остаточная часть усреднённой автокорреляционной функции, которую можно выразить как

$$\lambda(\xi)=\frac{1}{2T}\int_0^T\prod_{k=1-L}^{0}\left(\sum_{n=-(M-1)}^{M-1}P_n\exp\left\{j2\pi hn\left[\tfrac{1}{2}-q(t-kT)\right]\right\}\right)\times$$
$$\times\prod_{k=1-L}^{1}\left(\sum_{\substack{n=-(M-1)\\ n\text{ нечетные}}}^{M-1}P_n\exp\left[j2\pi hnq(t+\xi-kT)\right]\right)dt,\ m\geq L. \qquad (4.4.37)$$

Таким образом, $\bar{\phi}_{\upsilon\upsilon}(\tau)$ можно представить произведением $\lambda(\xi)$ и степени $\psi(jh)$, как указанно в (4.4.35) для $\tau=\xi+mT\geq LT$ и $0\leq\xi<T$. Эти свойства используются ниже.

Преобразование Фурье $\bar{\phi}_{\upsilon\upsilon}(\tau)$ даёт среднюю спектральную плотность мощности

$$\Phi_{\upsilon\upsilon}(f)=\int_{-\infty}^{\infty}\bar{\phi}_{\upsilon\upsilon}(\tau)\,e^{-j2\pi f\tau}d\tau=2\,\mathrm{Re}\left[\int_0^{\infty}\bar{\phi}_{\upsilon\upsilon}(\tau)e^{-j2\pi f\tau}d\tau\right]. \qquad (4.4.38)$$

Но

$$\int_0^{\infty}\bar{\phi}_{\upsilon\upsilon}(\tau)\,e^{-j2\pi f\tau}d\tau=\int_0^{LT}\bar{\phi}_{\upsilon\upsilon}(\tau)\,e^{-j2\pi f\tau}d\tau+\int_{LT}^{\infty}\bar{\phi}_{\upsilon\upsilon}(\tau)\,e^{-j2\pi f\tau}d\tau. \qquad (4.4.39)$$

С учётом (4.4.35) интеграл в области $LT\leq\tau<\infty$ можно выразить так:

$$\int_{LT}^{\infty}\bar{\phi}_{\upsilon\upsilon}(\tau)e^{-j2\pi f\tau}d\tau=\sum_{m=L}^{\infty}\int_{mT}^{(m+1)T}\bar{\phi}_{\upsilon\upsilon}(\tau)\,e^{-j2\pi f\tau}d\tau. \qquad (4.4.40)$$

Теперь пусть $\tau=\xi+mT$. Тогда (4.4.40) выражается как

$$\int_{LT}^{\infty}\bar{\phi}_{\upsilon\upsilon}(\tau)e^{-j2\pi f\tau}d\tau=\sum_{m=L}^{x}\int_0^T\bar{\phi}_{\upsilon\upsilon}(\xi+mT)\,e^{-j2\pi f(\xi+mT)}d\xi=$$
$$=\sum_{m=L}^{\infty}\int_0^T\lambda(\xi)\left[\psi(jh)\right]^{m-L}e^{-j2\pi f(\xi+mT)}d\xi=\sum_{n=0}^{\infty}\psi^n(jh)e^{-j2\pi fnT}\int_0^T\lambda(\xi)\,e^{-j2\pi f(\xi+LT)}d\xi. \qquad (4.4.41)$$

Характеристическая функция удовлетворяет условию $|\psi(jh)|\leq1$. Для значений h, для которых $|\psi(jh)|<1$, сумма в (4.4.41) сходится к результату

$$\sum_{n=0}^{\infty}\psi^n(jh)e^{-j2\pi fnT}=\frac{1}{1-\psi(ih)e^{-j2\pi fT}}. \qquad (4.4.42)$$

В этом случае (4.4.41) приводит к виду

$$\int_{LT}^{\infty}\bar{\phi}_{\upsilon\upsilon}(\tau)\,e^{-j2\pi f\tau}d\tau=\frac{1}{1-\psi(jh)e^{-j2\pi fT}}\int_0^T\bar{\phi}_{\upsilon\upsilon}(\xi+LT)\,e^{-j2\pi f(\xi+LT)}d\xi. \qquad (4.4.43)$$

Объединяя (4.4.38), (4.4.39) и (4.4.43), получаем формулу для спектральной плотности мощности сигнала МНФ в виде

$$\Phi_{\upsilon\upsilon}(f)=2\,\mathrm{Re}\left[\int_0^{LT}\bar{\phi}_{\upsilon\upsilon}(\tau)\,e^{-j2\pi f\tau}d\tau+\frac{1}{1-\psi(jh)e^{-j2\pi fT}}\int_{LT}^{(L+1)T}\bar{\phi}_{\upsilon\upsilon}(\tau)\,e^{-j2\pi f\tau}d\tau\right]. \qquad (4.4.44)$$

Это требуемый результат, когда $|\psi(jh)|<1$. В общем, спектральная плотность мощности вычисляется численно по формуле (4.4.44). Усреднённую автокорреляционную функцию $\bar{\phi}_{\upsilon\upsilon}(\tau)$ для области $0\leq\tau\leq(L+1)T$ можно вычислять численно из (4.4.34). Для значений h, для которых $|\psi(jh)|=1$, например $h=K$, где K – целое, можно положить

$$\psi(jh)=e^{j2\pi\nu},\quad 0\leq\nu<1. \qquad (4.4.45)$$

Тогда сумма в (4.4.41) даёт

$$\sum_{n=0}^{\infty} e^{-j2\pi T(f-\nu/T)n} = \tfrac{1}{2} + \frac{1}{2T}\sum_{n=-\infty}^{\infty} \delta\left(f - \frac{\nu}{T} - \frac{n}{T}\right) - j\tfrac{1}{2}\operatorname{ctg}\pi T\left(f - \frac{\nu}{T}\right). \qquad (4.4.46)$$

Таким образом, спектральная плотность мощности теперь содержит дискретные компоненты, локализованные на частотах

$$f_n = \frac{n+\nu}{T}, \qquad 0 \le \nu < 1, \qquad n = 0,1,2,.... \qquad (4.4.47)$$

Результат (4.4.46) можно объединить с (4.4.41) и (4.4.39), чтобы получить полную спектральную плотность мощности, которая включает компоненту с непрерывным спектром и компоненту с дискретным спектром.

Вернёмся к случаю $|\psi(jh)| < 1$. Если информационные символы равновероятны, т.е.

$$P_n = \frac{1}{M} \qquad \text{для всех } n,$$

то характеристическая функция упрощается до выражения

$$\psi(jh) = \frac{1}{M}\sum_{\substack{n=-(M-1)\\ n \text{ нечетные}}}^{M-1} e^{j\pi hn} = \frac{1}{M}\cdot\frac{\sin M\pi h}{\sin \pi h}. \qquad (4.4.48)$$

Заметим, что в этом случае $\psi(jh)$ вещественно. Усреднённая автокорреляционная функция, определяемая (4.4.34), также упрощается в этом случае:

$$\bar{\phi}_{uu}(\tau) = \frac{1}{2T}\int_0^T \prod_{k=1-L}^{[\tau/T]} \frac{1}{M}\frac{\sin 2\pi hM\left[q(t+\tau-kT) - q(t-kT)\right]}{\sin 2\pi h\left[q(t+\tau-kT) - q(t-kT)\right]}\,dt. \qquad (4.4.49)$$

Соответствующее выражение для спектральной плотности мощности

$$\Phi_{uu}(f) = 2\left[\int_0^{LT} \bar{\phi}_{uu}(\tau)\cos 2\pi f\tau\, d\tau + \right.$$
$$\left. + \frac{1-\psi(jh)\cos 2\pi fT}{1+\psi^2(jh)-2\psi(jh)\cos 2\pi fT}\int_{LT}^{(L+1)T} \bar{\phi}_{uu}(\tau)\cos 2\pi f\tau\, d\tau\right]. \qquad (4.4.50)$$

Спектральная плотность мощности для ЧМНФ. Замкнутое выражение для спектральной плотности мощности можно получить из (4.4.50) тогда, когда огибающая импульса $g(t)$ прямоугольная и равна нулю вне интервала $[0, T]$. В этом случае $q(t)$ линейно для $0 \le t \le T$. Результирующую спектральную плотность мощности можно выразить так:

$$\Phi_{uu}(f) = T\left[\frac{1}{M}\sum_{n=1}^{M} A_n^2(f) + \frac{2}{M^2}\sum_{n=1}^{M}\sum_{m=1}^{M} B_{nm}(f)A_n(f)A_m(f)\right], \qquad (4.4.51)$$

где

$$A_n(f) = \frac{\sin\pi\left[fT - \frac{1}{2}(2n-1-M)h\right]}{\pi\left[fT - \frac{1}{2}(2n-1-M)h\right]}, \qquad B_{nm}(f) = \frac{\cos(2\pi fT - \alpha_{nm}) - \psi\cos\alpha_{nm}}{1+\psi^2 - 2\psi\cos 2\pi fT}, \qquad (4.4.52)$$
$$\alpha_{nm} = \pi h(m+n-1-M), \qquad \psi \equiv \psi(jh) = \frac{\sin M\pi h}{M\sin\pi h}.$$

Спектральная плотность мощности ЧМНФ для $M = 2, 4$ и 8 показана соответственно на рис. 4.4.3–4.4.5 как функция нормированной частоты fT при индексе модуляции $h = 2f_dT$ в качестве параметра. Заметим, что на графиках показана только половина занимаемой полосы частот. Начало координат соответствует частоте несущей f_c.

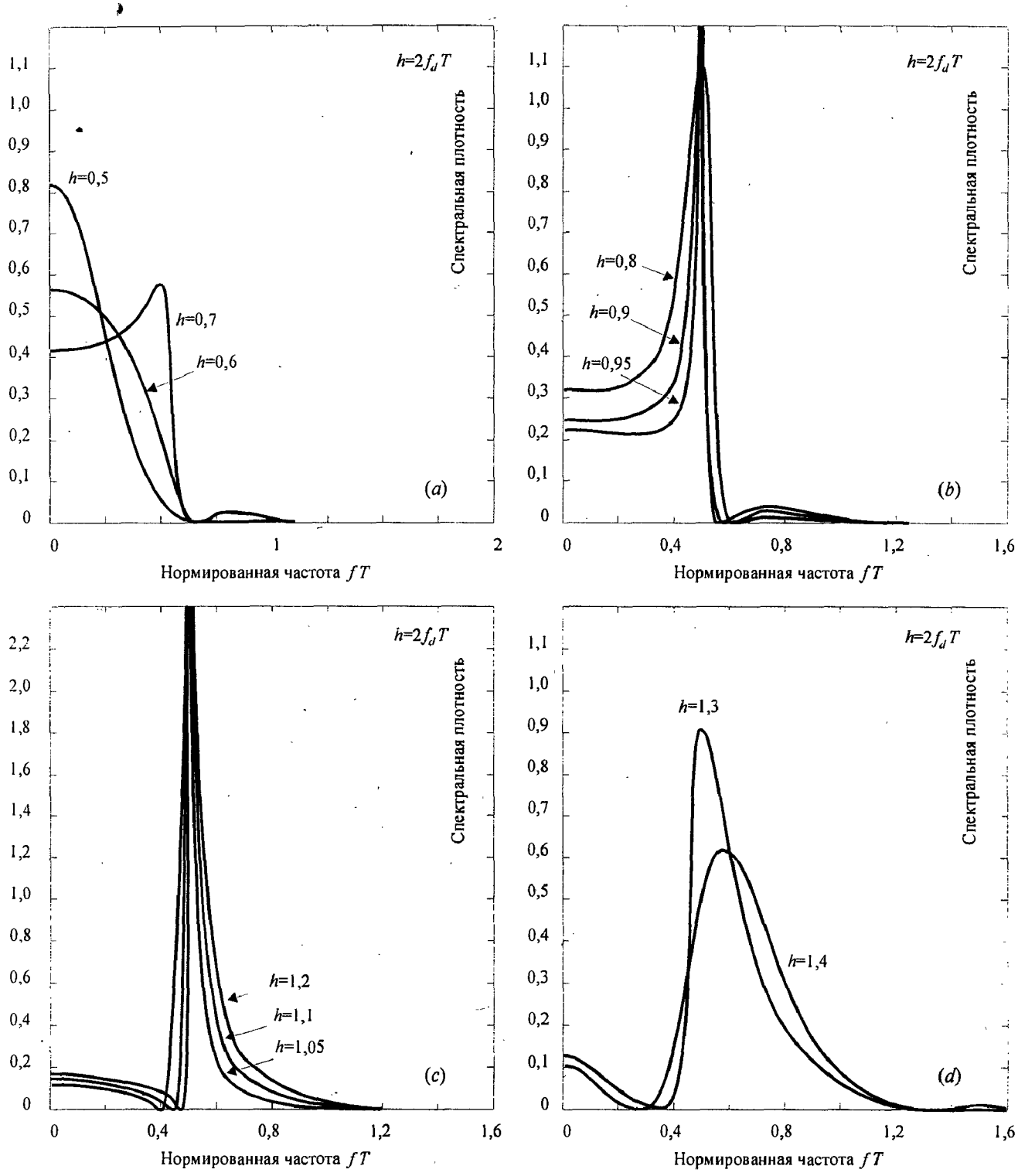


Рис. 4.4.3. Спектральная плотность мощности двоичной ЧМНФ

Графики показывают, что спектр ЧМНФ относительно узкий и хорошо ограничен при $h<1$. Когда h приближается к единице, в спектре отмечаются большие выбросы, и при $h=1$, когда $|\psi|=1$, мы находим, что пики возникают на M частотах. Когда $h>1$, спектр получается значительно шире. В системах связи, в которых используется ЧМНФ, индекс модуляции рассчитывается так, чтобы экономить полосу, так что $h<1$.

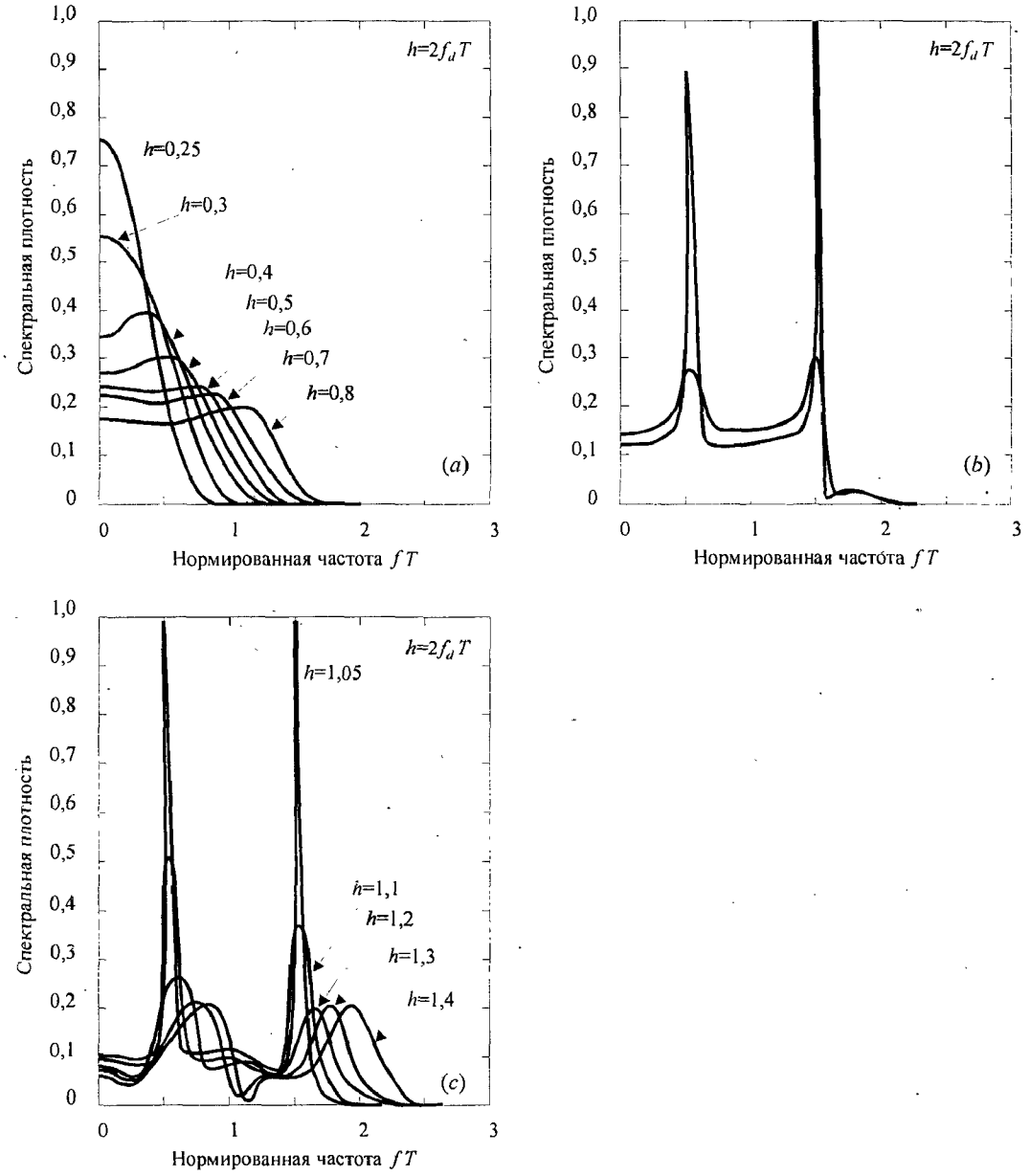


Рис. 4.4.4. Спектральная плотность мощности 4-позиционной ЧМНФ

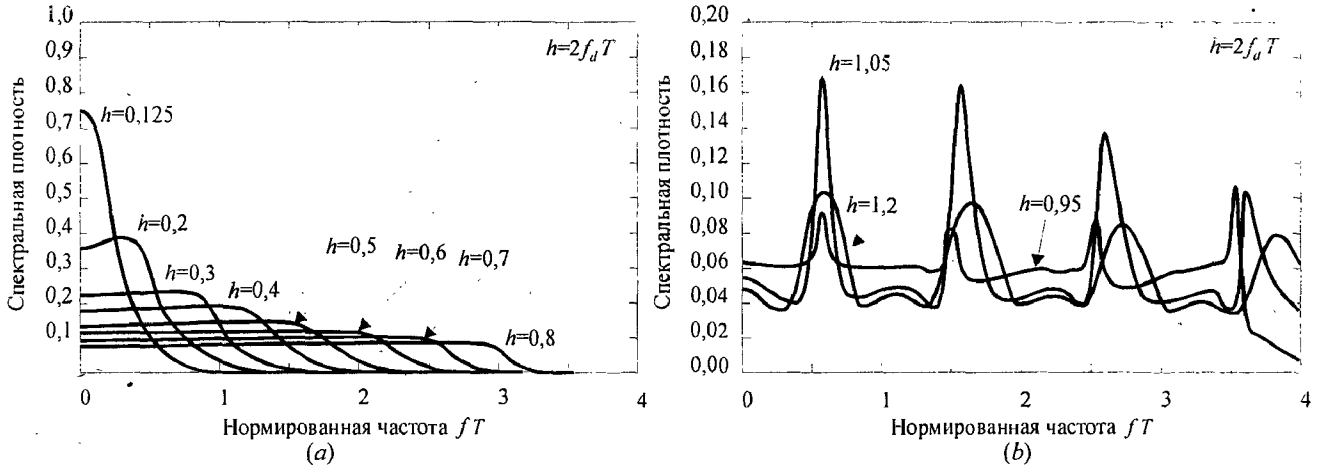


Рис. 4.4.5. Спектральная плотность мощности 8-позиционной ЧМНФ

Частный случай двоичной ЧМНФ с $h=\frac{1}{2}$ (или $f_d=1/(4T)$) и $\psi=0$ соответствует ММС. В этом случае спектр сигнала

$$\Phi_{\upsilon\upsilon}(f)=\frac{16A^2T}{\pi^2}\left(\frac{\cos 2\pi fT}{1-16f^2T^2}\right)^2. \qquad (4.4.53)$$

В (4.4.52) сигнальная амплитуда $A=1$. В противоположность этому спектр четырёхфазной офсетной (квадратурной) ФМ (ОКФМ) с прямоугольным импульсом $g(t)$ длительности T равен

$$\Phi_{\upsilon\upsilon}(f)=A^2T\left(\frac{\sin\pi fT}{\pi fT}\right)^2. \qquad (4.4.54)$$

Чтобы иметь возможность сравнить эти спектральные характеристики, мы должны нормировать частоту по битовой скорости или битовому интервалу T_b. Поскольку ММС двоичная ЧМ, то следует, что в (4.4.53) $T=T_b$. С другой стороны, для ОКФМ $T=2T_b$, так что (4.4.54) принимает вид

$$\Phi_{\upsilon\upsilon}(f)=2A^2T_b\left(\frac{\sin 2\pi fT_b}{2\pi fT_b}\right)^2. \qquad (4.4.55)$$

Спектры сигналов ММС и ОКФМ показаны на рис. 4.4.6.

Заметим, что главная доля спектра в системе ММС на 50 % шире, чем для ОКФМ. Однако боковые доли в ММС уменьшаются значительно быстрее, чем в ОКФМ. Например, если сравним полосу W, которая содержит 99 % общей мощности, найдём, что $W=1{,}2/T_b$ для ММС и $W\approx 8/T_b$ для ОКФМ. Следовательно, ММС имеет более узкую концентрацию спектра, если её оценить в долях мощности вне полосы $fT_b=1$. Графики для внеполосных долей мощности ОКФМ и ММС даны на рис. 4.4.7. Заметим, что ММС существенно эффективнее по полосе, чем ОКФМ. Эта эффективность объясняет популярность ММС во многих цифровых сетях связи.

Большую частотную эффективность, чем при ММС, можно достичь уменьшением индекса модуляции. Однако в этом случае сигналы ЧМ не будут больше ортогональными, и это приведёт к росту вероятности ошибки.

Спектральные характеристики МНФ. В общем занимаемая полоса частот зависит от выбора индекса модуляции h, формы огибающей импульса $g(t)$ и числа сигналов M. Как мы видели для ЧМНФ, малое значение h приводит к МНФ-сигналам с относительно узкой занимаемой полосой, в то время как большие значения h приводят к сигналам с большой занимаемой полосой. Здесь рассмотрим случай более общего сигнала МНФ.

Выбор гладкого импульса, такого как приподнятый косинус

$$g(t)=\begin{cases}\dfrac{1}{2LT}\left(1-\cos\dfrac{2\pi t}{LT}\right) & (0\le t\le LT),\\ 0 & (\text{для других } t),\end{cases} \qquad (4.4.56)$$

где $L=1$ для полного отклика и $L>1$ для парциального отклика, приводит к узкой занимаемой полосе и, следовательно, к большей частотной эффективности, чем при использовании прямоугольного импульса. Например, рис. 4.4.8 иллюстрирует спектральную плотность мощности для двоичной МНФ с различными парциальными откликами импульса приподнятого косинуса (LПК), когда $h=\frac{1}{2}$. Для сравнения также показаны характеристики двоичной ЧМНФ. Заметим, что с ростом L импульс $g(t)$ становится глаже, и соответственно занимаемая сигналом полоса сокращается.

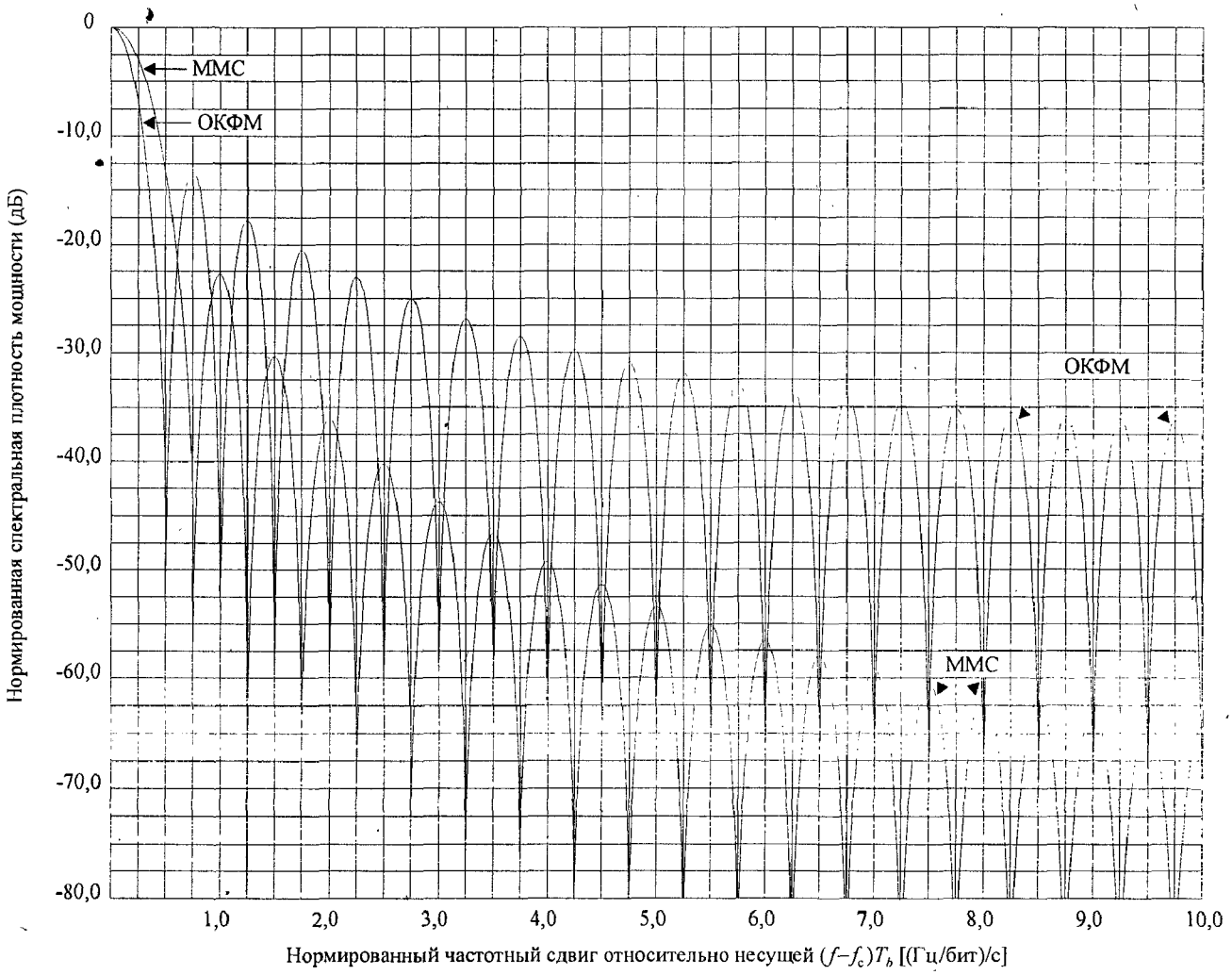


Рис. 4.4.6. Спектральная плотность мощности ММС и ОКФМ
[*Gronemeyer* и *McBride* (1976); © 1976 *IEEE*]

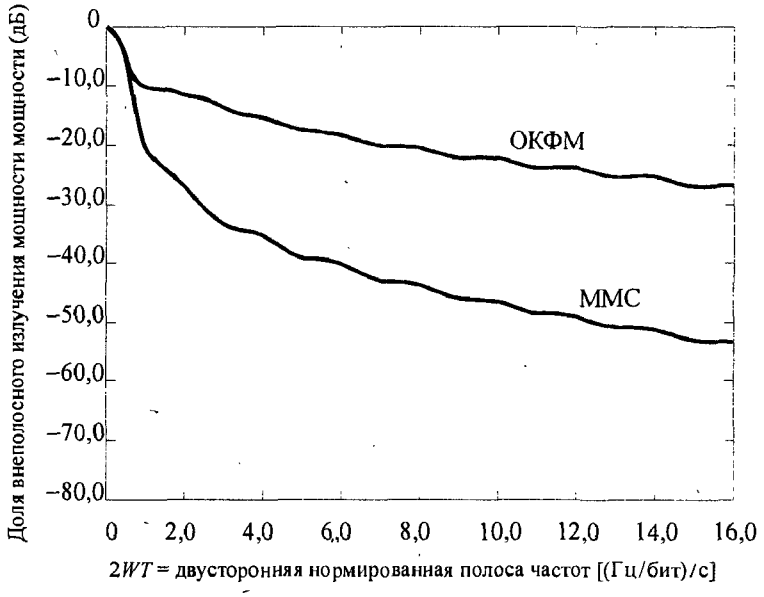


Рис. 4.4.7. Зависимость доли внеполосного излучения мощности от двусторонней нормированной полосы частот $2WT$
[*Gronemeyer* и *McBride* (1976); © 1976 *IEEE*]

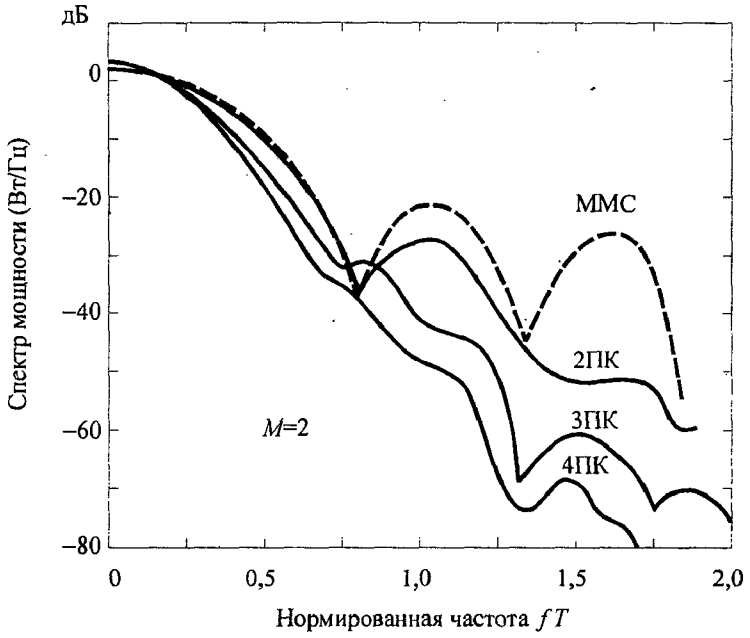


Рис. 4.4.8. Спектральная плотность мощности МНФ с h=1/2 и различной формой огибающей импульса
[*Aulin* и др. (1981); © 1981 *IEEE*]

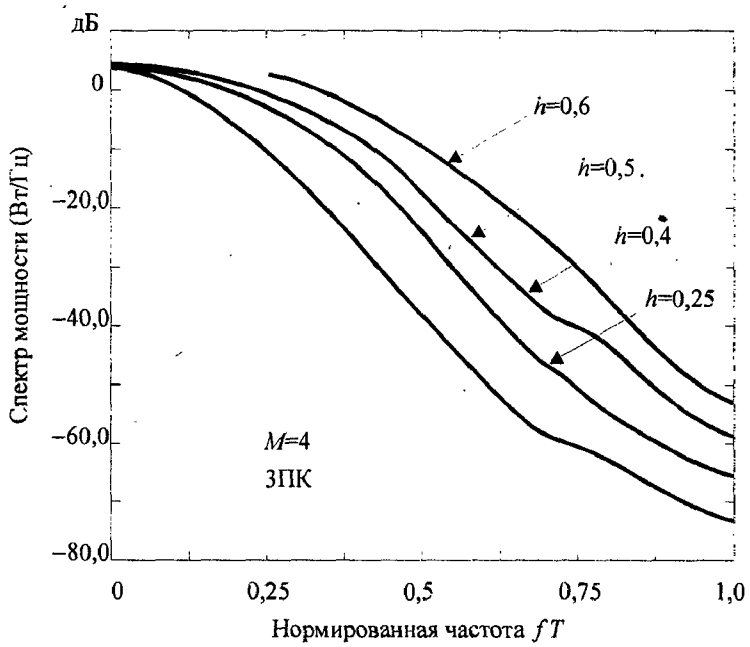


Рис. 4.4.9. Спектральная плотность мощности МНФ для M=4 с 3ПК и различными индексами модуляции
[*Aulin* и др. (1981); © 1981 *IEEE*]

Влияние изменения индекса модуляции в сигнале МНФ иллюстрируется на рис. 4.4.9 для случая $M = 4$ и импульса приподнятого косинуса формы данной (4.4.56) с $L = 3$. Заметим, эти спектральные характеристики похожи на те, которые ранее иллюстрировались для ЧМНФ, за исключением того, что этот спектр уже из-за использования гладкой огибающей импульса.

Наконец, на рис. 4.4.10 мы иллюстрируем зависимость доли внеполосной мощности от нормированной частоты для двухамплитудной ЧМНФ с несколькими различными значениями h.

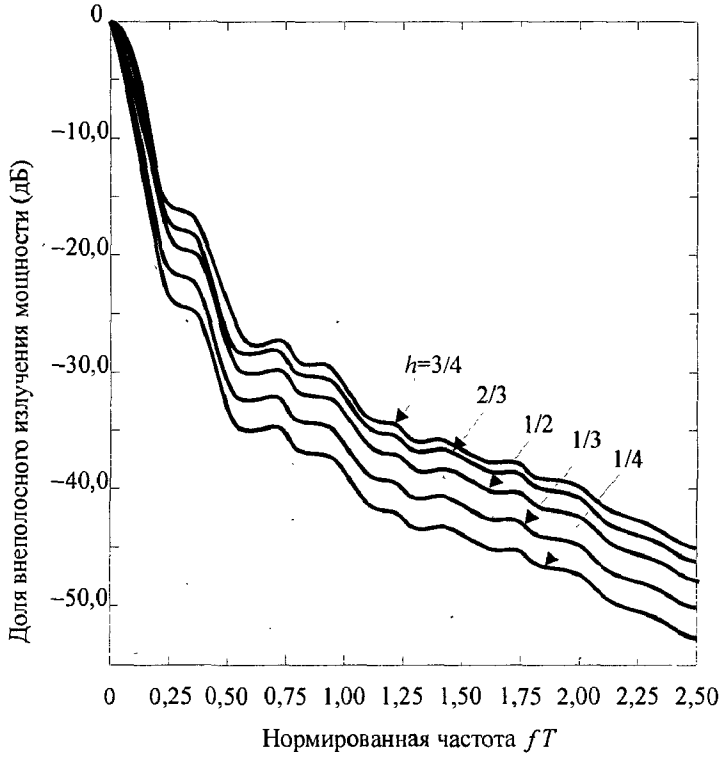


Рис. 4.4.10. Относительная величина внеполосной мощности для двухкомпонентной ЧМНФ (*Milligan*, 1988)

4.4.3. Спектр мощности для модулированных сигналов с памятью

В двух последних разделах мы определили спектральные характеристики для класса линейно модулированных сигналов без памяти и для класса модулированных по фазе сигналов, таких как ЧМНФ и МНФ, которые нелинейны и обладают памятью. В этом разделе рассмотрим спектральные характеристики линейно модулированных сигналов; которые обладают памятью и которые можно моделировать марковской цепью. Мы уже встречали такие сигналы в разд. 4.3.2, в котором описали несколько типов базовых сигналов.

Спектральную плотность мощности цифрового модулирующего сигнала, который описывается цепью Маркова, можно получить при помощи базовой процедуры, данной в предыдущих разделах. Сначала определяется автокорреляционная функция, а затем с помощью преобразования Фурье находится спектральная плотность мощности. Для сигналов, которые описываются цепью Маркова с матрицей переходных вероятностей **P**, спектральную плотность мощности сигнала можно выразить в общем виде (см. Тичворт и Велч, 1961):

$$\Phi(f)=\frac{1}{T^2}\sum_{n=-\infty}^{\infty}\left|\sum_{i=1}^{K}p_i S_i\left(\frac{n}{T}\right)\right|^2\delta\left(f-\frac{n}{T}\right)+\frac{1}{T}\sum_{i=1}^{K}p_i\left|S_i'(f)\right|^2+$$
$$+\frac{2}{T}\mathrm{Re}\left[\sum_{i=1}^{K}\sum_{j=1}^{K}p_i S_i'^{*}(f)S_j'(f)P_{ij}(f)\right], \qquad (4.4.57)$$

где $S_i(f)$ – преобразование Фурье для сигнала $s_i(t)$,

$$s_i'(t)=s_i(t)-\sum_{k=1}^{K}p_k s_k(t),$$

$P_{ij}(f)$ – преобразование Фурье дискретной во времени последовательности $p_{ij}(n)$, определенное выражением

$$P_{ij}(f)=\sum_{n=1}^{\infty}p_{ij}(n)\mathrm{e}^{-j2\pi nfT},\qquad(4.4.58)$$

K – номер состояния модулятора. Слагаемое $p_{ij}(n)$ определяет вероятность того, что сигнал $s_j(t)$ передаётся в n-м сигнальном интервале после передачи сигнала $s_i(t)$. Таким образом, $\{p_{ij}(n)\}$ являются вероятностями переходов в матрице вероятностей переходов $\mathbf{P}^n$. Заметим, что $p_{ij}(1)=p_{ij}$.

Если метод модуляции без памяти, то переданный в каждом сигнальном интервале сигнал не зависит от сигналов, переданных в предыдущих сигнальных интервалах. Спектральная плотность мощности результирующего сигнала в этом случае можно всё ещё выразить в виде (4.4.57), если матрицу переходных вероятностей заменить на

$$\mathbf{P}=\begin{bmatrix}p_1 & p_2 & \cdots & p_K\\ p_1 & p_2 & \cdots & p_K\\ \vdots & \vdots & \cdots & \vdots\\ p_1 & p_2 & \cdots & p_K\end{bmatrix}\qquad(4.4.59)$$

и навязать условие $\mathbf{P}^n=\mathbf{P}$ для всех $n\geq 1$. Только при этих условиях выражение для спектральной плотности мощности оказывается функцией стационарных вероятностей состояний $\{p_i\}$ и, следовательно, ведёт к простейшей форме

$$\Phi(f)=\frac{1}{T^2}\sum_{n=-\infty}^{\infty}\left|\sum_{i=1}^{K}p_iS_i\left(\frac{n}{T}\right)\right|^2\delta\left(f-\frac{n}{T}\right)+\frac{1}{T}\sum_{i=1}^{K}p_i(1-p_i)\left|S_i(f)\right|^2+$$
$$-\frac{2}{T}\sum_{\substack{i=1\\ i<j}}^{K}\sum_{j=1}^{K}p_ip_j\,\mathrm{Re}\left[S_i(f)S_j^*(f)\right].\qquad(4.4.60)$$

Видно, что наш предыдущий результат для спектральной плотности мощности линейной модуляции без памяти, определяемый (4.4.18), можно рассматривать как частный случай (4.4.60), в котором все сигналы идентичны, за исключением ряда скалярных множителей, которые передают цифровую информацию (задача 4.30).

Мы также видим, что первое слагаемое в выражении для спектральной плотности мощности (4.4.57) или (4.4.60) состоит из дискретных частотных компонент. Этот линейчатый спектр исчезает, когда

$$\sum_{i=1}^{K}p_iS_i\left(\frac{n}{T}\right)=0.\qquad(4.4.61)$$

Условие (4.4.61) обычно навязывается для практических систем связи и легко удовлетворяется подходящим выбором форм сигнала (задача 4.31).

Теперь определим спектральную плотность мощности базовых модулирующих сигналов, описанных в разд. 4.3.2. Сначала рассмотрим NRZ-сигнал, который характеризуется двумя сигналами $s_1(t)=g(t)$ и $s_2(t)=-g(t)$, где $g(t)$ – прямоугольный импульс амплитуды A. Для $K=2$ (4.4.60) даёт

$$\Phi(f)=\frac{(2p-1)^2}{T^2}\sum_{n=-\infty}^{\infty}\left|G\left(\frac{n}{T}\right)\right|^2\delta\left(f-\frac{n}{T}\right)+\frac{4p(1-p)}{T}\left|G(f)\right|^2,\qquad(4.4.62)$$

где

$$\left|G(f)\right|^2=(AT)^2\left(\frac{\sin\pi fT}{\pi fT}\right)^2.\qquad(4.4.63)$$

Заметим, что, когда $p=\frac{1}{2}$; линейчатый спектр исчезает и $\Phi(f)$ определяется так:

$$\Phi(f)=\frac{1}{T}|G(f)|^2. \tag{4.4.64}$$

NRZI-сигнал характеризуется матрицей переходных вероятностей

$$\mathbf{P}=\begin{bmatrix}\frac{1}{2} & \frac{1}{2}\\ \frac{1}{2} & \frac{1}{2}\end{bmatrix}. \tag{4.4.65}$$

Заметим, что в этом случае $\mathbf{P}^n=\mathbf{P}$ для всех $n\geq 1$. Следовательно, частная форма спектра плотности мощности, даваемая (4.4.62), хорошо подходит к модуляции по этому формату. Следовательно, спектральная плотность мощности NRZI-сигнала идентична спектру NRZ-сигнала.

Модуляция с задержкой имеет матрицу переходных вероятностей

$$\mathbf{P}=\begin{bmatrix}0 & \frac{1}{2} & 0 & \frac{1}{2}\\ 0 & 0 & \frac{1}{2} & \frac{1}{2}\\ \frac{1}{2} & \frac{1}{2} & 0 & 0\\ \frac{1}{2} & \frac{1}{2} & 0 & 0\end{bmatrix} \tag{4.4.66}$$

и стационарные вероятности состояний $p_i=\frac{1}{4}$ для $i=1, 2, 3, 4$. Степени $\mathbf{P}$ можно получить путём использования соотношения

$$\mathbf{P}^4\rho=-\tfrac{1}{4}\rho, \tag{4.4.67}$$

где ρ – матрица корреляции сигнала с элементами

$$\rho_{ij}=\frac{1}{T}\int_0^T s_i(t)s_j(t)\,dt, \tag{4.4.68}$$

а четыре сигнала $\{s_i(t), i=1, 2, 3, 4\}$ показаны на рис. 4.3.15. Легко видеть, что

$$\rho=\begin{bmatrix}1 & 0 & 0 & -1\\ 0 & 1 & -1 & 0\\ 0 & -1 & 1 & 0\\ -1 & 0 & 0 & 1\end{bmatrix}. \tag{4.4.69}$$

Следовательно, степени $\mathbf{P}$ можно получить из соотношения

$$\mathbf{P}^{k+4}\rho=-\tfrac{1}{4}\mathbf{P}^k\rho, \qquad k>1. \tag{4.4.70}$$

Используя (4.4.66), (4.4.69) и (4.4.70), в (4.4.57) можно найти спектральную плотность мощности при модуляции с задержкой. Её можно выразить в форме

$$\Phi(f)=\frac{1}{2\psi^2(17+8\cos 8\psi)}\left[23-2\cos\psi-22\cos 2\psi-12\cos 3\psi+\right.$$
$$\left.+5\cos 4\psi+12\cos 5\psi+2\cos 6\psi-8\cos 7\psi+2\cos 8\psi\right], \tag{4.4.71}$$

где $\psi=\pi f T$.

Спектр этих базовых модулирующих сигналов показан на рис. 4.4.11.

Видно, что спектр сигналов NRZ и NRZI имеет максимум при $f=0$. Модуляция с задержкой имеет более узкий спектр и относительно меньший уровень для нулевых частот. Занимаемая ею полоса частот существенно уже, чем у сигнала NRZ. Эти две характеристики делают модуляцию с задержкой привлекательным выбором для каналов, которые не пропускают постоянную составляющую, таких, как средства магнитной записи.

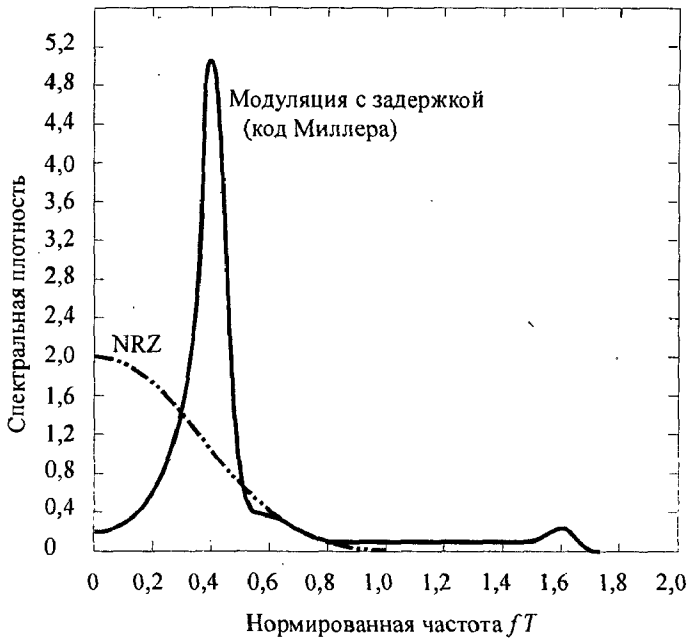


Рис. 4.4.11. Односторонняя спектральная плотность мощности для базовых сигналов кода Миллера (модуляция с задержкой) и NRZ/NRZI [*Hecht* и *Guida* (1969); © 1969 *IEEE*]

4.5. БИБЛИОГРАФИЧЕСКИЕ ЗАМЕЧАНИЯ И ССЫЛКИ

Характеристики сигналов и систем, данные в этой главе, очень полезны при конструировании техники оптимальной модуляции-демодуляции и кодирования-декодирования для различных моделей канала. В частности, методы цифровой модуляции, изложенные в этой главе, широко используются в цифровой связи. Следующая глава посвящена технике оптимальной демодуляции для таких сигналов и их характеристикам качества при аддитивном белом гауссовском шуме в канале. Общее освещение характеристик сигналов имеется в книге Френкса (1969).

Особенно важными при проектировании цифровых систем связи являются спектральные характеристики цифровых модулированных сигналов, которые представлены в этой главе с определённой глубиной. Из этой техники модуляции одна из наиболее важных – МНФ с учётом того, что она эффективно использует полосу частот. Из этих соображений она широко изучалась многими исследователями, и в технической литературе появилось большое число публикаций по этой теме. Наиболее исчерпывающие обсуждения МНФ, включая характеристики качества и спектральные характеристики, можно найти в книге Андерсена и др. (1986). В дополнение к этому материалу учебник Сандберга (1986) представляет базовые концепции и обзор характеристик качества различной техники МНФ. Эта публикация также содержит около 100 ссылок на опубликованные статьи по этой теме. Имеется большое число ссылок, связанных со спектральными характеристиками ЧМНФ и МНФ. Для начала поиска упомянем, что ММС была изобретена Дольцем и Хелдом в 1961. Ранние работы по спектральной плотности мощности для ЧМНФ и МНФ были сделаны Беннетом и Райсом (1963), Андерсоном и Сальцем (1965) и Беннетом и Давеем (1965). Книга Лакки и др.(1968) также содержит трактовку спектральных характеристик ЧМНФ. Большинство из недавних новых работ имеется в публикации Сандберга (1986). Мы должны также процитировать специальные исследования по частотной эффективности модуляции и кодирования, опубликованные IEEE Transactions on Communication (март 1981). Там имеются несколько статей по спектральным характеристикам и характеристикам качества МНФ.

Обобщение ММС на много амплитуд было исследовано Вебером и др. (1978). Комбинация многих амплитуд с общей системой МНФ была предложена Маллиганом (1988), который исследовал их спектральные характеристики и их характеристики качества (вероятности ошибки) в гауссовском шуме при отсутствии и наличии кодирования.

ЗАДАЧИ

4.1. Докажите следующие свойства преобразования Гильберта:

a) Если $x(t) = x(-t)$, тогда $\hat{x}(t) = -\hat{x}(-t)$;

b) Если $x(t) = -x(-t)$, тогда $\hat{x}(t) = \hat{x}(-t)$;

c) Если $x(t) = \cos\omega_0 t$, тогда $\hat{x}(t) = \sin\omega_0 t$;

d) Если $x(t) = \sin\omega_0 t$, тогда $\hat{x}(t) = -\cos\omega_0 t$;

e) $\hat{\hat{x}}(t) = -x(t)$;

f) $\int_{-\infty}^{\infty} x^2(t)\,dt = \int_{-\infty}^{\infty} \hat{x}^2(t)\,dt$;

g) $\int_{-\infty}^{\infty} x(t)\hat{x}(t)\,dt = 0$.

4.2. Если $X(t)$ – стационарный случайный процесс с автокорреляционной функцией $\phi_{xx}(\tau) = E\left[X(t)X(t+\tau)\right]$ и спектральной плотностью мощности $\Phi_{xx}(f)$, тогда покажите, что $\phi_{\hat{x}\hat{x}}(\tau) = \phi_{xx}(\tau)$, $\phi_{x\hat{x}}(\tau) = -\phi_{xx}(\tau)$ и $\Phi_{\hat{x}\hat{x}}(f) = \Phi_{xx}(f)$.

4.3. Предположим, что $N(t)$ – узкополосный стационарный случайный процесс с нулевым средним, представленный (4.1.37), (4.1.38) или (4.1.39). Автокорреляционная функция эквивалентного низкочастотного процесса $Z(t) = X(t) + jY(t)$ определяется так:

$$\Phi_{zz}(\tau) = \tfrac{1}{2} E\left[Z^*(t)Z(t+\tau)\right].$$

a) Покажите, что

$$E\left[Z(t)Z(t+\tau)\right] = 0.$$

b) Предположим, что $\phi_{zz}(\tau) = N_0\delta(\tau)$, и пусть

$$V = \int_0^T Z(t)\,dt.$$

Определите $E\left(V^2\right)$ и $E\left(VV^*\right) = E\left(|V|^2\right)$.

4.4. Определите автокорреляционную функцию случайного процесса

$$X(t) = A\sin(2\pi f_c t + \theta),$$

где f_c – константа, а θ – равномерно распределённая случайная фаза, т.е.

$p(\theta) = 1/2\pi$, $0 \le \theta \le 2\pi$.

4.5. Докажите, что $s_l(t)$ – в общем случае комплексный сигнал, и определите условия, когда он вещественен. Предположите, что $s(t)$ – вещественный полосовой сигнал.

4.6. Предположите, что $s(t)$ или вещественный, или комплексный сигнал, который представлен приближённо линейной комбинацией ортогональных функций $\{f_n(t)\}$, т.е.

$$\hat{s}(t) = \sum_{k=1}^{K} s_k f_k(t),$$

где

$$\int_{-\infty}^{\infty} f_n(t) f_m^*(t)\,dt = \begin{cases} 0 & (m \ne n), \\ 1 & (m = n). \end{cases}$$

Определите коэффициенты $\{s_k\}$ в выражении $\hat{s}(t)$ так, чтобы минимизировать энергию

$$\mathscr{E}_e = \int_{-\infty}^{\infty} |s(t) - \hat{s}(t)|^2 dt$$

и соответствующую остаточную ошибку $\mathscr{E}_e$.

4.7. Предположите, что имеется ансамбль из M комплексных сигналов $\{s_{lm}(t)\}$. Получите уравнения для процедуры Грама-Шмидта, которые приводят к ансамблю $N \le M$ ортонормированных сигналов.

4.8. Определите коэффициенты корреляции ρ_{km} четырёх сигналов $\{s_i(t)\}$, показанных на рис. 4.2.1, и соответствующие расстояния Евклида.

4.9. Рассмотрите ансамбль M ортогональных сигналов $\{s_m(t)\}$, $1 \le m \le M$, $0 \le t \le T$, каждый из которых имеет одинаковую энергию $\mathscr{E}$. Найдите новый ансамбль из M сигналов так, чтобы

$$s'_m(t) = s_m(t) - \frac{1}{M}\sum_{k=1}^{K} s_k(t), \qquad 1 \le m \le M, \quad 0 \le t \le T.$$

Покажите, что M сигналов $\{s'_m(t)\}$ имеют равную энергию, определяемую так:

$$\mathscr{E}' = (M-1)\mathscr{E}/M,$$

и они одинаково коррелированы с коэффициентом корреляции

$$\rho_{mn} = \frac{1}{\mathscr{E}'}\int_0^T s'_m(t)s'_n(t)dt = -\frac{1}{M-1}.$$

4.10. Рассмотрите три сигнала $f_n(t)$, показанных на рис. P4.10.

a) Покажите, что эти сигналы ортонормированы.

b) Выразите сигнал $x(t)$ как взвешенную линейную комбинацию $f_n(t)$, $n = 1, 2, 3$, если

$$x(t) = \begin{cases} -1 & (0 \le t < 1), \\ 1 & (1 \le t < 3), \\ -1 & (3 \le t < 4), \end{cases}$$

и определите коэффициенты взвешивания.

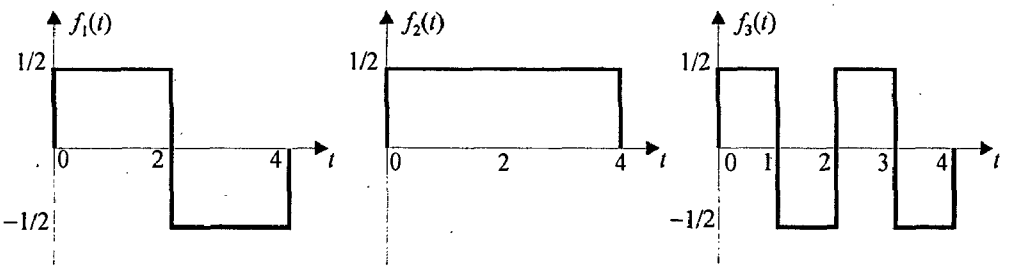


Рис. P4.10

4.11. Рассмотрите четыре формы сигналов, показанных на рис. P4.11.

a) Определите размерность сигналов и ансамбль базисных функций.

b) Используйте базисные функции для представления четырёх сигналов векторами $\mathbf{s}_1$, $\mathbf{s}_2$, $\mathbf{s}_3$ и $\mathbf{s}_4$.

c) Определите минимальное расстояние между любой парой векторов.

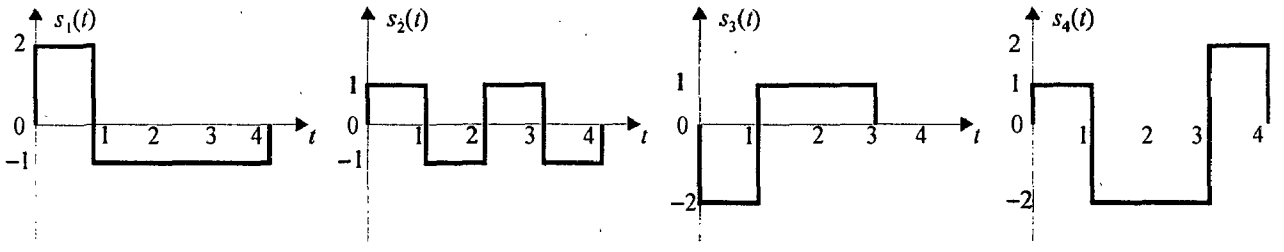


Рис. P4.11

4.12. Определите ансамбль ортогональных функций для четырёх сигналов, показанных на рис. P4.12.

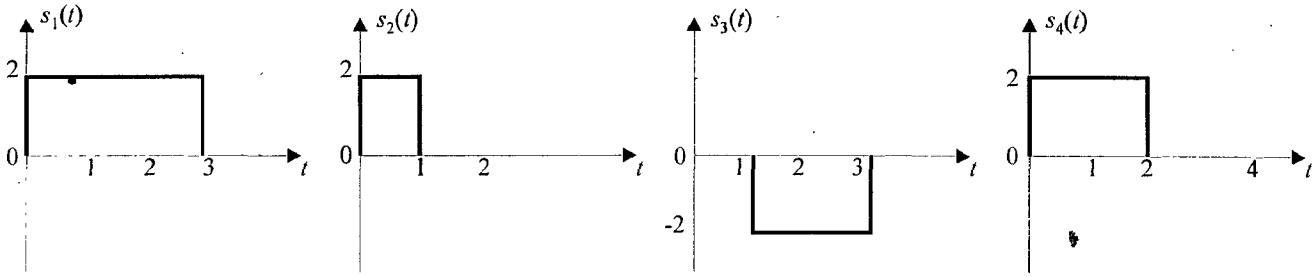


Рис. P4.12

4.13. Низкочастотный гауссовский случайный процесс $X(t)$ имеет спектральную плотность мощности

$$\Phi(f) = \begin{cases} N_0 & (|f| < B), \\ 0 & (|f| > B). \end{cases}$$

Определите спектральную плотность мощности и автокорреляционную функцию случайного сигнала $Y(t) = X^2(t)$.

4.14. Рассмотрите эквивалентный низкочастотный модулирующий сигнал в виде

$$u(t) = \sum_n \left[a_n g(t - 2nT) - jb_n g(t - 2nT - T)\right],$$

где $\{a_n\}$ и $\{b_n\}$ – две последовательности статистически независимых двоичных символов, а $g(t)$ – синусоидальный импульс, который определяется так:

$$g(t) = \begin{cases} \sin(\pi t/2T) & (0 < t < 2T), \\ 0 & \text{для других } t. \end{cases}$$

Этот вид сигнала рассматривается как четырёхфазовая ФМ, причём огибающая импульса $g(t)$ составляет полпериода синусоиды. Каждая из информационных последовательностей $\{a_n\}$ и $\{b_n\}$ передаётся со скоростью $1/2T$ бит/с, и, следовательно, общая скорость передачи равна $1/T$. Две последовательности синхронизированы во времени для передачи с задержкой на интервале T. Как следствие, сигнал $u(t)$ назван сигналом четырёхфазовой ФМ со сдвигом.

a) Докажите, что огибающая $|u(t)|$ – константа, независимо от информационных символов a_n в синфазной компоненте и информационных символов b_n в квадратурной компоненте. Другими словами, амплитуда несущей, используемая для передачи, постоянна.

b) Определите спектральную плотность мощности $u(t)$.

c) Сравните спектральную плотность мощности, полученную в (b), со спектральной плотностью мощности сигнала ММС. Какое заключение Вы можете сделать из этих сравнений?

4.15. Рассмотрите сигнал четырёхуровневой ФМ, представленной эквивалентным низкочастотным сигналом

$$u(t) = \sum_n \left[I_n g(t - 2nT)\right],$$

где I_n принимает одно из четырёх возможных значений $\sqrt{\frac{1}{2}}(\pm 1 \pm j)$ с равной вероятностью. Последовательность информационных символов $\{I_n\}$ статистически независима.

a) Определите и нарисуйте спектральную плотность мощности $u(t)$, когда

$$g(t) = \begin{cases} A & (0 \le t \le T), \\ 0 & \text{для других } t. \end{cases}$$

b) Повторите (a), когда

$$g(t) = \begin{cases} A\sin(\pi t/2T) & (0 \le t \le T), \\ 0 & \text{для других } t. \end{cases}$$

c) Сравните ширины полос спектров, полученных в (a) и (b), на уровне ослабления в 3 дБ и по полосе, определяемой первым нулём.

4.16. Случайный процесс $V(t)$ определён так:

$$V(t) = X\cos 2\pi f_c t - Y\sin 2\pi f_c t\,,$$

где X и Y – случайные величины. Покажите, что процесс $V(t)$ стационарен в широком смысле, если, и только если $E(X) = E(Y) = 0$, $E(X^2) = E(Y^2)$ и $E(XY) = 0$

4.17. Выполните ортогонализацию по Граму–Шмидту сигналов рис. 4.2.1 (a) в порядке $s_4(t)$, $s_3(t)$, $s_1(t)$ и затем получите ансамбль ортогональных функций $\{f_m(t)\}$. Определите векторные представления сигналов $\{s_n(t)\}$, используя ортонормированные функции $\{f_m(t)\}$. Определите также энергии сигналов.

4.18. Определите представление в пространстве сигналов четырёх сигналов $s_k(t)$, $k = 1, 2, 3, 4$, показанных на рис. Р4.18, используя ортонормальные базисные функции $f_1(t)$ и $f_2(t)$. Нарисуйте диаграмму пространства состояний для четырёх сигналов и покажите, что этот ансамбль сигналов эквивалентен ансамблю четырёхфазовой ФМ.

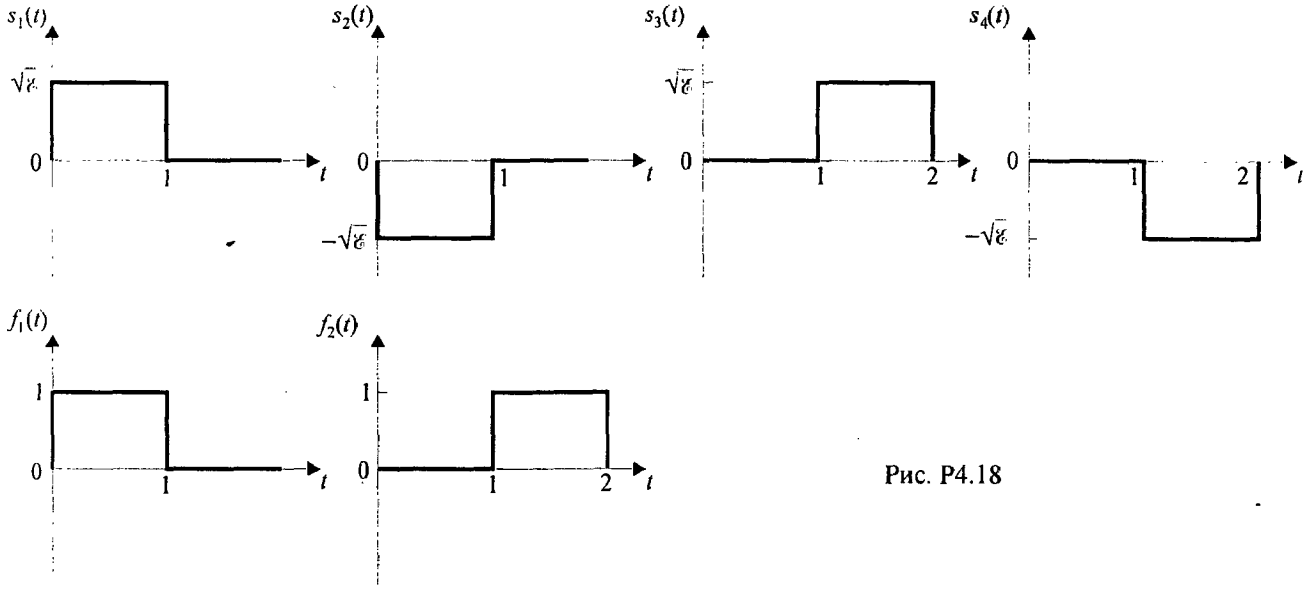


Рис. Р4.18

4.19. Спектральная плотность мощности циклостационарного случайного процесса

$$V(t) = \sum_{n=-\infty}^{\infty} I_n g(t - nT)$$

была получена в разд. 4.4.1 путём усреднения автокорреляционной функции $\phi_{vv}(t+\tau, t)$ за период T процесса, а затем вычислено преобразование Фурье от усреднённой корреляционной функции. Альтернативный подход заключается в превращении циклостационарного процесса в стационарный процесс $V_\Delta(t)$ путём добавления случайной величины Δ, равномерно распределённой на интервале $0 \le \Delta < T$, так что

$$V_\Delta(t) = \sum_{n=-\infty}^{\infty} I_n g(t - nT - \Delta)\,,$$

и нахождении спектральной плотности для $V(t)$ как преобразования Фурье автокорреляционной функции стационарного процесса $V_\Delta(t)$. Получите результат (4.4.11) путём вычисления автокорреляционной функции $V_\Delta(t)$ и её преобразования Фурье.

4.20. Сигнал АМ с парциальным откликом генерируется так, как показано на рис. Р4.20, – путём возбуждения идеального ФНЧ с полосой W последовательностью

$$B_n = I_n + I_{n-1}$$

со скоростью $1/T = 2W$ символов/с. Последовательность $\{I_n\}$ состоит из двоичных символов, выбираемых независимо из алфавита $\{1, -1\}$ с равной вероятностью. Следовательно, профильтрованный сигнал $V(t)$ имеет вид

$$V(t) = \sum_{n=-\infty}^{\infty} B_n g(t - nT), \quad T = \frac{1}{2W}.$$

a) Нарисуйте диаграмму пространства состояний сигнала $V(t)$ и определите вероятность появления каждого символа.

b) Определите автокорреляционную функцию и спектральную плотность мощности трёхуровневой последовательности $\{B_n\}$.

c) Сигнальные отсчёты последовательности $\{B_n\}$ образуют цепь Маркова. Нарисуйте эту марковскую цепь и укажите вероятности перехода отдельных состояний.

4.21. Эквивалентный низкочастотный сигнал АМ можно записать в виде

$$U(t) = \sum_{n} I_n g(t - nT).$$

Предположим, что $g(t)$ является прямоугольным импульсом, а

$$I_n = a_n - a_{n-2},$$

где $\{a_n\}$ – последовательность некоррелированных двоичных случайных величин $\{1, -1\}$, которые возникают с равной вероятностью.

a) Определите автокорреляционную функцию последовательности $\{I_n\}$.

b) Определите спектральную плотность мощности $U(t)$.

c) Повторите (b) для случая, когда $\{a_n\}$ принимает значения (0, 1).

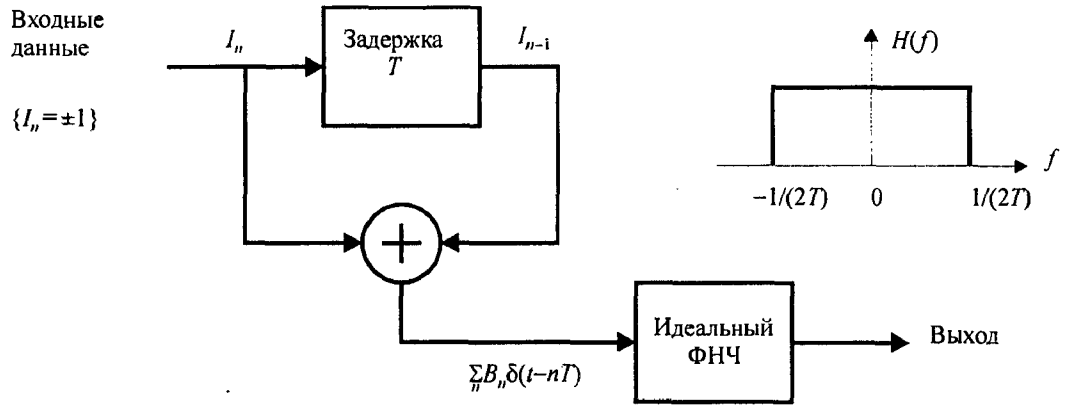


Рис. Р4.20

4.22. Покажите, что $x(t) = s(t)\cos 2\pi f_c t \pm \hat{s}(t)\sin 2\pi f_c t$ является однополосным сигналом, где сигнал $s(t)$ ограничен полосой $B \le f_c$, а $\hat{s}(t)$ – его преобразование по Гильберту.

4.23. Используйте результаты полученные в разд. 4.4.3 для того, чтобы определить спектральную плотность мощности сигнала двоичной МЧС, определяемые так

$$s_i(t) = \sin \omega_i t, \quad i = 1, 2, \quad 0 \le t \le T,$$

где $\omega_1 = n\pi/T$ и $\omega_2 = m\pi/T$, $n \neq m$, а m и n – произвольные положительные целые числа. Предположите, что $p_1 = p_2 = \frac{1}{2}$. Нарисуйте спектр и сравните этот результат со спектром сигнала ММС.

4.24. Используйте результаты, полученные в разделе 4.4.3, для того чтобы определить спектральную плотность мощности сигнала многоуровневой ЧМ, определяемого так:

$$s_n(t) = \sin \frac{2\pi nt}{T}, \quad i = 1, 2, ..., M, \quad 0 \le t \le T.$$

Предположите, что вероятности $p_i = 1/M$ для всех i. Нарисуйте график спектральной плотности мощности.

4.25. Квадратурный сигнал с парциальным откликом (КСПО) генерируется двумя отдельными сигналами с парциальным откликом вида, описанного в задаче 4.20. Следовательно, КСПО представляется так:

$$s(t) = \mathrm{Re}\left[\upsilon(t)e^{j2\pi f_c t}\right],$$

где

$$\upsilon(t) = \upsilon_c(t) + j\upsilon_s(t) = \sum_n B_n u(t - nT) + j\sum_n C_n u(t - nT),$$

$$B_n = I_n + I_{n-1},\ C_n = J_n + J_{n-1}.$$

Последовательности $\{B_n\}$ и $\{C_n\}$ не коррелированы, и $I_n = \pm 1$, $J_n = \pm 1$ с равной вероятностью.

a) Нарисуйте диаграмму пространства символов для КСПО и определите вероятность появления каждого символа.

b) Нарисуйте модель марковской цепи и укажите переходные вероятности для КСПО.

c) Определите автокорреляционную функцию и спектральную плотность мощности для $\upsilon_c(t)$, $\upsilon_s(t)$ и $\upsilon(t)$.

4.26. Определите автокорреляционные функции для модулированных сигналов ММС и КФМС, основываясь на предположении, что информационные последовательности для каждого из двух сигналов некоррелированы и с нулевым средним.

4.27. Постройте фазовое дерево, решётки состояний для МНФ с парциальным откликом с $h = \frac{1}{2}$ и

$$u(t) = \begin{cases} 1/4T & (0 \le t \le 2T), \\ 0 & (\text{для других } t). \end{cases}$$

4.28. Определите число конечных фазовых состояний на диаграмме в решётке состояний для двоичной ЧМНФ с полным откликом, когда $h = \frac{2}{3}$, $\frac{3}{4}$; в двоичной ЧМНФ с парциальным откликом и $L = 3$ при $h = \frac{2}{3}$, $\frac{3}{4}$.

4.29. Убедитесь, что 16 QAM можно представить как суперпозицию двух четырёхфазовых сигналов с постоянной огибающей, где каждая компонента отдельно усиливается до сложения, т.е.

$$s(t) = G\left[A_n \cos 2\pi f_c t + B_n \sin 2\pi f_c t\right] + \left[C_n \cos 2\pi f_c t + D_n \sin 2\pi f_c t\right],$$

где $\{A_n\}$, $\{B_n\}$, $\{C_n\}$ и $\{D_n\}$ – статистически независимые двоичные последовательности с элементами из ряда $\{+1, -1\}$, а G – коэффициент усиления. Затем покажите, что результирующий сигнал эквивалентен сигналу

$$s(t) = I_n \cos 2\pi f_c t + Q_n \sin 2\pi f_c t,$$

и определите I_n и Q_n через A_n, B_n, C_n и D_n.

4.30. Используйте результат (4.4.60), чтобы получить выражение для спектральной плотности мощности при линейной модуляции без памяти, определите (4.4.18) при условии, что

$$s_k(t) = I_k s(t),\quad k = 1, 2, \ldots, K,$$

где I_k – один из K возможных передаваемых символов, которые появляются с равными вероятностями.

4.31. Убедитесь, что достаточное условие отсутствия дискретных компонент спектра в (4.4.60) – это

$$\sum_{i=1}^{K} p_i s_i(t) = 0.$$

Является ли условие необходимым? Объясните ваш ответ.

4.32. Информационная последовательность $\{a_n\}_{n=-\infty}^{\infty}$ является последовательностью случайных величин, каждая из которых принимает значение +1 и –1 с равной вероятностью. Эта последовательность передаётся посредством базового модулирующего сигнала при помощи двухфазной схемы кодирования и определяется так:

$$s(t) = \sum_{n=-\infty}^{\infty} a_n g(t - nT),$$

а сигнал $g(t)$ показан на рис. Р4.32.

a) Найдите спектральную плотность мощности сигнала $s(t)$.

b) Предположите, что желательно иметь нуль в спектре мощности на частоте $f = 1/T$. С этой целью используйте схему предварительного кодирования $b_n = a_n + ka_{n-1}$, где k – некоторая постоянная, и далее

передайте последовательность $\{b_n\}$, используя тот же сигнал $g(t)$. Можно ли выбрать k так, чтобы образовать нуль на частоте $f = 1/T$? Если да, какова желательная величина k и каков результирующий спектр мощности?

c) Теперь предположите, что мы хотим иметь нуль на всех частотах, кратных $f_0 = 1/4T$. Возможно ли иметь эти нули при подходящем выборе k из предыдущей задачи? Если нет, какую схему предварительного кодированиия можете предложить, чтобы всё же получить требуемые нули?

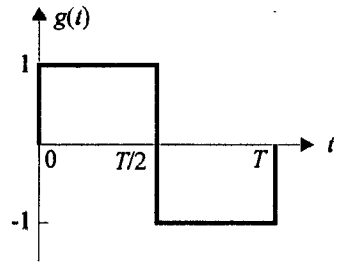


Рис. P4.32

4.33. Начиная с определения матрицы переходных вероятностей для модуляции с задержкой, данной (4.4.66), покажите, что соотношение

$$\mathbf{P}^4\rho = -\tfrac{1}{4}\rho$$

имеет место и, следовательно,

$$\mathbf{P}^{k+4}\rho = -\tfrac{1}{4}\mathbf{P}^k\rho, \qquad k \geq 1.$$

4.34. Два сигнала для передачи ЧМ с разрывом фазы определяются так:

$$s_0(t) = \sqrt{\frac{2\mathcal{E}_b}{T_b}}\cos\left[2\pi\left(f - \frac{\Delta f}{2}\right)t + \theta_0\right], \quad 0 \leq t < T,$$

$$s_0(t) = \sqrt{\frac{2\mathcal{E}_b}{T_b}}\cos\left[2\pi\left(f + \frac{\Delta f}{2}\right)t + \theta_1\right], \quad 0 \leq t \leq T,$$

где $\Delta f = 1/T << f_c$ и θ_0 и θ_1 – равномерно распределённые случайные величины на интервале $(0, 2\pi)$. Сигналы $s_0(t)$ и $s_1(t)$ имеют одинаковые вероятности.

a) Определите спектральную плотность мощности сигнала ЧМ

b) Покажите, что спектральная плотность мощности уменьшается пропорционально $1/f^2$ для $f >> f_c$.

5

ОПТИМАЛЬНЫЕ ПРИЁМНИКИ ДЛЯ КАНАЛА С АДДИТИВНЫМ БЕЛЫМ ГАУССОВСКИМ ШУМОМ

В гл. 4 мы рассмотрели различные методы модуляции, которые используются для передачи цифровой информации через канал связи. Как мы видели, модулятор на передаче формирует способ отображения цифровой последовательности в форме канального сигнала.

Эта глава имеет дело с проектированием (синтезом) и характеристиками качества оптимальных приёмников при различных методах модуляции, когда канал искажает передаваемый сигнал посредством аддитивного гауссовского шума. В разд. 5.1 сначала рассматриваем сигнал модуляции без памяти, затем сигнал модуляции с памятью. В разд. 5.2 оценим вероятность ошибки при различных методах модуляции. В разд. 5.3 рассмотрим оптимальный приёмник для сигналов МНФ и его характеристики качества. В разд. 5.4 обсудим оптимальный приёмник, когда фаза несущей сигнала неизвестна на приёмной стороне и она рассматривается как случайная величина. Наконец, в разд. 5.5 обсудим использование регенеративных повторителей для передачи сигналов и выполним анализ ресурсов линии связи с радиоканалом.

5.1. ОПТИМАЛЬНЫЙ ПРИЁМНИК ДЛЯ СИГНАЛОВ, ПОДВЕРЖЕННЫХ ВОЗДЕЙСТВИЮ АДДИТИВНОГО БЕЛОГО ГАУССОВСКОГО ШУМА

Начнём с разработки математической модели для сигнала на входе приёмника. Предположим, что передатчик передает цифровую информацию посредством M сигналов $\{s_m(t),\ \ m=1,2\ldots,M\}$.

Каждый сигнал передается на символьном интервале длительностью T. Для конкретности рассмотрим передачу информации на символьном интервале $0\le t\le T$. Предполагается, что канал искажает сигнал посредством аддитивного белого гауссовского шума (АБГШ), как показано на рис. 5.1.1.

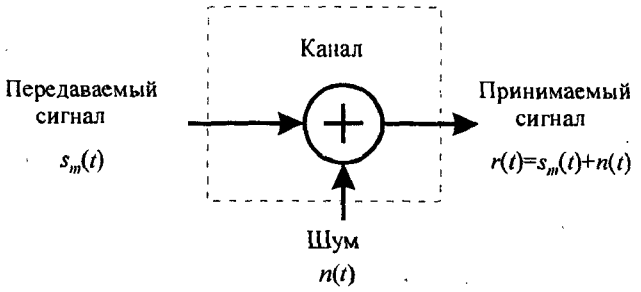


Рис. 5.1.1. Модель принимаемого сигнала, прошедшего через канал с АБГШ

Таким образом, принимаемый сигнал на интервале $0\le t\le T$ можно выразить так:[1]

$$r(t)=s_m(t)+n(t),\quad 0\le t\le T, \tag{5.1.1}$$

где $n(t)$ означает реализацию АБГШ со спектральной плотностью мощности

[1] Обычно в приёмнике цифровой связи выделяют демодулятор (первую решающую схему) и декодер (вторую решающую схему). Здесь под приёмником автор подразумевает только первую решающую схему, и целесообразность её деления на демодулятор и детектор представляется спорной (прп).

$\Phi_{nn}(f) = \frac{1}{2} N_0$ Вт / Гц. Основываясь на наблюдении $r(t)$ на сигнальном интервале, мы желаем найти приёмник, который оптимален в смысле минимизации средней вероятности ошибки.

Удобно разделить приёмник на две части: демодулятор сигнала и детектор – как показано на рис. 5.1.2. Функция демодулятора сигнала заключается в превращении сигнала $r(t)$ в N-мерный вектор $\mathbf{r} = [r_1\ r_2\ \ldots\ r_N]$, где N – размерность переданного сигнала. Задача детектора – решить, основываясь на векторе $\mathbf{r}$, какой из M возможных сигналов был передан.

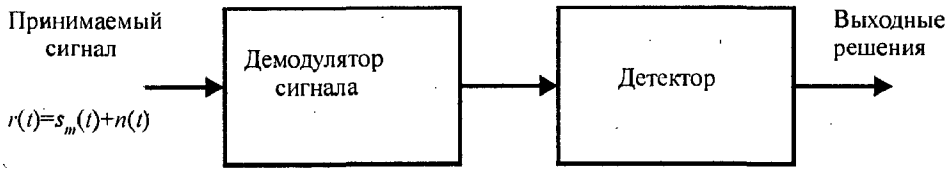


Рис. 5.1.2. Конфигурация приёмника

В следующих двух разделах описаны две реализации демодулятора сигнала. Одна основана на использовании корреляторов, вторая – на применении согласованных фильтров. Оптимальный детектор, который следует за демодулятором, проектируется так, чтобы минимизировать среднюю вероятность ошибки.

5.1.1. Корреляционный демодулятор

В этом разделе мы опишем корреляционный демодулятор, который разлагает принимаемый сигнал и шум на N-мерные векторы. Другими словами, сигнал и шум разлагаются в линейную взвешенную сумму ортонормированных базисных функций $\{f_n(t)\}$. Считается, что N базисных функций $\{f_n(t)\}$ покрывают пространство сигналов так, что каждый из возможных переданных сигналов из ансамбля $\{s_m(t),\ 1 \le m \le M\}$ может быть представлен как взвешенная линейная комбинация $\{f_n(t)\}$. Для шума функции не покрывают всё его пространство. Однако, как мы увидим ниже, компоненты шума, которые попадают вне пространства сигналов, не влияют на детектирование сигнала.

Предположим, что принимаемый сигнал $r(t)$ прошел через параллельный блок из N взаимных корреляторов, которые вычисляют его проекции на N базисных функций $\{f_n(t)\}$, как показано на рис. 5.1.3.

Эти проекции равны

$$\int_0^T r(t)\, f_k(t)dt = \int_0^T \left[s_m(t) + n(t)\right] f_k(t)dt,$$
$$r_k = s_{mk} + n_k, \quad k = 1, 2, \ldots, N, \tag{5.1.2}$$

где

$$s_{mk} = \int_0^T s_m(t)\, f_k(t)dt, \quad k = 1, 2, \ldots, N,$$
$$n_k = \int_0^T n(t)\, f_k(t)dt, \quad k = 1, 2, \ldots, N. \tag{5.1.3}$$

Сигнал теперь представлен вектором $\mathbf{s}_m$ с компонентами s_{mk}, $k = 1, 2, \ldots, N$. Их величины зависят от того, какой из M сигналов был передан. Компоненты $\{n_k\}$ являются случайными величинами, возникшими из-за присутствия аддитивного шума.

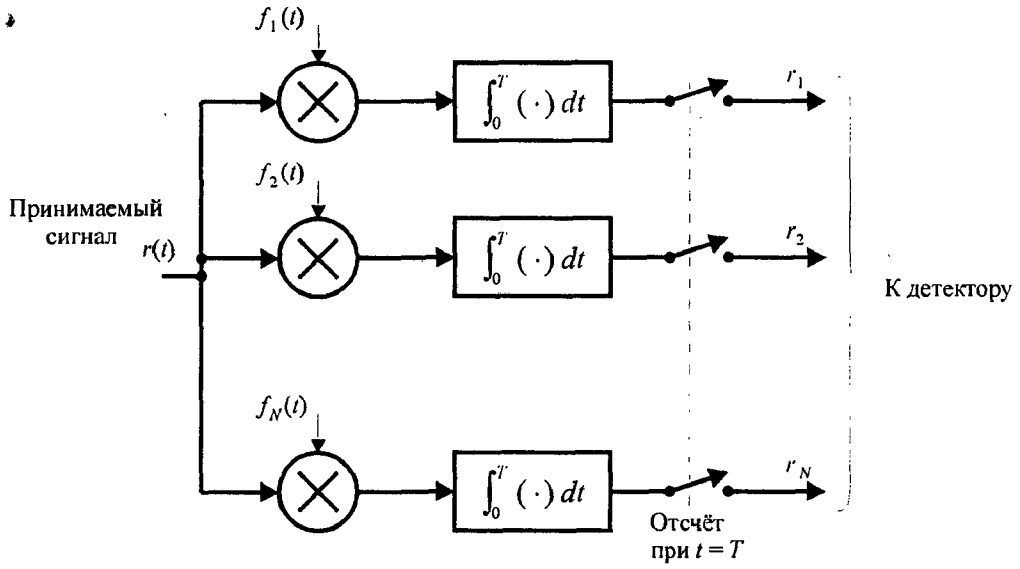


Рис. 5.1.3. Демодулятор по корреляционной схеме

Действительно, принимаемый сигнал $r(t)$ на интервале $0 \le t \le T$ можно выразить так:

$$r(t) = \sum_{k=1}^{N} s_{mk} f_k(t) + \sum_{k=1}^{N} n_k f_k(t) + n'(t) = \sum_{k=1}^{N} r_k f_k(t) + n'(t). \qquad (5.1.4)$$

Слагаемое $n'(t)$, определённое как

$$n'(t) = n(t) - \sum_{k=1}^{N} n_k f_k(t), \qquad (5.1.5)$$

является случайным гауссовским процессом с нулевым средним, который представляет разницу между действительным шумовым процессом $n(t)$ и той его частью, которая соответствует проекции $n(t)$ на базисные функции $\{f_k(t)\}$. Как увидим ниже, $n'(t)$ не влияет на качество решения о переданном сигнале. Следовательно, решение можно сделать, основываясь на выходах корреляторов $r_k = s_{mk} + n_k$, $k = 1, 2, \ldots, N$.

Поскольку сигналы $\{s_m(t)\}$ детерминированы, то сигнальные компоненты s_{mk} детерминированы. Компоненты шума $\{n_k\}$ гауссовские, их средние значения равны

$$E(n_k) = \int_0^T E[n(t)] f_k(t)\,dt = 0 \qquad (5.1.6)$$

для всех n. Их ковариации (в том числе дисперсии) равны

$$E(n_k n_m) = \int_0^T \int_0^T E[n(t)n(\tau)] f_k(t) f_m(\tau)\,dt\,d\tau = \\ = \tfrac{1}{2} N_0 \int_0^T \int_0^T \delta(t-\tau) f_k(t) f_m(\tau)\,dt\,d\tau = \tfrac{1}{2} N_0 \int_0^T f_k(t) f_m(\tau)\,dt = \tfrac{1}{2} N_0 \delta_{mk}, \qquad (5.1.7)$$

где $\delta_{mk} = 1$, когда $m = k$, и равно нулю, если это условие не выполняется. Следовательно, N шумовых компонент $\{n_k\}$ – некоррелированные гауссовские случайные величины с нулевыми средними и одинаковой дисперсией $\sigma_n^2 = \frac{1}{2} N_0$.

Из вышеизложенного следует, что выходы корреляторов $\{r_k\}$, определяемые m-м переданным сигналом, являются гауссовскими случайными величинами со средними

$$E(r_k) = E(s_{mk} + n_k) = s_{mk} \qquad (5.1.8)$$

и одинаковыми дисперсиями

$$\sigma_r^2 = \sigma_n^2 = \tfrac{1}{2} N_0 . \qquad (5.1.9)$$

Поскольку компоненты шума $\{n_k\}$ являются некоррелированными гауссовскими случайными величинами, они также статистически независимы. Как следствие, выходы корреляторов $\{r_k\}$, определяемые переданным m-м сигналом, – статистически независимые гауссовские случайные величины. Следовательно, условные плотности вероятности случайных величин $[r_1 r_2 \ldots r_N] = \mathbf{r}$ равны

$$p(\mathbf{r}|s_m) = \prod_{k=1}^{N} p(r_k|s_{mk}), \quad m = 1,2,\ldots,M, \qquad (5.1.10)$$

где

$$p(r_k|s_{mk}) = \frac{1}{\sqrt{\pi N_0}} \exp\left[-\frac{(r_k - s_{mk})^2}{N_0}\right], \quad k = 1,2,\ldots,N. \qquad (5.1.11)$$

Подставив (5.1.11) в (5.1.10), получим совместную условную ФПВ

$$p(\mathbf{r}|s_m) = \frac{1}{(\pi N_0)^{N/2}} \exp\left[-\sum_{k=1}^{N} \frac{(r_k - s_{mk})^2}{N_0}\right], \quad m = 1,2,\ldots,M. \qquad (5.1.12)$$

В заключение покажем, что выходы корреляторов $(r_1, r_2, \ldots r_N)$ являются достаточной статистикой для принятия решения о том, какой из M сигналов был передан, т.е. что никакая дополнительная полезная информация не может быть извлечена из остаточного шумового процесса $n'(t)$. В самом деле, процесс $n'(t)$ не коррелирован с N выходами корреляторов $\{r_k\}$, т.е.

$$\begin{aligned} E(n'(t)r_k) &= E(n'(t))s_{mk} + E(n'(t)n_k) = E[n'(t)n_k] = \\ &= E\left\{\left[n(t) - \sum_{j=1}^{N} n_j f_j(t)\right] n_k\right\} = \int_0^T E[n(t)n(\tau)] f_k(\tau)\,d\tau - \sum_{j=1}^{N} E(n_j n_k) f_j(t) = \\ &= \tfrac{1}{2} N_0 f_k(t) - \tfrac{1}{2} N_0 f_k(t) = 0. \end{aligned} \qquad (5.1.13)$$

Поскольку $n'(t)$ и $\{r_k\}$ являются гауссовскими и некоррелированными, они также статистически независимы. Следовательно, $n'(t)$ не содержит информацию, которая касается вопроса о том, какой сигнал передан. Вся относящаяся к делу информация находится в выходных данных коррелятора $\{r_k\}$. Следовательно, $n'(t)$ можно пренебречь.

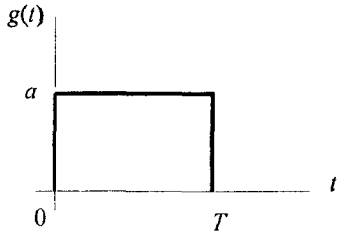


Рис. 5.1.4. Сигнальный импульс для примера 5.1.1

Пример 5.1.1. Рассмотрим M-позиционный ансамбль базовых сигналов, в котором огибающая базового импульса $g(t)$ прямоугольная, как показано на рис. 5.1.4. Аддитивный шум – белый гауссовский шум с нулевым средним. Определим базисную функцию $f(t)$ и выход демодулятора корреляционного типа.

Энергия прямоугольного импульса равна

$$\mathscr{E}_g = \int_0^T g^2(t)\,dt = \int_0^T a^2\,dt = a^2 T .$$

Поскольку ансамбль АМ сигналов имеет размерность $N = 1$, есть лишь одна базисная функция $f(t)$. Она определяется так

$$f(t)=\frac{1}{\sqrt{a^2T}}g(t)=\begin{cases}1/\sqrt{T} & (0\le t\le T),\\ 0 & (\text{для других } t).\end{cases}$$

Выход демодулятора корреляционного типа равен

$$r=\int_0^T r(t)f(t)dt=\frac{1}{\sqrt{T}}\int_0^T r(t)dt\,.$$

Интересно отметить, что когда $f(t)$ – прямоугольная функция, коррелятор оказывается простым интегратором. Если подставим выражение для $r(t)$, получим

$$r=\frac{1}{\sqrt{T}}\left\{\int_0^T\left[s_m(t)+n(t)\right]dt\right\}=\frac{1}{\sqrt{T}}\left[\int_0^T s_m(t)\,dt+\int_0^T n(t)\,dt\right],$$

$$r=s_m+n,$$

где для шумового слагаемого $E(n)=0$ и

$$\sigma_n^2=E\left[\frac{1}{T}\int_0^T\int_0^T n(t)n(\tau)dtd\tau\right]=\frac{1}{T}\int_0^T\int_0^T E\left[n(t)n(\tau)\right]dtd\tau=$$

$$=\frac{N_0}{2T}\int_0^T\int_0^T\delta(t-\tau)dtd\tau=\tfrac{1}{2}N_0.$$

ФПВ для выходного отсчёта равна

$$p(r|s_m)=\frac{1}{\sqrt{\pi N_0}}\exp\left[-\frac{(r-s_m)^2}{N_0}\right].$$

5.1.2. Согласованный фильтр как демодулятор

Вместо использования набора из N корреляторов для генерирования величин $\{r_k\}$ мы можем применить набор из N линейных фильтров. Для конкретности предположим, что импульсная характеристика фильтров такова:

$$h_k(t)=f_k(T-t),\quad 0\le t\le T,\qquad (5.1.14)$$

где $\{f_k(t)\}-N$ базисных функций, и $h_k(t)=0$ вне интервала $0\le t\le T$. Выходы этих фильтров равны

$$y_k(t)=\int_0^t r(\tau)h_k(t-\tau)d\tau=\int_0^t r(\tau)f_k(T-t+\tau)d\tau,\quad k=1,2,\ldots,N.\qquad (5.1.15)$$

Взяв отсчёт выхода в точке $t=T$, получим

$$y_k(T)=\int_0^T r(\tau)f_k(\tau)d\tau=r_k,\quad k=1,2,\ldots,N.\qquad (5.1.16)$$

Следовательно, отсчёт фильтров в точке $t=T$ точно определяет ансамбль величин $\{r_k\}$, полученных на выходе N линейных корреляторов. Фильтр, импульсная характеристика которого $h(t)=s(T-t)$, где $s(t)$ предполагается заключенным в интервале $0\le t\le T$, называется *согласованным фильтром* для сигнала $s(t)$.

Пример сигнала и импульсной характеристики согласованного с ним фильтра показан на рис. 5.1.5.

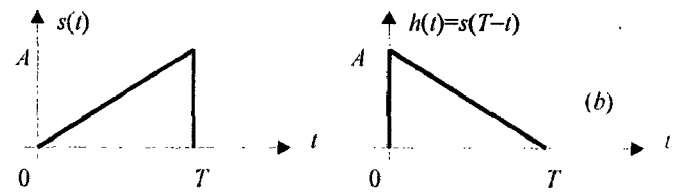


Рис. 5.1.5. Сигнал $s(t)$ (a) и импульсная характеристика фильтра, согласованного с $s(t)$ (b)

Отклик фильтра на сигнал $s(t)$ равен

$$y(t) = \int_0^t s(\tau)s(T-t+\tau)d\tau, \qquad (5.1.17)$$

что по определению является временной автокорреляционной функцией сигнала $s(t)$. Рисунок 5.1.6 иллюстрирует $y(t)$ для треугольного импульса, показанного на рис. 5.1.5, *а*.

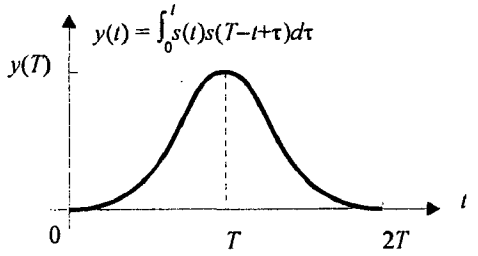


Рис. 5.1.6. Отклик согласованного фильтра определяет автокорреляционную функцию $s(t)$

Заметим, что автокорреляционная функция $y(t)$ является чётной функцией t и имеет пик в точке $t = T$.

В случае демодулятора, описанного выше, N согласованных фильтров согласованы с базисными функциями $f_k(t)$. Рис. 5.1.7 иллюстрирует демодулятор на согласованных фильтрах, который выдает наблюдаемые величины $\{r_k\}$.

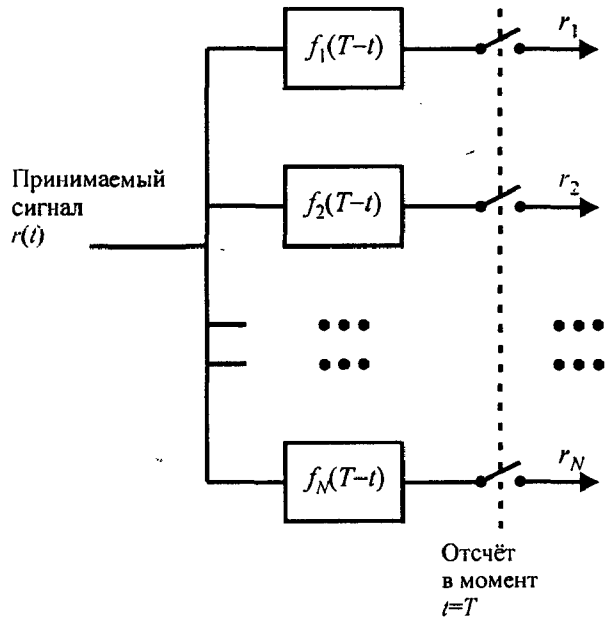


Рис. 5.1.7. Демодулятор на основе согласованных фильтров

Свойства согласованного фильтра. Согласованный фильтр имеет ряд интересных свойств. Докажем наиболее существенное свойство, которое состоит в следующем: если сигнал $s(t)$ подвергается воздействию АБГШ, то фильтр, согласованный с сигналом $s(t)$, максимизирует на выходе отношение сигнал / шум (ОСШ).

Чтобы доказать это свойство, предположим, что принимаемый сигнал $r(t)$ состоит из сигнала $s(t)$ и АБГШ $n(t)$ с нулевым средним и спектральной плотностью мощности $\Phi_{nn}(f) = \frac{1}{2}N_0$ Вт / Гц. Предположим, что сигнал $r(t)$ прошёл через фильтр с импульсной характеристикой $h(t)$, $0 \le t \le T$, и берётся отсчёт на выходе в точке $t = T$. Отклик фильтра на сигнальную и шумовую компоненту равен

$$y(t) = \int_0^t r(\tau)h(t-\tau)d\tau = \int_0^t s(\tau)h(t-\tau)d\tau + \int_0^t n(\tau)h(t-\tau)d\tau. \qquad (5.1.18)$$

В точке отсчёта $t = T$ сигнальная и шумовая компоненты равны

$$y(T) = \int_0^T s(\tau)h(T-\tau)d\tau + \int_0^T n(\tau)h(T-\tau)d\tau = y_s(T) + y_n(T), \qquad (5.1.19)$$

где $y_s(T)$ представляет сигнальную компоненту, а $y_n(T)$ – шумовую компоненту. Задача сводится к выбору импульсной характеристики фильтра, который максимизирует на выходе отношение сигнал / шум (ОСШ), определяемое так:

$$\text{ОСШ}_0 = \frac{y_s^2(T)}{E\left[y_n^2(T)\right]}. \qquad (5.1.20)$$

Знаменатель в (5.1.20) определяет дисперсию шумовой компоненты на выходе фильтра. Определим $E\left[y_n^2(T)\right]$, т.е. дисперсию выходного шума. Имеем

$$\begin{aligned} E\left[y_n^2(T)\right] &= \int_0^T \int_0^T E\left[n(\tau)n(t)\right]h(T-\tau)\,h(T-t)\,dtd\tau = \\ &= \tfrac{1}{2}N_0 \int_0^T \int_0^T \delta(t-\tau)\,h(T-\tau)\,h(T-t)\,dtd\tau = \tfrac{1}{2}N_0 \int_0^T h^2(T-\tau)dt. \end{aligned} \qquad (5.1.21)$$

Заметим, что эта дисперсия зависит от спектральной плотности шума на входе и энергии импульсной характеристики $h(t)$. Подставив $y_s(T)$ и $E\left[y_n^2(T)\right]$ в (5.1.20), получим для ОСШ на выходе фильтра выражение

$$\text{ОСШ}_0 = \frac{\left[\int_0^T s(\tau)h(T-\tau)d\tau\right]^2}{\frac{1}{2}N_0 \int_0^T h^2(T-t)dt} = \frac{\left[\int_0^T h(\tau)s(T-\tau)d\tau\right]^2}{\frac{1}{2}N_0 \int_0^T h^2(T-t)dt}. \qquad (5.1.22)$$

Так как знаменатель в ОСШ зависит от энергии $h(t)$, максимум ОСШ по $h(t)$ можно получить максимизацией числителя в предположении, что знаменатель фиксирован. Максимизация числителя выполняется легко использованием неравенства Коши–Шварца, которое в общем гласит, что если $g_1(t)$ и $g_2(t)$ – сигналы с ограниченной энергией, то

$$\left[\int_{-\infty}^{\infty} g_1(t)g_2(t)dt\right]^2 \le \int_{-\infty}^{\infty} g_1^2(t)dt \int_{-\infty}^{\infty} g_2^2(t)dt \qquad (5.1.23)$$

с равенством, когда $g_1(t) = Cg_2(t)$, C – произвольная константа. Если положим $g_1(t) = h(t)$ и $g_2(t) = s(T-t)$, то ясно, что ОСШ максимизируется, когда $h(t) = Cs(T-t)$, т.е. $h(t)$ согласовано с сигналом $s(t)$. Константа C^2 не входит в ОСШ, так как она одновременно присутствует в числителе и знаменателе. Выходное (максимальное) ОСШ, получаемое при помощи согласованного фильтра, равно

$$\text{ОСШ}_0 = \frac{2}{N_0}\int_0^T s^2(t)dt = 2\mathcal{E}/N_0. \qquad (5.1.24)$$

Заметим, что выходное ОСШ у согласованного фильтра зависит только от энергии сигнала $s(t)$, но не от детальных характеристик $s(t)$. Это другое интересное свойство согласованного фильтра.

Интерпретация согласованного фильтра в частотной области. Согласованный фильтр имеет интересную интерпретацию в частотной области. Поскольку $h(t) = s(T-t)$, преобразование Фурье такого сигнала

$$H(f) = \int_0^T s(T-t)\,e^{-j2\pi ft}dt = \left[\int_0^T s(\tau)\,e^{-j2\pi f\tau}d\tau\right]e^{-j2\pi fT} = S^*(f)e^{-j2\pi fT}. \qquad (5.1.25)$$

Видим, что согласованный фильтр имеет частотную характеристику, которая комплексно сопряжена с частотной характеристикой сигнала, и множитель $e^{-j2\pi fT}$,

который определяет задержку сигнала на T. Другими словами, $|H(f)| = |S(f)|$, так что амплитудно-частотная характеристика согласованного фильтра идентична амплитудно-частотной характеристике сигнала. С другой стороны, фазовая характеристика $H(f)$ противоположна по отношению к фазочастотной характеристике сигнала $S(f)$.

Если сигнал со спектром $S(f)$ проходит через согласованный фильтр, то отклик фильтра на этот сигнал имеет спектр $Y(f) = |S(f)|^2 e^{-j2\pi fT}$. Следовательно, сигнальная составляющая на выходе фильтра

$$y_s(t) = \int_{-\infty}^{\infty} Y(f) e^{j2\pi ft} df = \int_{-\infty}^{\infty} |S(f)|^2 e^{-j2\pi fT} e^{j2\pi ft} df . \quad (5.1.26)$$

В точке $t = T$ имеем

$$y_s(T) = \int_{-\infty}^{\infty} |S(f)|^2 df = \int_0^T s^2(t)\, dt = \mathscr{E}, \quad (5.1.27)$$

причём последний шаг преобразования следует из соотношения Парсеваля.

Шум на выходе фильтра имеет спектральную плотность мощности

$$\Phi_0(f) = \tfrac{1}{2} |H(f)|^2 N_0 . \quad (5.1.28)$$

Следовательно, суммарная мощность шума на выходе согласованного фильтра

$$P_n = \int_{-\infty}^{\infty} \Phi_0(f) df = \tfrac{1}{2} N_0 \int_{-\infty}^{\infty} |H(f)|^2 df = \tfrac{1}{2} N_0 \int_{-\infty}^{\infty} |S(f)|^2 df = \tfrac{1}{2} \mathscr{E} N_0 . \quad (5.1.29)$$

Выходное ОСШ равно отношению мощности сигнала P_s

$$P_s = y_s^2(T) = \mathscr{E}^2 \quad (5.1.30)$$

к мощности шума P_n. Следовательно,

$$\text{ОСШ}_0 = \frac{P_s}{P_n} = \frac{\mathscr{E}^2}{\frac{1}{2}\mathscr{E} N_0} = 2\mathscr{E}/N_0 , \quad (5.1.31)$$

что совпадает с результатом (5.1.24).

Пример 5.1.2. Рассмотрим $M = 4$ биортогональных сигналов для передачи информации по каналу с АБГШ. Два сигнала из этого ансамбля с положительной полярностью показаны на рис. 5.1.8, *а*. Считается, что шум имеет нулевое среднее и спектральную плотность мощности $\frac{1}{2} N_0$. Определим базисные функции для этого ансамбля сигналов, импульсную характеристику согласованного фильтра в качестве демодулятора и выходной сигнал согласованного фильтра-демодулятора, когда передан сигнал $s_1(t)$.

Ансамбль из $M = 4$ биортогональных сигналов имеет размерность $N = 2$. Следовательно, требуются две базисные функции для представления сигналов. Согласно рис. 5.1.8 выберем эти базисные функции $f_1(t)$ и $f_2(t)$ так:

$$f_1(t) = \begin{cases} \sqrt{2/T} & 0 \le t \le \frac{1}{2}T, \\ 0 & \text{(для других } t\text{)}; \end{cases}$$
$$f_2(t) = \begin{cases} \sqrt{2/T} & \frac{1}{2}T \le t \le T, \\ 0 & \text{(для других } t\text{)}. \end{cases} \quad (5.1.32)$$

Эти два сигнала иллюстрируются рис. 5.1.8, *а*. Импульсные характеристики фильтров, согласованных с этими сигналами, равны

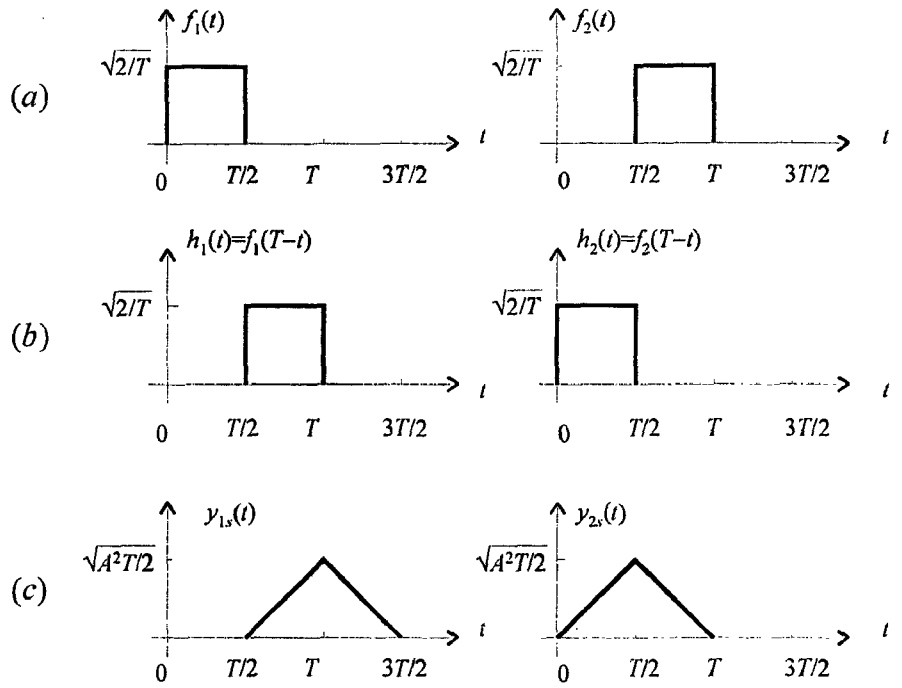


Рис. 5.1.8. Базисные функции и отклики согласованных фильтров на сигнал $s_1(t)=Af_1(t)$ для примера 5.1.2

$$h_1(t)=f_1(T-t)=\begin{cases}\sqrt{2/T} & \frac{1}{2}T\le t\le T,\\ 0 & \text{(для других } t\text{)},\end{cases}$$
$$h_2(t)=f_2(T-t)=\begin{cases}\sqrt{2/T} & 0\le t\le \frac{1}{2}T,\\ 0 & \text{(для других } t\text{)}.\end{cases} \qquad (5.1.33)$$

Они иллюстрируются на рис. 5.1.8, *b*.

Если передан $s_1(t)$, то сигналы на выходе согласованных фильтров $y_{1s}(t)$ и $y_{2s}(t)$ (без шумов) имеют вид, показанный на рис. 5.1.8, *c*. Поскольку берутся отсчёты выходных сигналов в точке $t=T$, то видим, что $y_{1s}(T)=\sqrt{\frac{1}{2}A^2T}$, а $y_{2s}(T)=0$. Заметим, что $\frac{1}{2}A^2T=\mathscr{E}$ – это энергия сигнала. Вектор сигналов, формируемый на выходе согласованных фильтров в точке $t=T$, равен

$$\mathbf{r}=\left[r_1\ \ r_2\right]=\left[\sqrt{\mathscr{E}}+n_1\ \ n_2\right], \qquad (5.1.34)$$

где $n_1=y_{1n}(T)$ и $n_2=y_{2n}(T)$ – шумовые компоненты на выходе согласованных фильтров, определяемые так:

$$y_{kn}(T)=\int_0^T n(t)f_k(t)\,dt,\quad k=1,2. \qquad (5.1.35)$$

Ясно, что $E(n_k)=E\left[y_{kn}(T)\right]=0$. Дисперсии шумовых компонент на выходе фильтра

$$\sigma_n^2=E\left[y_{kn}^2(T)\right]=\int_0^T\int_0^T E\left[n(t)n(\tau)\right]f_k(t)f_k(\tau)dtd\tau=$$
$$=\tfrac{1}{2}N_0\int_0^T\int_0^T\delta(t-\tau)f_k(t)f_k(\tau)dtd\tau=\tfrac{1}{2}N_0\int_0^T f_k^2(t)dt=\tfrac{1}{2}N_0. \qquad (5.1.36)$$

Видим, что ОСШ$_0$ для первого согласованного фильтра

$$\text{ОСШ}_0=\frac{P_s}{P_n}=\frac{\left(\sqrt{\mathscr{E}}\right)^2}{\frac{1}{2}N_0}=2\mathscr{E}/N_0, \qquad (5.1.37)$$

что совпадает с нашим предыдущим результатом. Заметим также, что четыре возможных выхода двух согласованных фильтров, соответствующих передаваемым сигналам, равны

$$(r_1, r_2) = \left(\sqrt{\mathscr{E}} + n_1,\, n_2\right),\ \left(n_1,\, \sqrt{\mathscr{E}} + n_2\right),\ \left(-\sqrt{\mathscr{E}} + n_1,\, n_2\right) \text{ и } \left(n_1,\, -\sqrt{\mathscr{E}} + n_2\right).$$

5.1.3. Оптимальный детектор

Мы показали, что при передаче сигналов через канал с АБГШ демодулятор на основе корреляционной схемы или на основе согласованных фильтров выдаёт вектор $\mathbf{r} = [r_1\, r_2 \ldots r_N]$, который содержит всю доступную информацию о принимаемом сигнале. В этом разделе мы опишем оптимальные правила решения, основанные на наблюдаемом векторе $\mathbf{r}$. Предположим, что сигналы, передаваемые на последовательных сигнальных интервалах, не имеют памяти.

Мы желаем синтезировать детектор сигнала, который выносит решение о передаваемом сигнале на каждом сигнальном интервале, основываясь на наблюдении вектора $\mathbf{r}$ в каждом интервале, так, чтобы максимизировать среднюю вероятность правильного решения. С этой целью рассмотрим правило[1] решения, базирующееся на вычислении *апостериорных вероятностей*

$$P(\mathbf{s}_m|\mathbf{r}),\quad m = 1, 2 \ldots, M,$$

и на выборе сигнала, соответствующего максимуму ряда апостериорных вероятностей $\{P(\mathbf{s}_m|\mathbf{r})\}$. Позже покажем, что это правило максимизирует среднюю вероятность правильного решения и, следовательно, минимизирует среднюю вероятность ошибки. Такое правило решения названо правилом максимума апостериорной вероятности (МАВ).

Используя правило Байеса, апостериорную вероятность можно выразить так:

$$P(\mathbf{s}_m|\mathbf{r}) = \frac{p(\mathbf{r}|\mathbf{s}_m)P(\mathbf{s}_m)}{p(\mathbf{r})}, \tag{5.1.38}$$

где $p(\mathbf{r}|\mathbf{s}_m)$ – условная ФПВ наблюдаемого вектора при передаче $\mathbf{s}_m$, а $P(\mathbf{s}_m)$ – вероятность m-го передаваемого сигнала. Знаменатель (5.1.38) можно выразить так:

$$p(\mathbf{r}) = \sum_{m=1}^{M} p(\mathbf{r}|\mathbf{s}_m)P(\mathbf{s}_m). \tag{5.1.39}$$

Из (5.1.38) и (5.1.39) видим, что вычисление апостериорных вероятностей $P(\mathbf{s}_m|\mathbf{r})$ требует знания *априорных вероятностей* $P(\mathbf{s}_m)$ и условных ФПВ $P(\mathbf{s}_m|\mathbf{r})$ для $m = 1, 2 \ldots, M$.

Некоторое упрощение имеет место при использовании правила МАВ, когда M сигналов имеют одинаковую априорную вероятность, т.е. $P(\mathbf{s}_m) = 1/M$ для всех M. Более того, заметим, что знаменатель (5.1.38) не зависит от того, какой сигнал передается. Следовательно, правило решения, основанное на нахождении сигнала, который максимизирует $P(\mathbf{s}_m|\mathbf{r})$, эквивалентно в рассмотренном случае нахождению сигнала, который максимизирует $p(\mathbf{r}|\mathbf{s}_m)$.

Условную ФПВ $p(\mathbf{r}|\mathbf{s}_m)$ или некоторую монотонную функцию от неё обычно называют *функцией правдоподобия*. Правило решения, основанное на максимизации $p(\mathbf{r}|\mathbf{s}_m)$ по всем

---

[1] Автор здесь и в дальнейшем использовал термин «критерий решения», который представляется не подходящим (прп).

M сигналам, называют правилом максимального правдоподобия (МП). Видим, что детектор, основанный на правиле МАВ и тот, который основан на правиле МП, обеспечивают одинаковое решение при одинаковых априорных вероятностях $P(\mathbf{s}_m)$, т.е. при равновероятных сигналах $\{\mathbf{s}_m\}$.

В случае АБГШ в канале функции правдоподобия $p(\mathbf{r}|\mathbf{s}_m)$ определяются (5.1.12). Для упрощения расчётов будем использовать натуральный логарифм от $p(\mathbf{r}|\mathbf{s}_m)$, который является монотонной функцией. Таким образом,

$$\ln p(\mathbf{r}|\mathbf{s}_m) = -\tfrac{1}{2} N \ln(\pi N_0) - \frac{1}{N_0}\sum_{k=1}^{N}(r_k - s_{mk})^2 . \qquad (5.1.40)$$

Максимизация $\ln p(\mathbf{r}|\mathbf{s}_m)$ по $\mathbf{s}_m$ эквивалентна нахождению сигнала $\mathbf{s}_m$, который минимизирует евклидово расстояние

$$D(\mathbf{r},\mathbf{s}_m) = \sum_{k=1}^{N}(r_k - s_{mk})^2 . \qquad (5.1.41)$$

$D(\mathbf{r},\mathbf{s}_m)$, $m=1,2...,M$, называют дистанционными характеристиками. Следовательно, для канала с АБГШ правило решения, основанное на правиле МП, сводится к нахождению сигнала $\mathbf{s}_m$, который наиболее близок по расстоянию к принимаемому сигнальному вектору $\mathbf{r}$. Мы будем ссылаться на это правило решения как на детектирование по минимуму расстояния.

Другую интерпретацию оптимального правила решения, основанного на правиле МП, можно получить путем раскрытия дистанционных метрик в (5.1.41):

$$D(\mathbf{r},\mathbf{s}_m) = \sum_{k=1}^{N} r_k^2 - 2\sum_{k=1}^{N} r_k s_{mk} + \sum_{k=1}^{N} s_{mk}^2 = |\mathbf{r}|^2 - 2\mathbf{r}\cdot\mathbf{s}_m + |\mathbf{s}_m|^2, \quad m=1,2...,M . \qquad (5.1.42)$$

Слагаемое $|\mathbf{r}|^2$ – общее для всех дистанционных метрик, и, следовательно, его можно не учитывать при вычислении метрик. Новый результат сводится к ряду модифицированных метрик

$$D'(\mathbf{r},\mathbf{s}_m) = -2\mathbf{r}\cdot\mathbf{s}_m + |\mathbf{s}_m|^2 . \qquad (5.1.43)$$

Заметим, что выбор сигнала $\mathbf{s}_m$, который минимизирует $D'(\mathbf{r},\mathbf{s}_m)$, эквивалентен выбору сигнала, который максимизирует метрику $C(\mathbf{r},\mathbf{s}_m) = -D'(\mathbf{r},\mathbf{s}_m)$, т.е.

$$C(\mathbf{r},\mathbf{s}_m) = 2\mathbf{r}\cdot\mathbf{s}_m - |\mathbf{s}_m|^2 . \qquad (5.1.44)$$

Слагаемое $\mathbf{r}\cdot\mathbf{s}_m$ представляет проекцию принимаемого вектора сигнала на сигнальные векторы всех M возможных для передачи сигналов. Величина каждой такой проекции является мерой корреляции между принятым вектором и m-м сигналом. Из этих соображений мы называем $C(\mathbf{r},\mathbf{s}_m)$, $m=1,2...,M$, *корреляционными метриками* для решения того, какой из M сигналов был передан. Наконец, слагаемые $|\mathbf{s}_m|^2 = \mathscr{E}_m$, $m=1,2...,M$ можно рассматривать как пороговые слагаемые, которые служат компенсацией для ансамбля сигналов с неравными энергиями сигналов, такого, как при АМ. Если все сигналы имеют одинаковую энергию, $|\mathbf{s}_m|^2$, можно не учитывать при вычислении корреляционных метрик $C(\mathbf{r},\mathbf{s}_m)$ и дистанционных метрик $D(\mathbf{r},\mathbf{s}_m)$ или $D'(\mathbf{r},\mathbf{s}_m)$.

Легко показать (см. задачу 5.5), что корреляционные метрики $C(\mathbf{r},\mathbf{s}_m)$ можно также выразить так:

$$C(\mathbf{r},\mathbf{s}_m)=2\int_0^T r(t)s_m(t)dt-\mathcal{E}_m,\quad m=0,1,\ldots,M. \qquad (5.1.45)$$

Следовательно, эти метрики можно генерировать демодулятором, который определяет корреляцию принимаемого сигнала с каждым из M возможных к передаче сигналов и устанавливает для выхода коррелятора вычитаемый порог в случае сигналов с неравными энергиями. Эквивалентно принимаемый сигнал можно пропустить через блок из M фильтров, согласованных с возможными к передаче сигналами $\{s_m(t)\}$, и взять отсчёты в конце символьного интервала $t=T$. Следовательно, оптимальный приёмник (демодулятор и детектор) можно выполнить по альтернативной схеме, показанной на рис. 5.1.9.

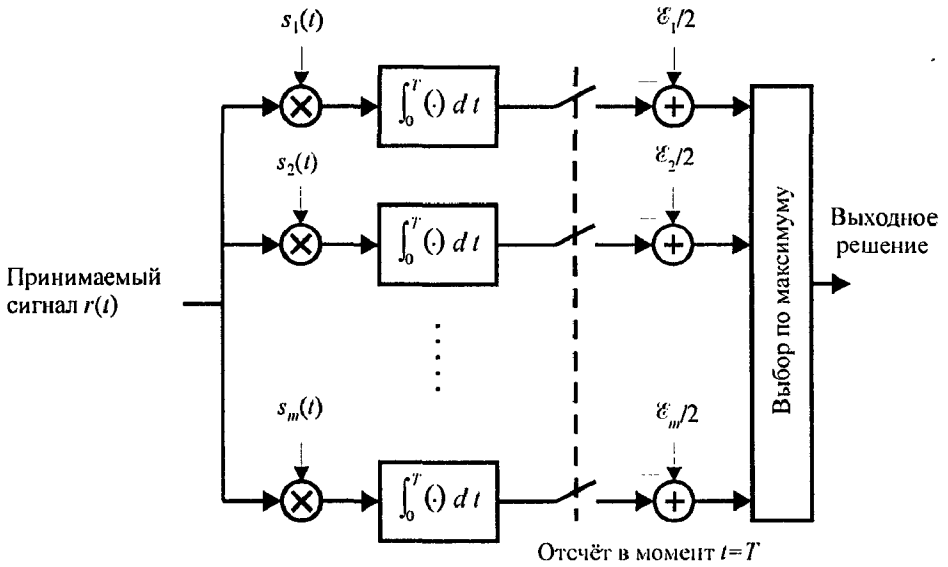


Рис. 5.1.9. Альтернативная реализация оптимального приемника при АБГШ

Суммируя, можно сказать, что мы показали, что оптимальный МП детектор вычисляет набор из M расстояний $D(\mathbf{r},\mathbf{s}_m)$ или $D'(\mathbf{r},\mathbf{s}_m)$ и выбирает сигнал, соответствующий минимальной (дистанционной) метрике. Эквивалентно оптимальный МП детектор вычисляет набор из M корреляционных метрик $C(\mathbf{r},\mathbf{s}_m)$ и выбирает сигнал, соответствующий наибольшей корреляционной метрике.

В вышеприведенном исследовании оптимального детектора рассмотрен важный случай, когда все сигналы равновероятны. В этом случае правило МАВ эквивалентно правилу МП. Однако, когда сигналы не равновероятны, оптимальный МАВ детектор основывает свои решения на вероятностях $P(\mathbf{s}_m|\mathbf{r})$, $m=1,2\ldots,M$, даваемых (5.1.38), или, что эквивалентно, на *метриках*

$$PM(\mathbf{r},\mathbf{s}_m)=p(\mathbf{r}|\mathbf{s}_m)P(\mathbf{s}_m).$$

Следующие примеры иллюстрируют эти расчеты для сигналов двоичной АМ.

Пример 5.1.3. Рассмотрим случай двоичных сигналов АМ, когда две возможные сигнальные точки равны $s_1=-s_2=\sqrt{\mathcal{E}_b}$, где $\mathcal{E}_b$ – энергия на бит. Априорные вероятности равны $P(s_1)=p$ и $P(s_2)=1-p$. Определим метрики для оптимального МАВ детектора, когда передаваемые сигналы искажаются АБГШ.

Вектор принимаемого сигнала (одномерный) для двоичной АМ равен

$$r = \pm\sqrt{\mathscr{E}_b} + y_n(T), \tag{5.1.46}$$

где $y_n(T)$ – гауссовская случайная величина с нулевым средним и с дисперсией $\sigma_n^2 = \frac{1}{2}N_0$. Следовательно, условные ФПВ $p(r|\mathbf{s}_m)$ для двух сигналов

$$p(r|s_1) = \frac{1}{\sqrt{2\pi}\sigma_n}\exp\left[-\frac{\left(r-\sqrt{\mathscr{E}_b}\right)^2}{2\sigma_n^2}\right], \tag{5.1.47}$$

$$p(r|s_2) = \frac{1}{\sqrt{2\pi}\sigma_n}\exp\left[-\frac{\left(r+\sqrt{\mathscr{E}_b}\right)^2}{2\sigma_n^2}\right]. \tag{5.1.48}$$

Поэтому метрики $PM(\mathbf{r},\mathbf{s}_1)$ и $PM(\mathbf{r},\mathbf{s}_2)$ равны

$$PM(\mathbf{r}|\mathbf{s}_1) = p\cdot p(r|s_1) = \frac{p}{\sqrt{2\pi}\sigma_n}\exp\left[-\frac{\left(r-\sqrt{\mathscr{E}_b}\right)^2}{2\sigma_n^2}\right], \tag{5.1.49}$$

$$PM(\mathbf{r}|\mathbf{s}_2) = (1-p)\cdot p(r|s_2) = \frac{1-p}{\sqrt{2\pi}\sigma_n}\exp\left[-\frac{\left(r+\sqrt{\mathscr{E}_b}\right)^2}{2\sigma_n^2}\right]. \tag{5.1.50}$$

Если $PM(\mathbf{r},\mathbf{s}_1) > PM(\mathbf{r},\mathbf{s}_2)$, выберем s_1 как переданный сигнал; в противном случае выберем s_2. Такое правило решения можно выразить так:

$$\frac{PM(\mathbf{r},\mathbf{s}_1)}{PM(\mathbf{r},\mathbf{s}_2)} \underset{s_2}{\overset{s_1}{\gtrless}} 1. \tag{5.1.51}$$

Но

$$\frac{PM(\mathbf{r},\mathbf{s}_1)}{PM(\mathbf{r},\mathbf{s}_2)} = \frac{p}{1-p}\exp\left[\frac{\left(r+\sqrt{\mathscr{E}_b}\right)^2 - \left(r-\sqrt{\mathscr{E}_b}\right)^2}{2\sigma_n^2}\right], \tag{5.1.52}$$

так что (5.1.51) можно выразить так:

$$\left[\frac{\left(r+\sqrt{\mathscr{E}_b}\right)^2 - \left(r-\sqrt{\mathscr{E}_b}\right)^2}{2\sigma_n^2}\right] \underset{s_2}{\overset{s_1}{\gtrless}} \ln\frac{1-p}{p} \tag{5.1.53}$$

или, что эквивалентно,

$$\sqrt{\mathscr{E}_b}\,r \underset{s_2}{\overset{s_1}{\gtrless}} \tfrac{1}{2}\sigma_n^2 \ln\frac{1-p}{p} = \tfrac{1}{4}N_0 \ln\frac{1-p}{p}. \tag{5.1.54}$$

Это окончательная формула, определяющая оптимальный детектор. Она предполагает вычисление корреляционной метрики $C(\mathbf{r},\mathbf{s}_1) = r\sqrt{\mathscr{E}_b}$ и ее сравнение с порогом $\frac{1}{4}N_0 \ln\left[(1-p)/p\right]$. Рисунок 5.1.10 иллюстрирует две сигнальные точки s_1 и s_2. Порог, обозначенный τ_h, делит вещественную ось на две области, скажем R_1 и R_2, где R_1 содержит совокупность точек, которые превышают τ_h, а R_2 содержит совокупность точек, которые меньше τ_h. Если $r\sqrt{\mathscr{E}_b} > \tau_h$, выносится решение, что был передан сигнал s_1, а если $r\sqrt{\mathscr{E}_b} < \tau_h$ – решение, что был передан сигнал s_2. Порог τ_h зависит от N_0 и p. Если $p = \frac{1}{2}$,

то $\tau_h = 0$. Если $p > \frac{1}{2}$, то сигнальная точка s_1 более вероятна и $\tau_h < 0$. В этом случае область R_1 больше, чем R_2, так что более вероятно выбрать решение s_1, чем s_2. Если $p < \frac{1}{2}$ – имеем противоположный случай. Таким образом минимизируется средняя вероятность ошибки.

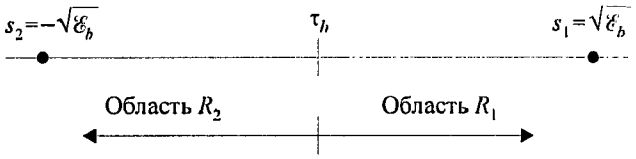


Рис. 5.1.10. Представление пространства сигналов, иллюстрирующее работу оптимального детектора для двоичной АМ

Интересно отметить, что в случае неравных вероятностей для вычисления порога необходимо знать не только априорные вероятности передачи символов, но и спектральную плотность шума N_0. Когда $p = \frac{1}{2}$, порог нулевой, и знание N_0 не требуется.

Мы завершаем этот раздел доказательством того, что правила решения, основанные на правиле максимального правдоподобия, минимизируют среднюю вероятность ошибки, когда все M сигналов равновероятны. Обозначим через R_m область в N-мерном пространстве, в котором мы принимаем решение о том, что передан сигнал $s_m(t)$, когда принят вектор $r = [r_1\ r_2\ \ldots r_N]$. Вероятность ошибочного решения при передаче $s_m(t)$ равна

$$P(e|\mathbf{s}_m) = \int_{R_m^c} p(\mathbf{r}|\mathbf{s}_m)\,dR, \qquad (5.1.55)$$

где R_m^c – дополнение R_m. Средняя вероятность ошибки

$$P(e) = \sum_{m=1}^{M} \frac{1}{M} P(e|\mathbf{s}_m) = \sum_{m=1}^{M} \frac{1}{M} \int_{R_m^c} p(\mathbf{r}|\mathbf{s}_m)\,dR = \sum_{m=1}^{M} \frac{1}{M}\left[1 - \int_{R_m} p(\mathbf{r}|\mathbf{s}_m)\,dR\right]. \qquad (5.1.56)$$

Замечаем, что $P(e)$ минимизируется, если выбирается сигнал $\mathbf{s}_m$ в том случае, когда $p(r|\mathbf{s}_m)$ больше, чем $p(r|\mathbf{s}_k)$, для всех $m \neq k$.

Если M сигналов не равновероятны, вышеприведённое доказательство можно обобщить и показать, что правило МАВ минимизирует среднюю вероятность ошибки.

5.1.4. Последовательный детектор максимального правдоподобия. Алгоритм Витерби

Если модулированный сигнал без памяти, то последовательный детектор, описанный в предыдущем разделе, является оптимальным в смысле минимизации средней вероятности ошибочного приёма символов. С другой стороны, если передаваемый сигнал имеет память, т.е. сигналы, переданные на последовательных сигнальных интервалах, между собой зависимы, оптимальный детектор – это детектор, который основывает свои решения на наблюдении последовательности принимаемых сигналов на последовательных сигнальных интервалах. Ниже опишем два различных типа алгоритмов детектирования последовательности символов. В этом разделе опишем алгоритм максимального правдоподобия для детектирования последовательности символов, который ищет минимум евклидова расстояния траекторий (путей) на решётке, которая характеризует память переданного сигнала. В следующем разделе мы опишем алгоритм, который выносит

посимвольное решение по правилу МАВ, основанное на наблюдении последовательности сигнальных интервалов.

Чтобы разработать алгоритм максимального правдоподобия для детектирования последовательности символов, рассмотрим в качестве примера ДБНП (NRZI)-сигнал, описанный в разд. 4.3.2. Его память характеризуется решёткой, показанной на рис. 4.3.14. На каждом сигнальном интервале имеем сигнал двоичной АМ. Следовательно, имеются два возможных передаваемых сигнала, соответствующих сигнальным точкам $s_1 = -s_2 = \sqrt{\mathcal{E}_b}$, где $\mathcal{E}_b$ – энергия на бит. Выход согласованного фильтра или корреляционного демодулятора для двоичной АМ на k-м сигнальном интервале можно выразить так:

$$r_k = \pm\sqrt{\mathcal{E}_b} + n_k, \tag{5.1.57}$$

где n_k – гауссовская случайная величина с нулевым средним и дисперсией $\sigma_n^2 = N_0/2$. Следовательно, условные ФПВ для двух возможных переданных сигналов равны

$$p(r_k|s_1) = \frac{1}{\sqrt{2\pi}\sigma_n}\exp\left[-\frac{\left(r_k - \sqrt{\mathcal{E}_b}\right)^2}{2\sigma_n^2}\right], \quad p(r_k|s_2) = \frac{1}{\sqrt{2\pi}\sigma_n}\exp\left[-\frac{\left(r_k + \sqrt{\mathcal{E}_b}\right)^2}{2\sigma_n^2}\right]. \tag{5.1.58}$$

Теперь предположим, что мы наблюдаем последовательность выходов согласованных фильтров $r_1\ r_2\ \ldots r_K$. Поскольку канальный шум считается гауссовским центрированным и белым, а сигналы $f(t-iT)$, $f(t-jT)$ ортогональны для $i \neq j$, следует $E(n_i n_j) = 0$, $i \neq j$. Следовательно, шумовая последовательность $n_1\ n_2\ \ldots n_K$ также белая. Поэтому для любой переданной последовательности $\mathbf{s}^{(m)}$ совместные ФПВ для $r_1\ r_2\ \ldots r_K$ можно выразить как произведение их собственных ФПВ, т.е.

$$\begin{aligned} p(r_1, r_2, \ldots r_K|s^{(m)}) &= \prod_{k=1}^{K} p(r_k|s^{(m)}) = \prod_{k=1}^{K}\frac{1}{\sqrt{2\pi}\sigma_n}\exp\left[-\frac{\left(r_k - s_k^{(m)}\right)^2}{2\sigma_n^2}\right] = \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma_n}\right)^K \exp\left[-\sum_{k=1}^{K}\frac{\left(r_k - s_k^{(m)}\right)^2}{2\sigma_n^2}\right], \end{aligned} \tag{5.1.59}$$

где $s_k^{(m)} = \sqrt{\mathcal{E}_b}$ или $s_k^{(m)} = -\sqrt{\mathcal{E}_b}$. Тогда по заданной последовательности $(r_1\ r_2\ \ldots r_K)$ на выходе согласованного фильтра или корреляционного демодулятора детектор определяет последовательность $\mathbf{s}^{(m)} = \left\{s_1^{(m)}, s_2^{(m)}, \ldots s_K^{(m)}\right\}$, которая максимизирует условную ФПВ $p(r_1, r_2, \ldots r_K|s^{(m)})$. Такой детектор называется *максимально правдоподобным* (МП) *последовательным детектором*.

Взяв логарифм от (5.1.59) и опуская слагаемое, не зависящее от $(r_1\ r_2\ \ldots r_K)$, находим, что эквивалент МП детектора последовательности – это детектор, выбирающий последовательность $\mathbf{s}^{(m)}$, которая минимизирует евклидово расстояние

$$D(\mathbf{r}, \mathbf{s}^{(m)}) = \sum_{k=1}^{K}\left(r_k - s_k^{(m)}\right)^2. \tag{5.1.60}$$

При поиске на решётке последовательности, которая минимизирует евклидово расстояние $D(\mathbf{r}, \mathbf{s}^{(m)})$, может показаться, что мы должны вычислить расстояния для всех возможных последовательностей. Например, для ДБНП, которая использует двоичную модуляцию, общее количество последовательностей равно 2^K, где K – число отдельных

выходов демодулятора. Однако это не так. Мы можем уменьшить число последовательностей при поиске по решётке, используя *алгоритм Витерби* для отсечения последовательностей по мере поступления новых данных от демодулятора.

Алгоритм Витерби является алгоритмом последовательного поиска на решётке для обеспечения МП декодирования последовательности. Он описывается в гл. 8 как алгоритм декодирования свёрточных кодов. Мы опишем его ниже применительно к сигналам ДБНП.

Предположим, что процесс поиска начинается первоначально с состояния S_0. Соответствующая решётка показана на рис. 5.1.11.

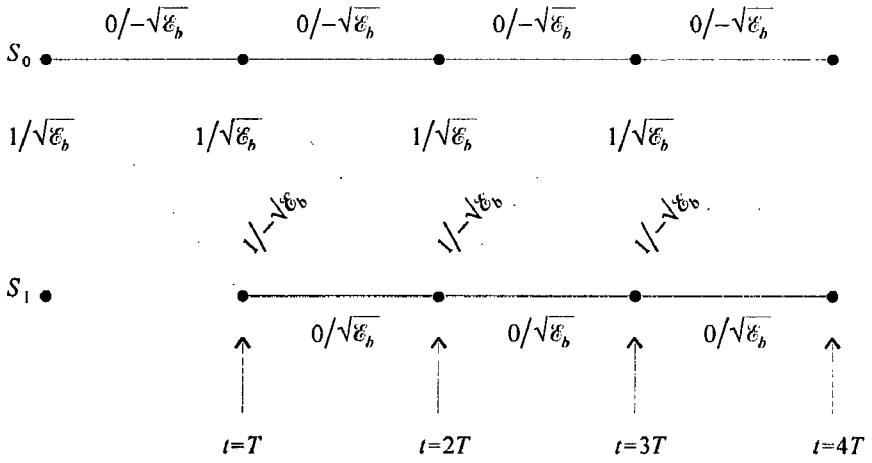


Рис. 5.1.11. Решётка для сигнала ДБНП (NRZI)

В точке $t=T$ принимаем от демодулятора $r_1 = s_1^{(m)} + n_1$, в точке $t=2T$ принимаем $r_2 = s_2^{(m)} + n_2$. Поскольку память сигнала равна одному биту, обозначим её $L=1$. Мы видим, что структура решётки регулярно повторяется (устойчивые состояния) после двух начальных переходов.

Так, при приёме r_2 в точке $t=2T$ (и так далее) мы видим, что имеется два сигнальных пути, входящие в каждый узел, и два сигнальных пути, уходящие от каждого узла.

Два пути, входящие в узел состояния S_0 при $t=2T$, соответствуют информационным битам (0, 0) и (1, 1) или, что эквивалентно, сигнальным точкам $\left(-\sqrt{\mathcal{E}_b},-\sqrt{\mathcal{E}_b}\right)$ и $\left(\sqrt{\mathcal{E}_b},-\sqrt{\mathcal{E}_b}\right)$.соответственно. Два пути, входящие в узел S_1 при $t=2T$, соответствуют информационным битам (0, 1) и (1, 0) или, что эквивалентно, сигнальным точкам $\left(-\sqrt{\mathcal{E}_b},\sqrt{\mathcal{E}_b}\right)$ $\left(\sqrt{\mathcal{E}_b},\sqrt{\mathcal{E}_b}\right)$ соответственно.

Для двух путей, входящих в узел S_0, вычислим две евклидовы метрики расстояний

$$D_0(0,0) = \left(r_1 + \sqrt{E_b}\right)^2 + \left(r_2 + \sqrt{E_b}\right)^2,\ D_0(1,1) = \left(r_1 - \sqrt{E_b}\right)^2 + \left(r_2 + \sqrt{E_b}\right)^2 \qquad (5.1.61)$$

при заданных выходах демодулятора r_1 и r_2.

Алгоритм Витерби сравнивает эти две метрики и отбрасывает путь, имеющий большую метрику (большее расстояние)[1]. Другой путь с меньшей метрикой накапливается, и его

[1] Заметим, что для ДБНП приём r_2 демодулятором не увеличивает и не уменьшает относительную разницу между двумя метриками $D_0(0,0)$ и $D_0(1,1)$. С этой точки зрения можно подумать о вовлечении этого наблюдения. Во всяком случае, мы продолжим описание детектора последовательности, основанном на алгоритме Витерби.

называют выжившим при $t = 2T$. Исключение одного из двух путей можно сделать, не нарушая оптимальность поиска по решётке, поскольку добавление пути с большим расстоянием за точки $t = 2T$ будет всегда иметь большую метрику, чем выживший путь, который сохранён после точки $t = 2T$.

Аналогично для двух путей, входящих в узел S_1 в точке $t = 2T$, мы вычислим две евклидовы метрики расстояний

$$D_1(0,1) = \left(r_1 + \sqrt{E_b}\right)^2 + \left(r_2 - \sqrt{E_b}\right)^2,$$
$$D_1(1,0) = \left(r_1 - \sqrt{E_b}\right)^2 + \left(r_2 - \sqrt{E_b}\right)^2, \qquad (5.1.62)$$

используя выходы демодулятора r_1 и r_2. Эти две метрики сравниваются, и сигнальный путь с большей метрикой исключается. Таким образом, в точке $t = 2T$ мы сохраняем два выживших пути, один в узле S_0 и другой в узле S_1 и их соответствующие метрики. Сигнальные пути в узлах S_0 и S_1 затем распределятся по двум выжившим путям.

При приёме r_3 в точке $t = 3T$ вычисляем метрики по двум путям, входящих в состояние S_0. Предположим, что выжившими при $t = 2T$ являются пути $(0,0)$ в состоянии S_0 и $(0,1)$ в состоянии S_1. Тогда две метрики для путей, входящих в S_0 при $t = 3T$, равны

$$D_0(0,0,0) = D_0(0,0) + \left(r_3 + \sqrt{E_b}\right)^2,$$
$$D_0(0,1,1) = D_1(0,1) + \left(r_3 + \sqrt{E_b}\right)^2. \qquad (5.1.63)$$

Эти две метрики сравниваются. и путь с большей метрикой (большим расстоянием) исключается. Аналогично, метрики для двух путей, входящих в S_1 при $t = 3T$, равны

$$D_1(0,0,1) = D_0(0,0) + \left(r_3 - \sqrt{E_b}\right)^2,$$
$$D_1(0,1,0) = D_1(0,1) + \left(r_3 - \sqrt{E_b}\right)^2. \qquad (5.1.64)$$

Эти две метрики сравниваются, и путь с большей метрикой (большим расстоянием) исключается.

Этот процесс продолжается, пока каждый новый сигнальный отсчёт принимается демодулятором. Таким образом, алгоритм Витерби вычисляет две метрики для двух сигнальных путей, входящих в узел на каждом шаге поиска по решётке, и исключает один из двух путей на каждом узле. Два выживших пути затем продвигаются вперёд до следующего состояния. Следовательно, число путей поиска по решётке сокращается в два раза на каждом шаге.

Относительно просто отобразить поиск по решётке по алгоритму Витерби для M-позиционной модуляции. Для примера на рис. 5.1.12 показана решётка на четыре состояния, характеризующая модуляцию с задержкой, использующую $M = 4$ сигнала.

Мы видим, что для каждого состояния два сигнальных пути входят в узел и два из него выходят.

Память сигнала равна $L = 1$. Следовательно, алгоритм Витерби будет иметь четыре выживших пути на каждом шаге и их соответствующие метрики. Две метрики, соответствующие двум входным путям, вычисляются на каждом узле, и один из двух сигнальных путей, входящих в узел, исключается в каждом состоянии решётки. Таким образом, алгоритм Витерби минимизирует число путей поиска по решётке при осуществлении МП детектора последовательности.

Из описания алгоритма Витерби, данного выше, не ясно, как выносится решение об индивидуальном детектируемом информационном символе по данным выжившим последовательностям.

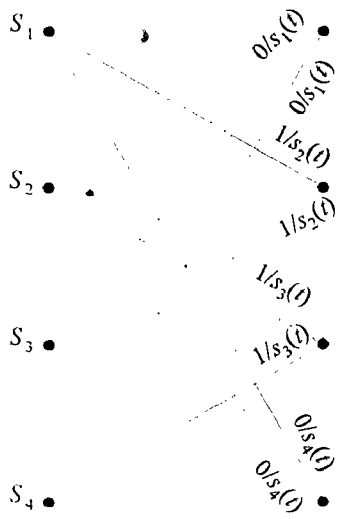


Рис. 5.1.12. Один шаг по решётчатой диаграмме для модуляции с задержкой

Если мы продвигаемся на некоторое число шагов, скажем K, где $K >> L$, по решётке и сравниваем выжившие последовательности, мы найдём, что в вероятностном смысле все выжившие последовательности сближаются в битах (или символах) на $K - 5L$ позициях или раньше[1]. При практических применениях алгоритма Витерби решения о каждом информационном бите (или символе) осуществляются после задержки на $5L$ бит (или символов), и, следовательно, выжившие последовательности запоминаются на $5L$ бит (или символов). Таким образом, избегается переменная задержка при детектировании символов. Потеря в качестве, возникающая при такой субоптимальной процедуре детектирования, несущественна, если задержка по крайнеё мере равна $5L$.

Пример 5.1.4. Рассмотрим правило решения при детектировании данных последовательности сигналов ДБНП с алгоритмом Витерби и задержкой решения на $5L$ символов.

Решётка для ДБНП показана на рис. 5.1.11. В этом случае $L = 1$, следовательно, задержка при детектировании символа простирается на 5 бит. Следовательно, при $t = 6T$ имеем две выжившие последовательности, одну для каждого из двух состояний, и соответствующие метрики $\mu_6(b_1, b_2, b_3, b_4, b_5, b_6)$ и $\mu_6(b_1', b_2', b_3', b_4', b_5', b_6')$. На этом шаге с вероятностью, близкой к единице, символ b_1 будет такой же, что и b_1'; это значит, что обе выжившие последовательности будут иметь общую первую ветвь. Если $b_1 \neq b_1'$, мы можем выбрать символ b_1 или b_1', соответствующий меньшей из двух метрик. Затем первый бит исключается из двух выживших последовательностей. При $t = 7T$, две метрики $\mu_7(b_2, b_3, b_4, b_5, b_6, b_7)$ и $\mu_7(b_2', b_3', b_4', b_5', b_6', b_7')$ используются для определения решения по символу b_2. Этот процесс повторяется на каждом шаге при поиске по решётке наименьших по расстоянию последовательностей. Таким образом, в нашем примере задержка детектора фиксирована на 5 символов[2].

5.1.5. Посимвольное детектирование для сигналов с памятью

В противоположность МП детектору последовательности для детектирования переданной информации теперь опишем детектор, который выполняет посимвольные решения, основанные на вычислении максимума апостериорной вероятности (МАВ) для каждого детектируемого символа. Следовательно, этот детектор оптимален в том смысле, что он минимизирует среднюю вероятность ошибочного приёма символа. Алгоритм детектирования, который представлен ниже, принадлежит Абенду и Фритчману (1970), которые разработали его как алгоритм детектирования для каналов с межсимвольной интерференцией, т.е. каналов с памятью.

---

[1] Задержка решения относительно отдельных кодовых символов при использовании строгого АВ может быть различной, даже неограниченной, что практически крайне нежелательно (прп).

[2] Отметим, что в этом примере последовательный МП детектор и посимвольный детектор, который игнорирует память ДБНП сигнала, принимают одинаковые решения. Поэтому нет нужды в задержке решения. Тем не менее процедуру, описанную выше, применяют в общем случае.

Мы проиллюстрируем этот алгоритм применительно к детектированию сигнала АМ с M возможными уровнями. Предположим, что необходимо детектировать информационный символ, переданный на k-м сигнальном интервале, и пусть $r_1, r_2, \ldots, r_{k+D}$ – наблюдаемая принятая последовательность, а D – параметр задержки, который выбирается так, чтобы превысить память сигнала, т.е. $D \geq L$, где L – присущая сигналу память. На основе принятой последовательности вычисляем апостериорные вероятности

$$P\left(s^{(k)} = A_m | r_{k+D}, r_{k+D-1}, \ldots r_1\right) \tag{5.1.65}$$

для M возможных позиций символов и выбираем символ с наибольшей вероятностью. Так как

$$P\left(s^{(k)} = A_m | r_{k+D}, r_{k+D-1}, \ldots r_1\right) = \frac{p\left(r_{k+D}, r_{k+D-1}, \ldots r_1 | s^{(k)} = A_m\right) P\left(s^{(k)} = A_m\right)}{p\left(r_{k+D}, r_{k+D-1}, \ldots r_1\right)}, \tag{5.1.66}$$

а знаменатель общий для всех символов, правило максимума апостериорной вероятности (МАВ) эквивалентно выбору $s^{(k)}$, который максимизирует числитель (5.1.66). Таким образом, правило выбора символа $s^{(k)}$ можно записать так:

$$\tilde{s}^{(k)} = \arg\left\{\max_{s^{(k)}} p\left(r_{k+D}, r_{k+D-1}, \ldots r_1 | s^{(k)} = A_m\right) P\left(s^{(k)} = A_m\right)\right\}. \tag{5.1.67}$$

Если символы равновероятны, вероятности $P\left(s^{(k)} = A_m\right)$ можно исключить из вычислений.

Алгоритм для вычисления вероятностей в (5.1.67) рекуррентно начинается с первого символа $s^{(1)}$. Имеем

$$\begin{aligned}
\tilde{s}^{(1)} &= \arg\left\{\max_{s^{(1)}} p\left(r_{1+D}, \ldots, r_1 | s^{(1)} = A_m\right) P\left(s^{(1)} = A_m\right)\right\} = \\
&= \arg\left\{\max_{s^{(1)}} \sum_{s^{(1+D)}} \cdots \sum_{s^{(2)}} p\left(r_{1+D}, \ldots, r_1 | s^{(1+D)}, \ldots s^{(1)}\right) P\left(s^{(1+D)}, \ldots, s^{(1)}\right)\right\} = \\
&= \arg\left\{\max_{s^{(1)}} \sum_{s^{(1+D)}} \cdots \sum_{s^{(2)}} p_1\left(s^{(1+D)}, \ldots, s^{(2)}, s^{(1)}\right)\right\},
\end{aligned} \tag{5.1.68}$$

где $\tilde{s}^{(1)}$ означает решение об $s^{(1)}$, и для математического удобства мы обозначили

$$p_1\left(s^{(1+D)}, \ldots, s^{(2)}, s^{(1)}\right) \equiv p\left(r_{1+D}, \ldots, r_1 | s^{(1+D)}, \ldots s^{(1)}\right) P\left(s^{(1+D)}, \ldots, s^{(1)}\right). \tag{5.1.69}$$

Совместную вероятность $P\left(s^{(1+D)}, \ldots, s^{(2)}, s^{(1)}\right)$ можно опустить, если все символы равновероятны и статистически независимы. Как следствие статистической независимости последовательности отсчётов аддитивного шума, имеем

$$\begin{aligned}
&p\left(r_{1+D}, \ldots, r_1 | s^{(1+D)}, \ldots s^{(1)}\right) = \\
&= p\left(r_{1+D} | s^{(1+D)}, \ldots, s^{(1+D-L)}\right) p\left(r_D | s^{(D)}, \ldots, s^{(D-L)}\right) \ldots p\left(r_2 | s^{(2)}, s^{(1)}\right) p\left(r_1 | s^{(1)}\right),
\end{aligned} \tag{5.1.70}$$

где мы предположили, что $s^{(k)} = 0$ для $k \leq 0$.

Для детектирования символа $s^{(2)}$ имеем

$$\begin{aligned}
\tilde{s}^{(2)} &= \arg\left\{\max_{s^{(2)}} p\left(r_{2+D}, \ldots, r_1 | s^{(2)} = A_m\right) P\left(s^{(2)} = A_m\right)\right\} = \\
&= \arg\left\{\max_{s^{(2)}} \sum_{s^{(2+D)}} \cdots \sum_{s^{(3)}} p\left(r_{2+D}, \ldots, r_1 | s^{(2+D)}, \ldots s^{(2)}\right) P\left(s^{(2+D)}, \ldots, s^{(2)}\right)\right\}.
\end{aligned} \tag{5.1.71}$$

Совместные условные плотности вероятности в суммах можно выразить так:

$$p\left(r_{2+D},\ldots,r_1|s^{(2+D)},\ldots s^{(2)}\right)=p\left(r_{2+D}|s^{(2+D)},\ldots,s^{(2+D-L)}\right)p\left(r_{1+D},\ldots,r_1|s^{(1+D)},\ldots,s^{(2)}\right). \quad (5.1.72)$$

Далее совместную плотность вероятности

$$p\left(r_{1+D},\ldots,r_1|s^{(1+D)},\ldots s^{(2)}\right)P\left(s^{(1+D)},\ldots,s^{(2)}\right)$$

можно получить через плотности вероятности, вычисленные ранее при детектировании $s^{(1)}$. Это даёт

$$\begin{aligned}&p\left(r_{1+D},\ldots,r_1|s^{(1+D)},\ldots,s^{(2)}\right)=\\&=\sum_{s^{(1)}}p\left(r_{1+D},\ldots,r_1|s^{(1+D)},\ldots,s^{(1)}\right)P\left(s^{(1+D)},\ldots,s^{(1)}\right)=\sum_{s^{(1)}}p_1\left(s^{(1+D)},\ldots,s^{(2)},s^{(1)}\right).\end{aligned} \quad (5.1.73)$$

Комбинируя (5.1.73) и (5.1.72), а затем подставив их в (5.1.71), получим

$$\tilde{s}^{(2)}=\arg\left\{\max_{s^{(2)}}\sum_{s^{(2+D)}}\cdots\sum_{s^{(3)}}p_2\left(s^{(2+D)},\ldots,s^{(3)},s^{(2)}\right)\right\}, \quad (5.1.74)$$

где по определению

$$p_2\left(s^{(2+D)},\ldots,s^{(3)},s^{(2)}\right)=p\left(r_{2+D}|s^{(2+D)},\ldots,s^{(2+D-L)}\right)P\left(s^{(2+D)}\right)\sum_{s^{(1)}}p_1\left(s^{(1+D)},\ldots,s^{(2)},s^{(1)}\right). \quad (5.1.75)$$

В общем рекуррентный алгоритм для детектирования символа $s^{(k)}$ после приёма $r_{k+D},r_{k+D-1},\ldots r_2,r_1$ можно записать в виде

$$\begin{aligned}\tilde{s}^{(k)}&=\arg\left\{\max_{s^{(k)}}p\left(r_{k+D},r_{k+D-1},\ldots r_1|s^{(k)}\right)P\left(s^{(k)}\right)\right\}=\\&=\arg\left\{\max_{s^{(k)}}\sum_{s^{(k+D)}}\cdots\sum_{s^{(k)}}p_k\left(s^{(k+D)},\ldots,s^{(k+1)},s^{(k)}\right)\right\},\end{aligned} \quad (5.1.76)$$

где по определению

$$\begin{aligned}&p_k\left(s^{(k+D)},\ldots,s^{(k+1)},s^{(k)}\right)=\\&=p\left(r_{k+D}|s^{(k+D)},\ldots,s^{(k+D-L)}\right)P\left(s^{(k+D)}\right)\sum_{s^{(k-1)}}p_{k-1}\left(s^{(k-1+D)},\ldots,s^{(k)},s^{(k-1)}\right).\end{aligned} \quad (5.1.77)$$

Таким образом, рекуррентный характер алгоритма выражается соотношениями (5.1.76) и (5.1.77).

Основная проблема с этим алгоритмом – вычислительная сложность. В частности, усреднение, выполняемое над символами $s^{(k+D)},\ldots,s^{(k+1)},s^{(k)}$ в (5.1.76)? требует большое число вычислений на одном такте принимаемого сигнала, особенно если число M уровней амплитуд $\{A_m\}$ велико. С другой стороны, если M мало и память относительно невелика, этот алгоритм легко выполняется.

5.2. ХАРАКТЕРИСТИКИ КАЧЕСТВА ОПТИМАЛЬНОГО ПРИЁМНИКА ДЛЯ МОДУЛЯЦИИ БЕЗ ПАМЯТИ

В этом разделе мы определим вероятность ошибки для модулированных сигналов без памяти, описанных в разд. 4.3.1. Сначала рассмотрим сигналы двоичной АМ, а затем M-позиционные сигналы различных видов.

5.2.1. Вероятность ошибки при двоичной модуляции

Рассмотрим сигнал двоичной АМ, где два сигнала – $s_1(t)=g(t)$ и $s_2(t)=-g(t)$, а $g(t)$ – произвольный импульс, который отличен от нуля на интервале $0\le t\le T_b$ и равен нулю в остальной области.

Поскольку $s_1(t) = -s_2(t)$, эти сигналы называют *противоположными*. Пусть энергия импульса $g(t)$ равна $\mathscr{E}_b$. Как указано в разд. 4.3.1, сигналы АМ является одномерными, и, следовательно, их геометрическое представление определяется одномерными векторами $s_1 = \sqrt{\mathscr{E}_b}$, $s_2 = -\sqrt{\mathscr{E}_b}$. Рисунок 5.2.1 иллюстрирует две сигнальные точки.

Предположим, что два сигнала равновероятны и что передан сигнал $s_1(t)$. Тогда выходной сигнал (на выходе корреляционной схемы или согласованного фильтра) демодулятора равен

$$r = s_1 + n = \sqrt{\mathscr{E}_b} + n, \tag{5.2.1}$$

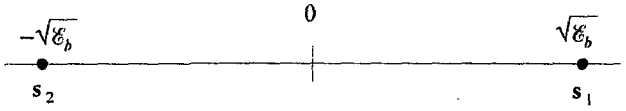


Рис. 5.2.1. Сигнальные точки для двоичных противоположных сигналов

где n представляет компоненту аддитивного гауссовского шума, которая имеет нулевое среднее и дисперсию $\sigma^2 = \frac{1}{2}N_0$. В этом случае правило решения, основанное на корреляционной метрике (5.1.44), сравнивает r с нулевым порогом. Если $r > 0$, то решение принимается в пользу $s_1(t)$, а если $r < 0$ – в пользу $s_2(t)$. Ясно, что условные ФПВ для r равны

$$p(r|s_1) = \frac{1}{\sqrt{\pi N_0}} e^{-\left(r-\sqrt{\mathscr{E}_b}\right)^2/N_0}, \tag{5.2.2}$$

$$p(r|s_2) = \frac{1}{\sqrt{\pi N_0}} e^{-\left(r+\sqrt{\mathscr{E}_b}\right)^2/N_0}. \tag{5.2.3}$$

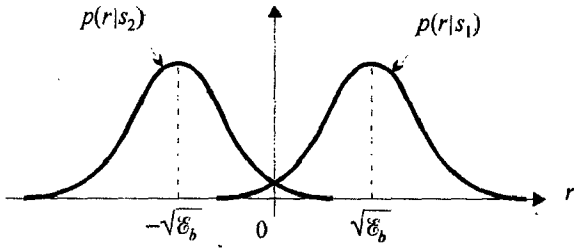


Рис. 5.2.2. Условные ФПВ для двух сигналов

Эти две условные ФПВ показаны на рис. 5.2.2. При условии, что передан сигнал $s_1(t)$, вероятность ошибки определяется вероятностью того, что $r < 0$, т.е.

$$P(e|s_1) = \int_{-\infty}^{0} p(r|s_1)\,dr = \frac{1}{\sqrt{\pi N_0}} \int_{-\infty}^{0} \exp\left[-\frac{\left(r-\sqrt{\mathscr{E}_b}\right)^2}{N_0}\right] dr =$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-\sqrt{2\mathscr{E}_b/N_0}} e^{-x^2/2}\,dx = \frac{1}{\sqrt{2\pi}} \int_{\sqrt{2\mathscr{E}_b/N_0}}^{\infty} e^{-x^2/2}\,dx = Q\left(\sqrt{\frac{2\mathscr{E}_b}{N_0}}\right), \tag{5.2.4}$$

где $Q(x)$ – Q-функция, определённая (2.1.97). Аналогично, если предположим, что был передан сигнал $s_2(t)$, то $r = -\sqrt{\mathscr{E}_b} + n$ и вероятность того, что $r > 0$, также равна $P(e|s_2) = Q\left(\sqrt{2\mathscr{E}_b/N_0}\right)$. Поскольку сигналы $s_1(t)$ и $s_2(t)$ равновероятны, то средняя вероятность ошибки

$$P_b = \tfrac{1}{2}P(e|s_1) + \tfrac{1}{2}P(e|s_2) = Q\left(\sqrt{\frac{2\mathscr{E}_b}{N_0}}\right). \tag{5.2.5}$$

Мы хотим отметить два важных свойства этой характеристики качества. Во-первых, заметим, что вероятность ошибки зависит только от отношения $\mathscr{E}_b/N_0$ и не зависит от

других более детальных характеристик сигналов и шума. Во-вторых, заметим, что $2\mathscr{E}_b/N_0$ также выходное отношение сигнал/шум (ОСШ$_0$) согласованного фильтра (и коррелятора) демодулятора. Отношение $\mathscr{E}_b/N_0$ обычно называют *отношением сигнал/шум на бит.*

Отметим также, что вероятность ошибки можно выразить через расстояние между сигналами s_1 и s_2. Из рис. 5.2.1 видно, что два сигнала находятся на расстоянии $d_{12}=2\sqrt{\mathscr{E}_b}$. Подставив $\mathscr{E}_b=\frac{1}{4}d_{12}^2$ в (5.2.5), получим

$$P_b=Q\left(\sqrt{\frac{d_{12}^2}{2N_0}}\right). \tag{5.2.6}$$

Это выражение иллюстрирует зависимость вероятности ошибки от расстояния между двумя сигнальными точками.

Далее определим вероятность ошибки для двоичных ортогональных сигналов. Повторим, что в этом случае сигнальные векторы $\mathbf{s}_1$ и $\mathbf{s}_2$ являются двухмерными, как показано на рис. 5.2.3, и их можно выразить согласно (4.3.30) так:

$$\begin{aligned}\mathbf{s}_1&=\left[\sqrt{\mathscr{E}_b}\quad 0\right],\\ \mathbf{s}_2&=\left[0\quad\sqrt{\mathscr{E}_b}\right],\end{aligned} \tag{5.2.7}$$

где $\mathscr{E}_b$ обозначает энергию для каждого из сигналов. Заметим, что расстояние между сигнальными точками теперь $d_{12}=\sqrt{2\mathscr{E}_b}$.

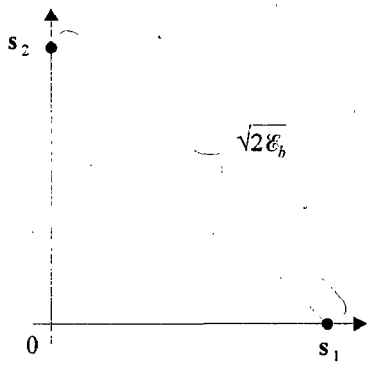


Рис. 5.2.3. Сигнальные точки для двоичных ортогональных сигналов

Для расчёта вероятности ошибки предположим, что передаётся сигнал $\mathbf{s}_1$. Тогда принимаемый вектор на выходе демодулятора

$$\mathbf{r}=\left[\sqrt{\mathscr{E}_b}+n_1\quad n_2\right]. \tag{5.2.8}$$

Мы можем теперь подставить $\mathbf{r}$ в корреляционные метрики, определяемые (5.1.44), чтобы получить $C(\mathbf{r},\mathbf{s}_1)$ и $C(\mathbf{r},\mathbf{s}_2)$ Вероятность ошибки – это вероятность того, что $C(\mathbf{r},\mathbf{s}_2)>C(\mathbf{r},\mathbf{s}_1)$. Таким образом,

$$P(e|\mathbf{s}_1)=P\left[C(\mathbf{r},\mathbf{s}_2)>C(\mathbf{r},\mathbf{s}_1)\right]=P\left[n_2-n_1>\sqrt{\mathscr{E}_b}\right]. \tag{5.2.9}$$

Поскольку n_1 и n_2 – статистически независимые гауссовские случайные величины с нулевым средним и дисперсией $\frac{1}{2}N_0$, то $x=n_2-n_1$ – гауссовская случайная величина с нулевым средним и дисперсией N_0. Следовательно,

$$P\left[n_2-n_1>\sqrt{\mathscr{E}_b}\right]=\frac{1}{\sqrt{2\pi N_0}}\int_{\sqrt{2\mathscr{E}_b}}^{\infty}e^{-x^2/2N_0}\,dx=\frac{1}{\sqrt{2\pi}}\int_{\sqrt{2\mathscr{E}_b/N_0}}^{\infty}e^{-x^2/2}\,dx=Q\left(\sqrt{\frac{\mathscr{E}_b}{N_0}}\right). \tag{5.2.10}$$

Вследствие симметрии та же вероятность ошибки получается в предположении, что передаётся $\mathbf{s}_2$. Следовательно, средняя вероятность ошибки для двоичных ортогональных сигналов

$$P_b=Q\left(\sqrt{\frac{\mathscr{E}_b}{N_0}}\right)=Q\left(\sqrt{\gamma_b}\right), \tag{5.2.11}$$

где по определению γ_b – это ОСШ на бит.

Если сравним вероятность ошибки для двоичных противоположных сигналов с вероятностью ошибки для двоичных ортогональных сигналов, то находим, что ортогональные сигналы требуют удвоения энергии сигнала для достижения той же вероятности ошибки, что в системе с противоположными сигналами. Поскольку $10\lg 2 = 3$ дБ, то видим, что ортогональные сигналы на 3 дБ хуже, чем противоположные сигналы. Разница в 3 дБ объясняется тем, что расстояние между двумя сигнальными точками ортогональной системы равно $d_{12}^2 = 2\mathscr{E}_b$, в то время как расстояние между точками противоположных сигналов равно $d_{12}^2 = 4\mathscr{E}_b$.

Зависимость вероятности ошибки от $10\lg \mathscr{E}_b / N_0$ для этих двух типов сигналов показана на рис. 5.2.4. Как видно из рисунка, для любой заданной вероятности ошибки требуемое значение $\mathscr{E}_b / N_0$ для ортогональных сигналов больше, чем для противоположных сигналов.

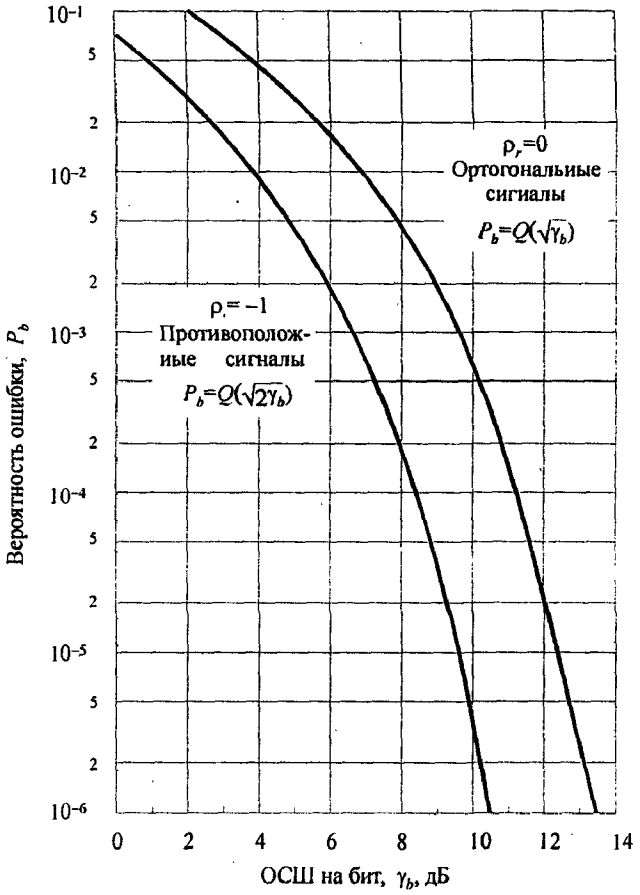


Рис. 5.2.4. Вероятность ошибки для двоичных сигналов

5.2.2. Вероятность ошибки для M-позиционных ортогональных сигналов

Для ортогональных сигналов равной энергии оптимальный детектор выберет сигнал, который приводит к наибольшей корреляции между принимаемым вектором $\mathbf{r}$ и каждым из M возможных к передаче сигнальных векторов $\{\mathbf{s}_m\}$, т.е.

$$C(\mathbf{r}, \mathbf{s}_m) = \mathbf{r} \cdot \mathbf{s}_m = \sum_{k=1}^{M} r_k s_{mk}, \quad m = 1, 2, \ldots, M. \tag{5.2.12}$$

Чтобы рассчитать вероятность ошибки, предположим, что передаётся сигнал $\mathbf{s}_1$. Тогда принимаемый сигнальный вектор

$$\mathbf{r} = \left[\sqrt{\mathscr{E}_s} + n_1 \;\; n_2 \;\; n_3 \;\; \ldots n_M\right], \tag{5.2.13}$$

где $n_1, n_2, n_3, \ldots n_M$ – взаимно независимые случайные гауссовские величины с нулевыми средними и дисперсией $\sigma_n^2 = \frac{1}{2} N_0$. В этом случае выходы набора M корреляторов равны

$$\begin{aligned} C(\mathbf{r}, \mathbf{s}_1) &= \sqrt{\mathscr{E}_s}\left(\sqrt{\mathscr{E}_s} + n_1\right), \\ C(\mathbf{r}, \mathbf{s}_2) &= \sqrt{\mathscr{E}_s} n_2, \\ &\vdots \qquad \vdots \\ C(\mathbf{r}, \mathbf{s}_M) &= \sqrt{\mathscr{E}_s} n_M. \end{aligned} \tag{5.2.14}$$

Заметим, что скалярный множитель $\mathscr{E}_s$ можно исключить путём деления всех выходов на $\sqrt{\mathscr{E}_s}$. Тогда с учётом нормирования ФПВ сигнала на выходе первого коррелятора $\left(r_1 = \sqrt{\mathscr{E}_s} + n_1\right)$ равна

$$p_{r_1}(x_1)=\frac{1}{\sqrt{\pi N_0}}\exp\left[-\frac{\left(x_1-\sqrt{\mathscr{E}_S}\right)^2}{N_0}\right], \qquad (5.2.15)$$

а ФПВ сигналов на выходах остальных $M-1$ корреляторов равны

$$p_{r_m}(x_m)=\frac{1}{\sqrt{\pi N_0}}\exp\left[-\frac{x_m^2}{N_0}\right], \quad m=2,3,\ldots,M. \qquad (5.2.16)$$

Математически удобно сначала найти вероятность того, что детектор осуществляет правильный приём. Это вероятность того, что r_1 больше, чем каждый из $M-1$ выходов корреляторов $n_2,n_3,\ldots,n_m$. Вероятность этого события определяется так:

$$P_c=\int_{-\infty}^{\infty}P\left(n_2<r_1,n_3<r_1,\ldots,n_M<r_1\middle|r_1\right)p(r_1)dr_1, \qquad (5.2.17)$$

где $P\left(n_2<r_1,n_3<r_1,\ldots,n_M<r_1\middle|r_1\right)$ – совместная вероятность того, что $n_2,n_3,\ldots,n_m$ меньше, чем r_1 при данном r_1. Затем эта совместная вероятность усредняется по всем r_1. Так как $\{r_m\}$ статистически независимы, то совместная вероятность определяется произведением $M-1$ собственных вероятностей вида

$$P\left(n_m<r_1\middle|r_1\right)=\int_{-\infty}^{r_1}p_{r_m}(x_m)dx_m=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{r_1\sqrt{2/N_0}}e^{-x^2/2}dx. \qquad (5.2.18)$$

Эти вероятности одинаковы для $m=2,3,\ldots,M$, и, следовательно, совместная вероятность приводит к (5.2.18) в степени $M-1$. Таким образом, вероятность правильного приёма

$$P_c=\int_{-\infty}^{\infty}\left(\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{r_1\sqrt{2/N_0}}e^{-x^2/2}dx\right)^{M-1}p(r_1)dr_1, \qquad (5.2.19)$$

а вероятность ошибки (k-битового) символа равна

$$P_M=1-P_c, \qquad (5.2.20)$$

где

$$P_M=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\left[1-\left(\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y}e^{-x^2/2}dx\right)^{M-1}\right]\exp\left[-\frac{1}{2}\left(y-\sqrt{\frac{2\mathscr{E}_s}{N_0}}\right)^2\right]dy. \qquad (5.2.21)$$

Такое же выражение для вероятности ошибки получим при передаче любого из других $M-1$ сигналов. Поскольку все сигналы равновероятны, то выражение P_M по формуле (5.2.21) определяет и среднюю вероятность ошибки. Расчёт по этой формуле можно выполнить численно.

Для сравнения качества различных методов цифровой модуляции желательно иметь зависимость вероятности ошибки от ОСШ на бит $\mathscr{E}_b/N_0$, вместо ОСШ на символ $\mathscr{E}_s/N_0$. При $M=2^k$ каждый символ передаёт k бит информации, и, следовательно, $\mathscr{E}_s=k\mathscr{E}_b$. Таким образом, (5.2.21) можно выразить через $\mathscr{E}_b/N_0$ подстановкой $\mathscr{E}_s$.

Иногда также желательно выразить вероятность ошибки символа через эквивалентную вероятность ошибки на бит[1].

[1] Более общее понятие «эквивалентной вероятности ошибки», которое позволяет сравнивать между собой самые различные цифровые системы (с различными кодами и методами модуляции) в произвольных каналах введено Л.М. Финком [ ] (прп).

Для равновероятных ортогональных сигналов все вероятности ошибки на символ равновероятны, и они возникают с вероятностью

$$\frac{P_M}{M-1}=\frac{P_M}{2^k-1}. \qquad (5.2.22)$$

Далее имеется $\binom{k}{n}$ возможностей путей, при которых из k переданных битов n приняты с ошибкой. Следовательно, среднее число ошибочных битов на k-битовый символ равно

$$\sum_{n=1}^{k} n\binom{k}{n}\frac{P_M}{2^k-1}=k\frac{2^{k-1}}{2^k-1}P_M, \qquad (5.2.23)$$

а средняя вероятность ошибки на бит точно определяется делением (5.2.23) на k – число бит на символ. Таким образом,

$$P_b=\frac{2^{k-1}}{2^k-1}P_M\approx\frac{P_M}{2}, \quad k>>1. \qquad (5.2.24)$$

Кривые зависимости вероятности ошибки на бит от ОСШ на бит $\mathscr{E}_b/N_0$ даны на рис. 5.2.5 для $M=2, 4, 8, 16, 32$ и 64.

Эти кривые показывают, что с увеличением числа сигналов M можно уменьшить ОСШ на бит, требуемое для заданной вероятности ошибки на бит. Например, чтобы достичь $P_b=10^{-5}$, для $M=2$ требуется ОСШ на бит немного больше, чем 12 дБ, но если M увеличить до 64 сигналов (k=6 бит/символ), требуемое ОСШ на бит станет равным примерно 6 дБ. Таким образом, реализуется экономия выше 6 дБ (сокращение в 4 раза) в передаваемой мощности (или энергии) для достижения $P_b=10^{-5}$ при увеличении числа сигналов M от 2 до 64. Каково минимальное значение $\mathscr{E}_b/N_0$ для достижения произвольной малой вероятности ошибки при $M\to\infty$? На этот вопрос ответим ниже.

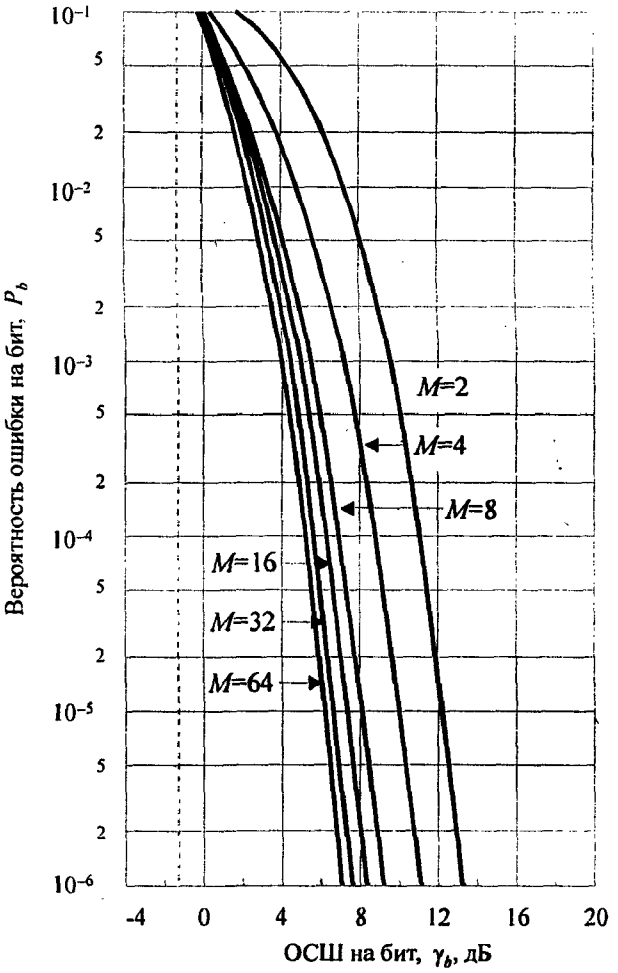


Рис. 5.2.5. Вероятность ошибки на бит для когерентного детектирования ортогональных сигналов

Объединённая граница для вероятности ошибки[1]. Рассмотрим влияние роста M на вероятность ошибки для ортогональных сигналов. Чтобы облегчить математический анализ, сначала найдём верхнюю границу для вероятности ошибки на символ, которая намного проще, чем точная формула (5.2.21).

Напомним, что вероятность ошибки для двоичных ортогональных сигналов даётся формулой (5.2.11). Теперь будем рассматривать детектор для M ортогональных сигналов

[1] Эту границу в литературе чаще всего именуют аддитивной верхней границей (прп).

как такой, который выполняет $M-1$ двоичных решений между выходом коррелятора $C(\mathbf{r},\mathbf{s}_1)$, который содержит сигнал, и остальными $M-1$ выходами корреляторов $C(\mathbf{r},\mathbf{s}_m)$, $m=2, 3, \ldots, M$.

Вероятность ошибки ограничена сверху объединённой границей для вероятности $M-1$ событий. Это означает, что если E_i представляет событие, что $C(\mathbf{r},\mathbf{s}_i) > C(\mathbf{r},\mathbf{s}_1)$ для $i \neq 1$, тогда имеем $P_M = P\left(\bigcup_{i=1}^{n} E_i\right) \leq \sum_{i=1}^{n} E_i$. Следовательно,

$$P_M \leq (M-1)P_2 = (M-1)Q\left(\sqrt{\mathscr{E}_s / N_0}\right) < MQ\left(\sqrt{\mathscr{E}_s / N_0}\right). \quad (5.2.25)$$

Эту границу можно упростить посредством верхней границы для $Q\left(\sqrt{\mathscr{E}_s / N_0}\right)$. Имеем

$$Q\left(\sqrt{\mathscr{E}_s / N_0}\right) < \mathrm{e}^{-\mathscr{E}_s/2N_0}. \quad (5.2.26)$$

Таким образом,

$$P_M < M\mathrm{e}^{-\mathscr{E}_s/2N_0} = 2^k \mathrm{e}^{-k\mathscr{E}_b/2N_0},$$
$$P_M < \mathrm{e}^{-k\left(\mathscr{E}_b/N_0 - 2\ln 2\right)/2}. \quad (5.2.27)$$

При $k \to \infty$, что эквивалентно $M \to \infty$, вероятность ошибки экспоненциально стремится к нулю при условии, что $\mathscr{E}_b / N_0$ больше, чем $2\ln 2$, т.е.

$$\mathscr{E}_b / N_0 > 2\ln 2 = 1{,}39 \quad (1{,}42\text{ дБ}). \quad (5.2.28)$$

Простая верхняя граница для вероятности ошибки, определяемая (5.2.27), подразумевает, что когда ОСШ на бит больше, чем 1,42 дБ, то мы можем достичь произвольно малую вероятность ошибки P_M. Однако эта объединённая граница не является очень плотной границей при достаточно низком ОСШ, что объясняется тем фактом, что верхняя граница для Q-функции в (5.2.26) является неточной. Действительно, посредством более тщательного исследования границ в гл. 7 показано, что верхняя граница (5.2.27) достаточно плотная при $\mathscr{E}_b / N_0 > 4\ln 2$. Для $\mathscr{E}_b / N_0 < 4\ln 2$ плотная верхняя граница для P_m определяется так:

$$P_M < 2\mathrm{e}^{-k\left(\sqrt{\mathscr{E}_b/N_0} - \sqrt{\ln 2}\right)^2}. \quad (5.2.29)$$

Следовательно, $P_M \to 0$ при $k \to \infty$ при условии, что

$$\mathscr{E}_b / N_0 > \ln 2 = 0{,}693 \quad (-1{,}6\text{ дБ}). \quad (5.2.30)$$

Таким образом, 1,6 дБ – это минимальное ОСШ на бит, требуемое для достижения произвольной сколь угодно малой вероятности ошибки в пределе, когда $k \to \infty$ $(M \to \infty)$. Это минимальное значение ОСШ на бит (–1,6 дБ) названо *пределом Шеннона* для канала с аддитивным белым гауссовским шумом.

5.2.3. Вероятность ошибки для M-позиционной биортогональной системы сигналов

Как указано в разд. 4.3, ансамбль из $M = 2^k$ биортогональных сигналов конструируется из $\frac{1}{2}M$ ортогональных сигналов путем его дополнения сигналами, которые противоположны ортогональным сигналам. Так мы достигаем уменьшения сложности демодулятора для биортогональных сигналов относительно демодулятора такого же количества ортогональных сигналов, так как он требует лишь $\frac{1}{2}M$ взаимных корреляторов или согласованных фильтров вместо M согласованных фильтров или взаимных корреляторов.

Чтобы рассчитать вероятность ошибки для оптимального детектора, предположим, что

был передан сигнал $s_1(t)$, которому соответствует вектор $\mathbf{s}_1 = \left[\sqrt{\mathscr{E}_b} \;\; 0 \;\; 0 \ldots 0\right]$. Тогда вектор принимаемого сигнала

$$\mathbf{r} = \left[\sqrt{\mathscr{E}_s} + n_1 \;\; n_2 \;\; \ldots \;\; n_{M/2}\right], \tag{5.2.31}$$

где $\{n_m\}$ – взаимно независимые гауссовские случайные величины с нулевыми средними и дисперсией $\sigma_n^2 = \frac{1}{2} N_0$. Оптимальный детектор выбирает решение в пользу сигнала, которому соответствует максимальное значение взаимной корреляции

$$C(\mathbf{r}, \mathbf{s}_m) = \mathbf{r} \cdot \mathbf{s}_m = \sum_{k=1}^{M/2} r_k s_{mk}, \quad m = 1, 2, \ldots \tfrac{1}{2} M, \tag{5.2.32}$$

причём знак наибольшего слагаемого используется для решения о том, передан ли сигнал $s_m(t)$ или $-s_m(t)$. Согласно этому правилу решения вероятность правильного решения равна вероятности того, что $r_1 = \sqrt{\mathscr{E}_s} + n_1 > 0$, и r_1 превышает по модулю $|r_m| = |n_m|$, $m = 2, 3, \ldots, \frac{1}{2} M$. Но

$$P\left[|n_m| < r_1 | r_1 > 0\right] = \frac{1}{\sqrt{\pi N_0}} \int_{r_1}^{r_1} e^{-x^2/N_0}\, dx = \frac{1}{\sqrt{2\pi}} \int_{r_1/\sqrt{N_0/2}}^{r_1/\sqrt{N_0/2}} e^{-x^2/2}\, dx. \tag{5.2.33}$$

Вероятность правильного решения равна

$$P_c = \int_0^\infty \left(\frac{1}{\sqrt{2\pi}} \int_{-r_1/\sqrt{N_0/2}}^{r_1/\sqrt{N_0/2}} e^{-x^2/2}\, dx \right)^{M/2-1} p(r_1) dr_1.$$

Из этой формулы, подставив выражение для $p(r_1)$, получим

$$P_c = \int_{\sqrt{2\mathscr{E}_s/N_0}}^\infty \left(\frac{1}{\sqrt{2\pi}} \int_{\left(v+\sqrt{2\mathscr{E}_s/N_0}\right)}^{v+\sqrt{2\mathscr{E}_s/N_0}} e^{-x^2/2}\, dx \right)^{M/2-1} e^{-v^2/2} dv, \tag{5.2.34}$$

где мы использовали ФПВ, определяемое (5.2.15). Окончательно вероятность ошибки на символ $P_M = 1 - P_c$.

Вероятность P_c и, следовательно, P_M можно численно рассчитать для различных значений M по (5.2.34). Кривые на рис. 5.2.6 иллюстрируют зависимость P_M как функцию от $\mathscr{E}_b / N_0$, где $\mathscr{E}_s = k\mathscr{E}_b$, для M=2, 4, 8, 16, 32.

Видим, что эти кривые похожи на те, которые определяют систему ортогональных сигналов (см. рис. 5.2.5). Однако в этом случае вероятность ошибки для $M = 4$ больше, чем при $M = 2$. Это объясняется тем, что на рис. 5.2.6 показана зависимость для P_M. Если бы мы показали зависимость эквивалентной вероятности ошибки, то можно было бы видеть, что кривые при $M = 2$ и $M = 4$ совпадают. Как и в случае ортогональных сигналов при $M \to \infty$ (при $k \to \infty$) требуемое минимальное значение $\mathscr{E}_b / N_0$ для достижения произвольно малой вероятности ошибки равна –1,6 дБ, т.е. пределу Шеннона.

5.2.4. Вероятность ошибки для симплексных сигналов

Теперь рассмотрим вероятность ошибки для M симплексных сигналов. Напомним из разд. 4.3, что симплексные сигналы образуют ансамбль из M одинаково коррелированных сигналов с коэффициентом взаимной корреляции $\rho_{mn} = -1/(M-1)$. Эти сигналы имеют одинаковое минимальное расстояние $\sqrt{2\mathscr{E}_s}$ между соседними сигнальными точками в M-мерном пространстве, как у ортогональных сигналов. Они достигают такое взаимное разделение посредством передаваемой энергии $\mathscr{E}_s(M-1)/M$, которая меньше, чем требуется для ортогональных сигналов, в $(M-1)/M$ раз. Следовательно, формула для вероятности ошибки системы симплексных сигналов такая же, как для ортогональных сигналов, но достигается экономия в ОСШ на

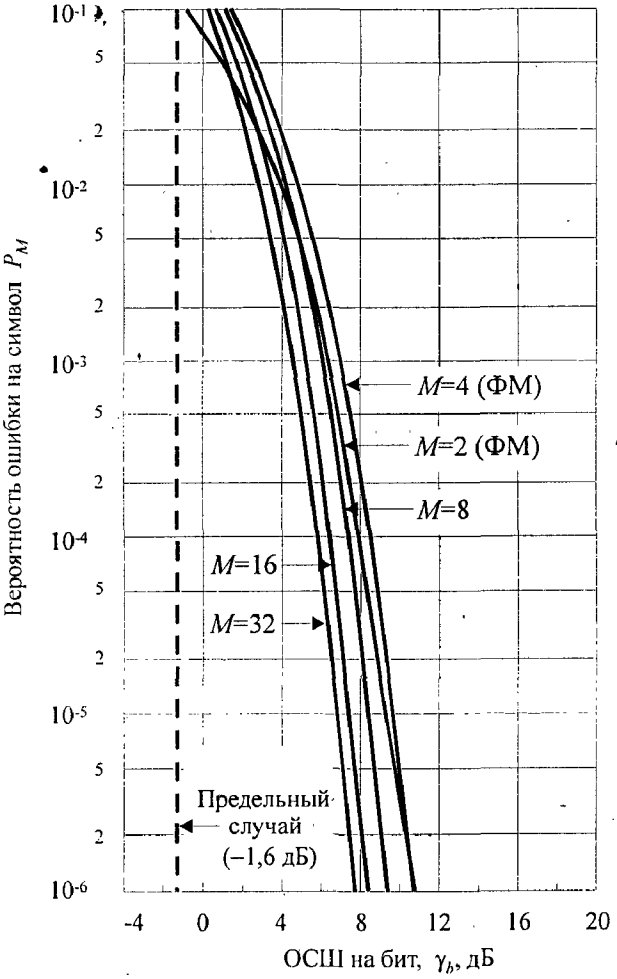


Рис. 5.2.6. Вероятность ошибки на символ для биортогональных сигналов

$$10\lg(1-\rho) = 10\lg\frac{M}{M-1} \text{ дБ}. \quad (5.2.35)$$

Для $M=2$ экономия составит 3 дБ. Однако по мере увеличения M экономия в ОСШ стремится к 0 дБ.

5.2.5. Вероятность ошибки для M-позиционной системы с двоичными кодовыми сигналами

Мы видели в разд. 4.3, что двоичные кодированные сигналы можно представить сигнальным вектором

$$\mathbf{s}_m = [s_{m_1}\ s_{m_2}\ \dots\ s_{mN}],\quad m = 1, 2, \dots, M,$$

где $s_{m_j} = \pm\sqrt{\mathscr{E}/N}$ для всех m и j, N – длина кодового блока, и она является также размерностью M-позиционного сигнала.

Если $d_{\min}^{(e)}$ – минимальное евклидово расстояние между парой сигналов из M возможных, тогда вероятность ошибки приёма кодового блока определяется верхней границей так:

$$P_m < (M-1)P_b = (M-1)Q\left(\sqrt{\frac{\left(d_{\min}^{(e)}\right)^2}{2N_0}}\right) < 2^k \exp\left[-\frac{\left(d_{\min}^{(e)}\right)^2}{4N_0}\right]. \quad (5.2.36)$$

Величина минимального евклидова расстояния будет зависеть от выбора кодовых слов, т.е. синтеза кода.

5.2.6. Вероятность ошибки для M-позиционной АМ

Напомним, что M-позиционные сигналы АМ представляются геометрически как M одномерных сигнальных точек со значениями

$$s_m = \sqrt{\tfrac{1}{2}\mathscr{E}_g}\,A_m,\quad m = 1,2,\dots,M, \quad (5.2.37)$$

где $\mathscr{E}_g$ – энергия базового сигнального импульса $g(t)$. Значения амплитуд можно выразить так:

$$A_m = (2m-1-M)d,\quad m = 1,2,\dots,M, \quad (5.2.38)$$

где евклидово расстояние между соседними сигнальными точками равно $d\sqrt{2\mathscr{E}_g}$.

Средняя энергия сигнала

$$\mathscr{E}_{\text{ср}} = \frac{1}{M}\sum_{m=1}^{M}\mathscr{E}_m = \frac{d^2\mathscr{E}_g}{2M}\sum_{m=1}^{M}(2m-1-M)^2 = \frac{d^2\mathscr{E}_g}{2M}\left[\tfrac{1}{3}M\left(M^2-1\right)\right] = \tfrac{1}{6}\left(M^2-1\right)d^2\mathscr{E}_g. \quad (5.2.39)$$

Эквивалентно мы можем характеризовать эти сигналы их средней мощностью, которая равна

$$P_{\text{ср}} = \frac{\mathscr{E}_{\text{ср}}}{T} = \tfrac{1}{6}\left(M^2 - 1\right)\frac{d^2\mathscr{E}_g}{T}. \qquad (5.2.40)$$

Средняя вероятность ошибки для M-позиционной АМ можно определить из правила выбора решения по максимуму метрик корреляции, определяемых (5.1.44). Эквивалентно детектор сравнивает выход демодулятора r с рядом M–1 порогов, которые располагаются в средних точках между соседними уровнями амплитуд, как показано на рис. 5.2.7.

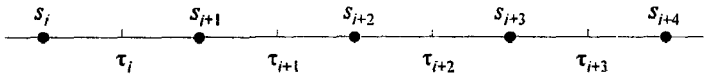


Рис. 5.2.7. Расположение порогов в средних точках между соседними уровнями амплитуд

Таким образом, решение выбирается в пользу уровня амплитуды, который расположен ближе всех к r.

Расположение порогов, указанных на рис. 5.2.7, помогает в вычислении вероятности ошибки. Заметим, что если передаётся m-й амплитудный уровень, то выход демодулятора равен

$$r = s_m + n = \sqrt{\tfrac{1}{2}\mathscr{E}_g}\,A_m + n, \qquad (5.2.41)$$

где компонента шума n имеет нулевое среднее и дисперсию $\sigma_n^2 = \frac{1}{2}N_0$. В предположении, что все уровни амплитуд априори равновероятны, средняя вероятность ошибочного приёма символа равна вероятности того, что компонента шума n превосходит по амплитуде половину расстояния между соседними уровнями. Однако, когда передаётся один из двух крайних уровней $\pm(M-1)$, ошибка возникает только в одном направлении. Таким образом, мы имеем

$$P_{\text{M}} = \frac{M-1}{M}P\left(|r - s_m| > d\sqrt{\tfrac{1}{2}\mathscr{E}_g}\right) = \frac{M-1}{M}\frac{2}{\sqrt{\pi N_0}}\int_{d\sqrt{\mathscr{E}_g/2}}^{\infty} \mathrm{e}^{-x^2/N_0}dx =$$

$$= \frac{M-1}{M}\frac{2}{\sqrt{2\pi}}\int_{\sqrt{d^2\mathscr{E}_g/N_0}}^{\infty} \mathrm{e}^{-x^2/2}dx = \frac{2(M-1)}{M}Q\left(\sqrt{\frac{d^2\mathscr{E}_g}{N_0}}\right). \qquad (5.2.42)$$

Вероятность ошибки (5.2.42) можно также выразить через среднюю переданную мощность. Из (5.2.40) видно, что

$$d^2\mathscr{E}_g = \frac{6}{M^2-1}P_{\text{ср}}T. \qquad (5.2.43)$$

Подставив $d^2\mathscr{E}_g$ в (5.2.42), получим среднюю вероятность ошибки на символ для АМ через среднюю мощность:

$$P_M = \frac{2(M-1)}{M}Q\left(\sqrt{\frac{6P_{\text{ср}}T}{(M^2-1)N_0}}\right) \qquad (5.2.44)$$

или, что эквивалентно,

$$P_M = \frac{2(M-1)}{M}Q\left(\sqrt{\frac{6\mathscr{E}_{\text{ср}}}{(M^2-1)N_0}}\right), \qquad (5.2.45)$$

где $\mathscr{E}_{ср} = P_{ср}T$ – средняя энергия. При построении зависимости вероятности ошибки на символ M-позиционной системы АМ обычно используется ОСШ на бит как базовый параметр. Так как $T = kT_b$ и $k = \log_2 M$, (5.2.45) можно преобразовать к

$$P_M = \frac{2(M-1)}{M} Q\left(\sqrt{\frac{(6\log_2 M)\mathscr{E}_{bср}}{(M^2-1)N_0}}\right), \qquad (5.2.46)$$

где $\mathscr{E}_{bср} = P_{ср}T_b$ – средняя энергия на бит, а $\mathscr{E}_{bср}/N_0$ – средняя ОСШ на бит.

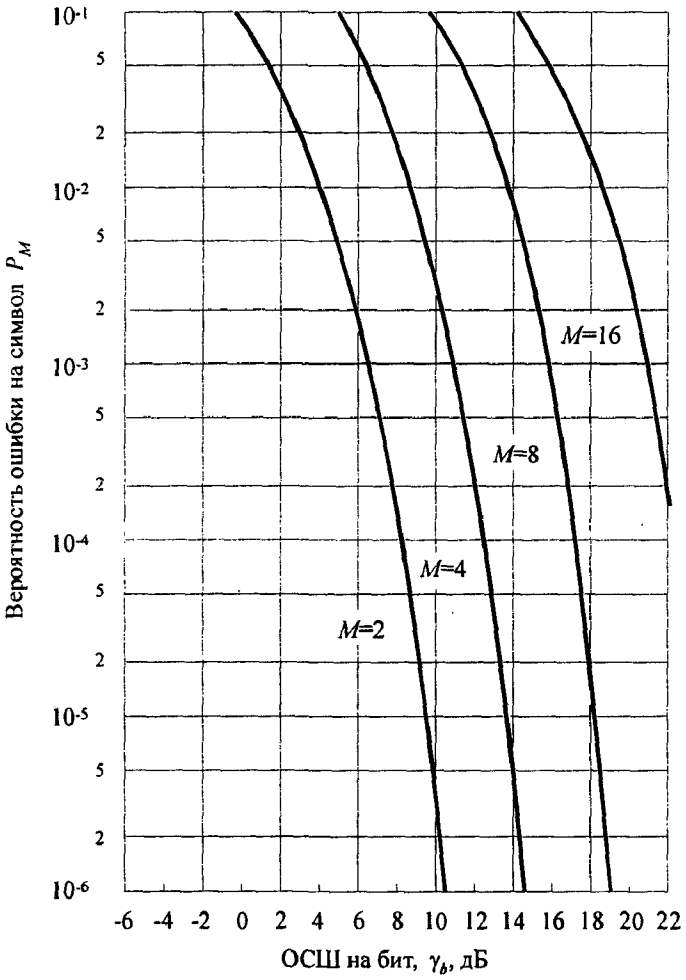


Рис. 5.2.8. Вероятность ошибки на символ для АМ

Рисунок 5.2.8 иллюстрирует зависимость вероятности ошибки на символ от $10\lg\left(\mathscr{E}_{bср}/N_0\right)$ со значением M в качестве параметра.

Заметим, что случай $M = 2$ соответствует вероятности ошибки для двоичной системы противоположных сигналов. Также видим, что при фиксированной вероятности ошибки P_M ОСШ на бит возрастает более чем на 4 дБ при каждом удвоении числа M. При очень больших M требуемый рост ОСШ при удвоении числа M приближается к 6 дБ.

5.2.7. Вероятность ошибки для M-позиционной ФМ

Напомним из разд. 4.3, что цифровой сигнал ФМ можно выразить так:

$$s_m(t) = g(t)\cos\left[2\pi f_c t + \frac{2\pi}{M}(m-1)\right], \quad 1 \le m \le M, \quad 0 \le t \le T, \qquad (5.2.47)$$

и он имеет векторное представление

$$\mathbf{s}_m = \left[\sqrt{\mathscr{E}_s}\cos\frac{2\pi}{M}(m-1) \quad \sqrt{\mathscr{E}_s}\sin\frac{2\pi}{M}(m-1)\right], \qquad (5.2.48)$$

где $\mathscr{E}_s = \frac{1}{2}\mathscr{E}_g$ – энергия каждого сигнала, а $g(t)$ – огибающая импульса передаваемого сигнала. Поскольку сигналы имеют одинаковую энергию, оптимальный детектор в канале с АБГШ, определяемый (5.1.44), вычисляет корреляционные метрики

$$C(\mathbf{r}, \mathbf{s}_m) = \mathbf{r} \cdot \mathbf{s}_m, \quad m = 1, 2, \ldots, M. \tag{5.2.49}$$

Другими словами, принимаемый сигнальный вектор $\mathbf{r} = [r1 \quad r2]$ проектируется на M возможных сигнальных векторов, и решение принимается в пользу сигнала с наибольшей проекцией.

Корреляционный детектор, описанный выше, эквивалентен фазовому детектору, который определяет фазу принимаемого сигнала $\mathbf{r}$ и выбирает сигнальный вектор $\mathbf{s}_m$, фаза которого ближе всего к фазе $\mathbf{r}$. Поскольку фаза $\mathbf{r}$ равна

$$\Theta_r = \operatorname{arctg} \frac{r_2}{r_1}, \tag{5.2.50}$$

мы хотим определить ФПВ Θ_r, по которой сможем вычислить вероятность ошибки.

Рассмотрим случай, когда фаза передаваемого сигнала $s_1(t)$ равна $\Theta_s = 0$. Следовательно, вектор переданного сигнала

$$\mathbf{s}_1 = \left[\sqrt{\mathscr{E}_s} \quad 0\right], \tag{5.2.51}$$

а вектор принимаемого сигнала имеет компоненты

$$\begin{aligned} r_1 &= \sqrt{\mathscr{E}_s} + n_1, \\ r_2 &= n_2. \end{aligned} \tag{5.2.52}$$

Поскольку n_1 и n_2 являются совместно гауссовскими случайными величинами[1] с нулевыми средними, следует, что r_1 и r_2 являются совместно гауссовскими случайными величинами с $E(r_1) = \sqrt{\mathscr{E}_s}$, $E(r_2) = 0$ и $\sigma_{r_1}^2 = \sigma_{r_2}^2 = \frac{1}{2} N_0 = \sigma_r^2$. Следовательно,

$$p_r(r_1, r_2) = \frac{1}{2\pi\sigma_r^2} \exp\left[-\frac{\left(r_1 - \sqrt{\mathscr{E}_s}\right)^2 + r_2^2}{2\sigma_r^2}\right]. \tag{5.2.53}$$

ФПФ фазы Θ_r можно получить заменой переменных (r_1, r_2) на

$$\begin{aligned} V &= \sqrt{r_1^2 + r_2^2}, \\ \Theta_r &= \operatorname{arctg}(r_2 / r_1). \end{aligned} \tag{5.2.54}$$

Это даёт совместную ФПВ

$$p_{V,\Theta_r}(V, \Theta_r) = \frac{V}{2\pi\sigma_r^2} \exp\left[-\frac{V^2 + \mathscr{E}_s - 2\sqrt{\mathscr{E}_s} V \cos\Theta_r}{2\sigma_r^2}\right].$$

Интегрирование $p_{V,\Theta_r}(V, \Theta_r)$ по области V даёт

$$p_{\Theta_r}(\Theta_r) = \int_0^\infty p_{V,\Theta_r}(V, \Theta_r) dV = \frac{1}{2\pi} e^{-2\gamma_s \sin^2\Theta_r} \int_0^\infty V e^{-(V - \sqrt{4\gamma_s}\cos\Theta_r)^2/2} dV, \tag{5.2.55}$$

где для удобства мы обозначили ОСШ символом $\gamma_s = \mathscr{E}_s / N_0$. Рисунок 5.2.9 иллюстрирует $p_{\Theta_r}(\Theta_r)$ для различных значений параметра ОСШ γ_s, когда фаза переданного сигнала равна нулю Заметим, что $p_{\Theta_r}(\Theta_r)$ становится уже и более концентрированной около фазы $\Theta_r = 0$ по мере увеличения параметра ОСШ γ_s.

Когда передаётся $s_1(t)$, ошибочное решение произойдёт, если шум вызовет нахождение фазы γ_s вне области $-\pi/M \le \Theta_r \le \pi/M$.

---

[1] Автор на протяжении всей книги обозначает случайные величины (процессы) то прописными (как обычно принято), то строчными буквами. Ввиду большого объёма материала мы не стали вносить соответствующих изменений (прп).

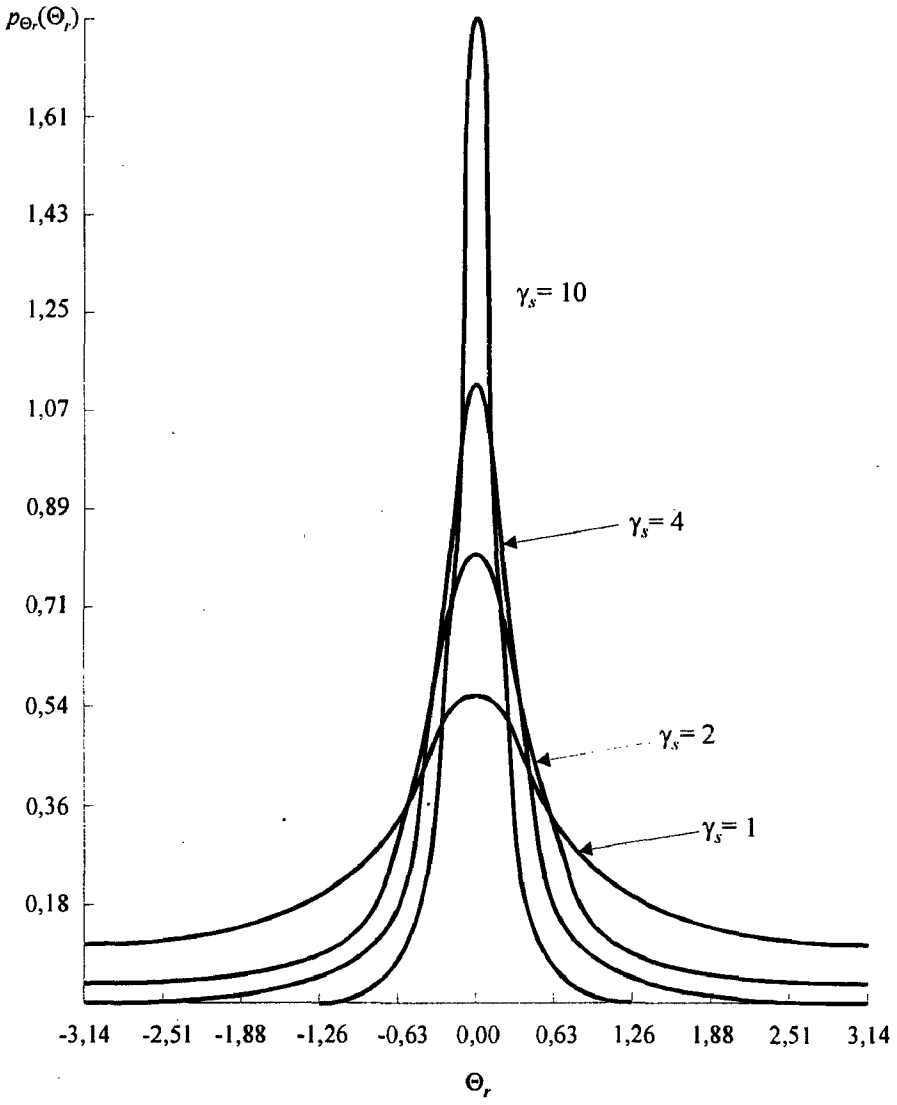


Рис. 5.2.9. Функция плотности вероятности $p_{\Theta_r}(\Theta_r)$ для $\gamma_s = 1, 2, 4$ и 10

Следовательно, вероятность ошибочного приёма символа

$$P_M = 1 - \int_{\pi/M}^{\pi/M} p_{\Theta_r}(\Theta_r) d\Theta_r . \qquad (5.2.56)$$

В общем, интегрирование $p_{\Theta_r}(\Theta_r)$ не приводится к простой форме и следует выполнить численное интегрирование, исключая случаи $M = 2$ и $M = 4$.

Для двоичной фазовой модуляции два сигнала $s_1(t)$ и $s_2(t)$ противоположны, и, следовательно, вероятность ошибки

$$P_2 = Q\left(\sqrt{\frac{2\mathscr{E}_b}{N_0}}\right). \qquad (5.2.57)$$

Когда $M = 4$, имеем случай двух двоичных фазово-модулированных сигналов в квадратуре. Поскольку здесь нет переходных помех или интерференции между сигналами на двух квадратурных несущих, вероятность ошибки на бит идентична той, которая определяется (5.2.57). С другой стороны, вероятность ошибки на символ при $M = 4$ определяется с учётом того, что

$$P_c = (1 - P_2)^2 = \left[1 - Q\left(\sqrt{\frac{2\mathscr{E}_b}{N_0}}\right)\right]^2, \qquad (5.2.58)$$

где P_c – вероятность правильного приёма для двух битовых символов. Результат (5.2.58) следует из статистической независимости шума на квадратурных несущих. Следовательно, вероятность ошибки на символ для $M = 4$ равна

$$P_4 = 1 - P_c = 2Q\left(\sqrt{\frac{2\mathscr{E}_b}{N_0}}\right)\left[1 - \frac{1}{2}Q\left(\sqrt{\frac{2\mathscr{E}_b}{N_0}}\right)\right]^2. \qquad (5.2.59)$$

Для $M > 4$ вероятность ошибки на символ P_M получена численным интегрированием (5.2.55). Рисунок 5.2.10 иллюстрирует эти вероятности ошибки как функции ОСШ на бит для $M = 2, 4, 8, 16$ и 32.

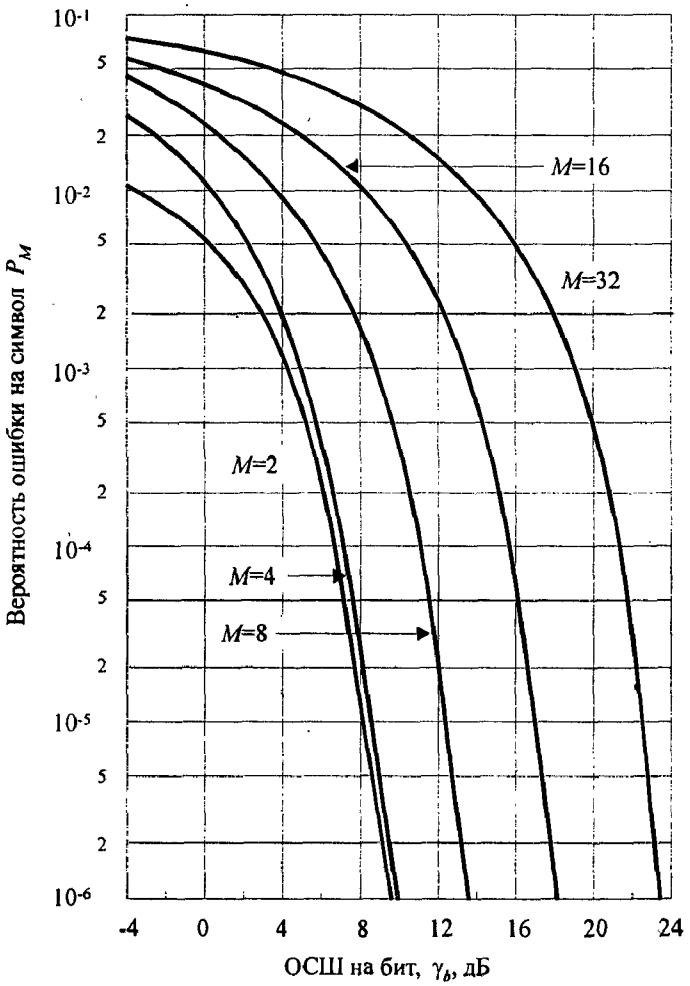


Рис. 5.2.10. Вероятность ошибки на символ для сигналов ФМ

Кривые явно иллюстрируют потери в ОСШ на бит по мере роста $M > 4$. Например, при $P_M = 10^{-5}$ разница в ОСШ между $M = 4$ и $M = 8$ приблизительно равна 4 дБ, а разница между $M = 8$ и $M = 16$ приблизительно равна 5 дБ. Для больших значений M рост числа фаз вдвое требует дополнительного увеличения ОСШ на 6 дБ/бит для достижения того же качества.

Аппроксимация вероятности ошибки для больших значений M и для больших ОСШ можно получить по первой аппроксимации $p_{\Theta_r}(\Theta_r)$. Для $\mathscr{E}_s/N_0 >> 1$ и $|\Theta_r| \le \frac{1}{2}\pi$ $p_{\Theta_r}(\Theta_r)$ хорошо аппроксимируется так:

$$p_{\Theta_r}(\Theta_r) \approx \sqrt{\frac{2\gamma_s}{\pi}} \cos\Theta_r e^{-2\gamma_s \sin^2\Theta_r}. \qquad (5.2.60)$$

Поставив (5.2.60) в (5.2.56) и выполнив замену переменной Θ_r на $u = \sqrt{2\gamma_s}\sin\Theta_r$, найдем

$$P_M \approx 1 - \int_{-\pi/M}^{\pi/M} \sqrt{\frac{2\gamma_s}{\pi}} \cos\Theta_r e^{-2\gamma_s \sin^2\Theta_r} d\Theta_r \approx \sqrt{\frac{2}{\pi}} \int_{\sqrt{2\gamma_s}\sin(\pi/M)}^{\infty} e^{-u^2/2} du =$$
$$= 2Q\left(\sqrt{2\gamma_s}\sin\frac{\pi}{M}\right) = 2Q\left(\sqrt{2k\gamma_b}\sin\frac{\pi}{M}\right), \qquad (5.2.61)$$

где $k = \log_2 M$ $\gamma_s = k\gamma_b$. Заметим, что эта аппроксимация вероятности ошибки хороша для всех значений M. Например, когда $M = 2$ и $M = 4$, мы имеем $P_2 = P_4 = 2Q\left(\sqrt{2\gamma_s}\right)$, что хорошо совпадает (за исключением множителя 2) с точным значением вероятности, данной (5.2.57).

Эквивалентную вероятность ошибки на бит для M-позиционной ФМ скорее утомительно вычислить с учетом её зависимости от отображения k-битового блока в соответствующее значение фазы сигнала. Если для такого отображения используется код Грея, два k-битовых блока, соответствующие сигналам с соседними значениями фаз, отличаются только на один бит. Поскольку более вероятные ошибки, обусловленные действием шума, приводят к выбору сигнала с соседним значением фазы вместо верного выбора, большинство k-битовых блоков содержат ошибки только в одном бите. Следовательно, эквивалентная вероятность ошибки на бит для M-позиционной ФМ хорошо аппроксимируется выражением

$$P_b = P_M/k. \qquad (5.2.62)$$

Наша трактовка демодуляции сигналов ФМ предполагает, что демодулятор располагает совершенной оценкой фазы несущей. На практике, однако, фаза несущей определяется по принятому сигналу путем использования некоторых нелинейных операций, которые приводят к неоднозначности фазы. Для примера в двоичной ФМ сигнал часто подвергается квадратированию, чтобы снять модуляцию, затем образованный сигнал с удвоенной частотой фильтруется и делится по частоте на 2 для того, чтобы получить оценку частоты несущей и фазы ϕ. Эти операции приводят к неоднозначности фазы несущей на 180°. Аналогично в четырехфазовой ФМ принимаемый сигнал возводится в четвертую степень, чтобы снять цифровую модуляцию, а затем четвёртая гармоника частоты несущей фильтруется и делится на 4 для того, чтобы выделить компоненту несущей. Эти операции приводят к компоненте частоты несущей, содержащей оценку фазы несущей ϕ, но возникают неоднозначности фазы на ±90° и на 180° при оценке фазы. Следовательно, мы не имеем точную оценку фазы несущей в демодуляторе.

Проблема неоднозначности фазы, возникающей при оценке фазы несущей ϕ, может быть преодолена путём использования дифференциальной ФМ (ДФМ) вместо абсолютной ФМ. При дифференциальной ФМ кодирование информации осуществляется посредством разности фаз между соседними переданными сигналами[1], а не самой абсолютной фазы, как при обычной ФМ. Например, в двоичной ДФМ информационный символ 1 передаётся со

[1] Метод однократной фазоразностной (или относительной фазовой) модуляции был впервые предложен в 1959 г. Н.Т. Петровичем [ ], а многократной фазоразностной модуляцин в 1972 г. Ю.Б. Окуневым [ ].

сдвигом фазы несущей на 180^0 относительно предыдущего значения фазы несущей, в то время как информационный символ 0 передаётся без сдвига фазы. В четырёхфазной ДФМ относительный сдвиг фаз между соседними сигнальными интервалами равен 0, $90^\circ, 180^\circ$, и -90° в зависимости от информационных символов 00, 01, 11 и 10 соответственно. Обобщение на случай $M > 4$ очевидно. Сигналы ФМ, получаемые при таком процессе кодирования, называют *дифференциально-кодированными*. Такое кодирование выполняется относительно простой логической схемой, предшествующей модулятору.

Демодуляция сигнала при дифференциальном кодировании ФМ может выполняться, как описано выше, с игнорированием неоднозначности фазы. Так, принимаемый сигнал демодулируется и детектируется на каждом сигнальном интервале в одно из M возможных значений фазы. За детектором имеется относительно простое устройство сравнения фаз, которое сравнивает фазы демодулированных сигналов на двух соседних сигнальных интервалах с тем, чтобы извлечь информацию.

Когерентная демодуляция для ФМ с дифференциальным кодированием приводит к большей вероятности ошибки, чем вероятность ошибки, достигаемая при абсолютном фазовом кодировании. При ФМ с дифференциальным кодированием ошибка при демодуляции фазы сигнала на данном интервале будет обычно возникать при ошибочном декодировании на любом из двух соседних сигнальных интервалов. Это особенно характерно для ошибок с вероятностью ниже 0,1. Следовательно, вероятность ошибки M-позиционной ФМ при дифференциальном кодировании приблизительно вдвое больше вероятности ошибки для M-позиционной ФМ с абсолютным кодированием фазы. Однако увеличение вероятности ошибки вдвое ведёт к относительно малым потерям в ОСШ.

5.2.8. Дифференциальная ФМ (ДФМ) и её характеристики качества

Сигнал дифференциально-кодированный фазовой модуляции (ДФМ) позволяет использовать вид демодуляции, который не требует оценки фазы несущей[1]. Вместо этого принимаемый сигнал на заданном сигнальном интервале сравнивается по фазе с принятым сигналом на предыдущем сигнальном интервале. Для детальной разработки предположим, что мы демодулируем сигнал ДФМ путём умножения $r(t)$ на $\cos 2\pi f_c t$ и $\sin 2\pi f_c t$ и интегрируем произведения на интервале T. На k-м сигнальном интервале выход демодулятора равен

$$\mathbf{r}_k = \left[\sqrt{\mathscr{E}_s}\cos(\theta_k - \phi) + n_{k1} \quad \sqrt{\mathscr{E}_s}\sin(\theta_k - \phi) + n_{k2}\right]$$

или, что эквивалентно,

$$r_k = \sqrt{\mathscr{E}_s}e^{j(\theta_k - \phi)} + n_k, \tag{5.2.63}$$

где r_k – комплексное число, θ_k – фазовый угол переданного сигнала на k-м тактовом интервале, ϕ – фаза несущей и $n_k = n_{k1} + jn_{k2}$ – вектор шума. Аналогично принятый сигнальный вектор на выходе демодулятора на предыдущем сигнальном интервале равен

$$r_{k-1} = \sqrt{\mathscr{E}_s}e^{j(\theta_{k-1} - \phi)} + n_{k-1}. \tag{5.2.64}$$

В фазовом детекторе решение принимается по разности фаз между двумя комплексными величинами. Эквивалентно мы можем проектировать r_k на r_{k-1} и использовать фазу результирующего комплексного числа, т.е. получим

---

[1] Поскольку не требуется оценки фазы, ДФМ часто рассматривается как техника некогерентной связи. Мы придерживаемся точки зрения, что ДФМ представляет вид цифровой модуляции фазы в экстремальном случае, когда требуется оценка фазы только от предыдущего символьного интервала.

$$r_k r_{k-1}^* = \mathscr{E}_s e^{j(\theta_k - \theta_{k-1})} + \sqrt{\mathscr{E}_s} e^{j(\theta_k - \phi)} n_{k-1}^* + \sqrt{\mathscr{E}_s} e^{-j(\theta_{k-1} - \phi)} n_k + n_k n_{k-1}^*, \quad (5.2.65)$$

откуда при отсутствии шума может быть точно определена разность фаз $\theta_k - \theta_{k-1}$. Таким образом, значение $r_k r_{k-1}^*$ не зависит от фазы несущей. Дифференциально кодированные сигналы ФМ, которые демодулируются и декодируется описанным выше способом, называют обычно *дифференциальной* ФМ (ДФМ).

Демодуляция и детектирование ДФМ с использованием согласованных фильтров иллюстрируются рис. 5.2.11. Если импульс $g(t)$ прямоугольный, согласованные фильтры можно заменить интеграторами со сбросом.

Теперь рассмотрим эволюцию характеристики качества (вероятности ошибки) демодулятора и детектора ДФМ. Точный расчёт величины вероятности ошибки для M-позиционной ДФМ очень сложен, исключая случай M=2. Наибольшая трудность связана с определением ФПВ для фазы случайной величины $r_k r_{k-1}^*$, определяемой (5.2.65). Однако, как мы теперь покажем, легко получить аппроксимацию характеристики качества ДФМ.

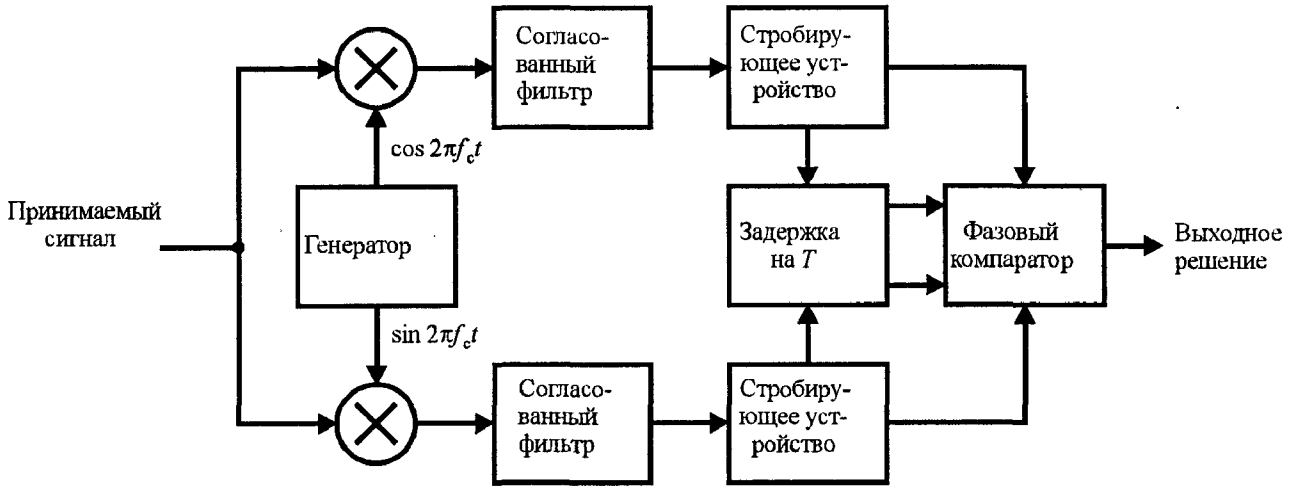


Рис. 5.2.11. Блок-схема демодулятора ДФМ

Без потери общности предположим, что разность фаз $\theta_k - \theta_{k-1} = 0$. Далее, экспоненциальные множители $e^{-j(\theta_{k-1} - \phi)}$ и $e^{j(\theta_k - \phi)}$ в (5.2.65) можно включить в гауссовские шумовые компоненты n_{k-1} и n_k без изменения их статистических свойств. Следовательно, $r_k r_{k-1}^*$ в (5.2.65) можно выразить так:

$$r_k r_{k-1}^* = \mathscr{E}_s + \sqrt{\mathscr{E}_s}\left(n_k + n_{k-1}^*\right) + n_k n_{k-1}^*. \quad (5.2.66)$$

Сложность определения ФПВ фазы определяется слагаемым $n_k n_{k-1}^*$. Однако при больших ОСШ, представляющих практический интерес, слагаемое $n_k n_{k-1}^*$ мало по сравнению с доминирующей компонентой шума $\sqrt{\mathscr{E}_s}\left(n_k + n_{k-1}^*\right)$ Если мы пренебрежём слагаемым $n_k n_{k-1}^*$ и нормируем $r_k r_{k-1}^*$ делением на $\sqrt{\mathscr{E}_s}$, то получим новый ряд метрик, по которым выносится решение:

$$\begin{aligned} x &= \sqrt{\mathscr{E}_s} + \mathrm{Re}\left(n_k + n_{k-1}^*\right), \\ y &= \mathrm{Im}\left(n_k + n_{k-1}^*\right). \end{aligned} \quad (5.2.67)$$

Здесь x и y являются некоррелированными гауссовскими случайными величинами с одинаковой дисперсией $\sigma_n^2 = N_0$. Фаза равна

$$\Theta_r = \operatorname{arctg}(y/x). \qquad (5.2.68)$$

На этой стадии мы имеем проблему, которая идентична той, которую мы решали ранее для фазово-когерентной демодуляции. Единственная разница в том, что дисперсия шума теперь в два раза больше, чем в случае ФМ. На этом основании заключаем, что характеристика качества ДФМ на 3 дБ хуже, чем для ФМ. Этот результат относительно хорош для $M \geq 4$, но он пессимистичен для $M = 2$ в том смысле, что действительная потеря ДФМ относительно ФМ менее 3 дБ при больших ОСШ. Это мы покажем ниже.

В двоичной ДФМ два возможных значения фазы передаваемого сигнала равны 0 и π. Как следствие, только реальная часть $r_k r_{k-1}^*$ необходима для извлечения информации. Используя (5.2.67), выразим реальную часть так:

$$\operatorname{Re}\left(r_k r_{k-1}^*\right) = \tfrac{1}{2}\left(r_k r_{k-1}^* + r_k^* r_{k-1}\right).$$

Поскольку мы предполагали, что разность фаз между сигналами на соседних интервалах равна 0, ошибка возникает, если $\operatorname{Re}\left(r_k r_{k-1}^*\right) < 0$. Вероятность того, что $r_k r_{k-1}^* + r_k^* r_{k-1} < 0$, это специальный случай исследования, данного в приложении В, где обсуждается вероятность того, что общая квадратичная форма комплексных случайных гауссовских величин меньше нуля. Общая форма для этой вероятности даётся (В.21) в приложении В, и она зависит всецело от первого и второго моментов комплексных гауссовских случайных величин r_k и r_{k-1}. Вычислив моменты и параметры, которые являются функциями моментов, получим вероятность ошибки двоичной ДФМ в виде[1]

$$P_b = \tfrac{1}{2} e^{-\mathcal{E}_b/N_0}, \qquad (5.2.69)$$

где $\gamma_b = \mathcal{E}_b/N_0$ – это ОСШ на бит.

Графики $P_b(\gamma_b)$ показаны на рис. 5.2.12. На этом графике показана также вероятность ошибки двоичной когерентной ФМ.

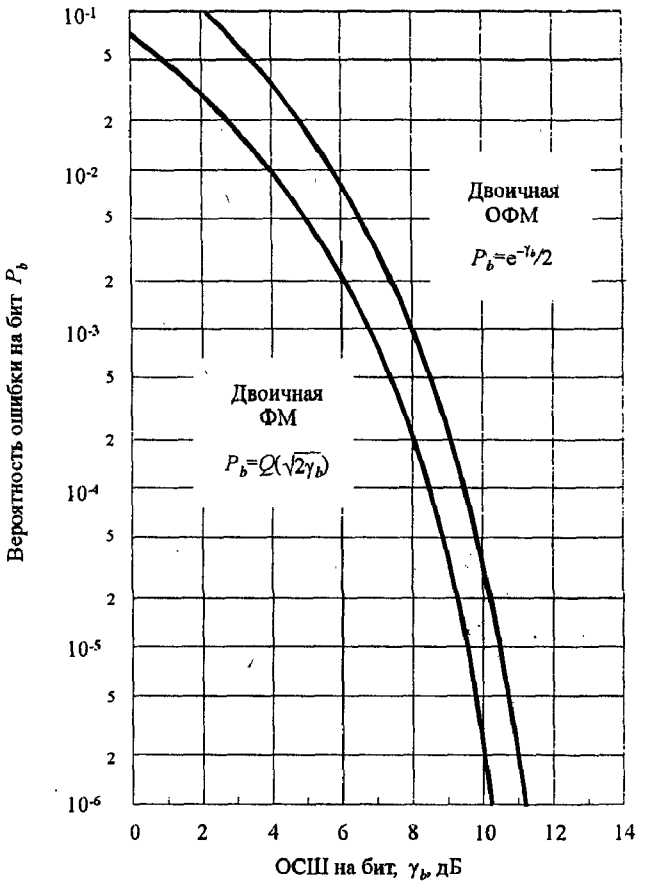


Рис. 5.2.12. Вероятность ошибки для двоичной ФМ и ДФМ

Видно, что при вероятности ошибки $P_b \leq 10^{-3}$ разница в ОСШ между ФМ и ДФМ менее 3 дБ. При $P_b \leq 10^{-5}$ разница в ОСШ меньше 1 дБ.

---

[1] Этот результат можно получить значительно проще (см. Л.М. Финк [ ]) из формулы для вероятности ошибки двоичной системы сигналов, ортогоиальных в усиленном смысле, если учесть, что сигналы двоичной ДФМ на интервале $2T$ ортогональны в усилеином смысле (прп).

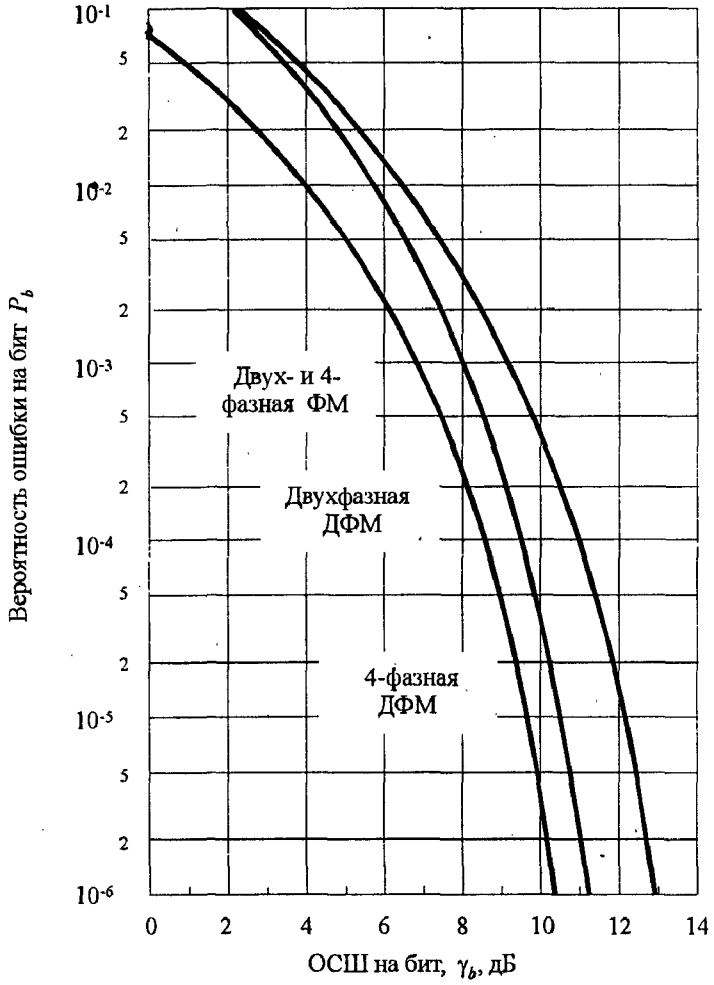


Рис. 5.2.13. Вероятность ошибки на бит для двоичной и четырёхфазной ФМ и ДФМ

Вероятность ошибки на бит для четырёхфазной ДФМ с кодом Грея можно выразить через известные функции.

Мы просто сформулируем здесь результат, а читателю, интересующемуся деталями, рекомендуем приложение С. Результат выражается в виде

$$P_b = Q_1(a,b) - \tfrac{1}{2} I_0(a \cdot b) \exp\left[-\tfrac{1}{2}(a^2 + b^2)\right],$$

где $Q_1(ab)$ – это Q-функция Маркума, определенная (2.1.122) и (2.1.123), $I_0(x)$ – модифицированная функция Бесселя нулевого порядка, определённая (2.1.120), а параметры a и b определяются так:

$$a = \sqrt{2\gamma_b\left(1 - \sqrt{1/2}\right)}, \qquad b = \sqrt{2\gamma_b\left(1 + \sqrt{1/2}\right)}. \tag{5.2.71}$$

Рисунок 5.2.13 иллюстрирует зависимость вероятности ошибки на бит для сигналов двух- и четырёхфазной ДФМ и когерентной ФМ, полученную расчётом по точным формулам этого раздела.

Поскольку двоичная ДФМ мало уступает двоичной ФМ при больших ОСШ и не требует разработки специального метода оценки фазы несущей, она часто используется в цифровых системах. С другой стороны, четырёхфазная ДФМ приблизительно на 2,3 дБ хуже по качеству, чем четырёхфазная ФМ при больших значениях ОСШ. Следовательно, выбор между этими двумя четырёхфазными системами неоднозначен. Надо взвесить потери в 2,3 дБ и упрощения в реализации устройства.

5.2.9. Вероятность ошибки для КАМ

Напомним из раздела 4.3, что сигнал КАМ можно выразить так

$$s_m(t) = A_{mc} g(t) \cos 2\pi f_c t - A_{ms} g(t) \sin 2\pi f_c t, \tag{5.2.72}$$

где A_{mc} и A_{ms} – содержащие информацию амплитуды квадратурных несущих, а $g(t)$ – сигнальный импульс. Векторное представление этих сигналов

$$\mathbf{s}_m = \left[A_{mc}\sqrt{\tfrac{1}{2}\mathscr{E}_g} \quad A_{ms}\sqrt{\tfrac{1}{2}\mathscr{E}_g}\right]. \tag{5.2.73}$$

Чтобы определить вероятность ошибки при КАМ, мы должны конкретизировать точки сигнального созвездия. Начнём с сигнального ансамбля КАМ, который имеет $M = 4$ точки. Рис. 5.2.14 иллюстрирует два таких ансамбля. Первый (a) – это четырёхфазный модулированный сигнал, а второй (b) – это четырёхфазный сигнал КАМ с двумя уровнями амплитуд, обозначенными A_1 и A_2, и четырьмя значениями фаз. Поскольку вероятность ошибки определяется минимальным расстоянием между парой сигнальных точек, примем

условие, что $d_{\min}^{(e)} = 2A$ для обоих сигнальных созвездий, и рассчитаем среднюю переданную мощность, основываясь на предположении, что все сигнальные точки равновероятны. Для четырёхфазного сигнала имеем

$$P_{ср} = \tfrac{1}{4}(4)2A^2 = 2A^2. \qquad (5.2.74)$$

Для двухамплитудной четырёхфазной КАМ мы разместим точки на окружностях радиуса A и $\sqrt{3}A$. Поскольку $d_{\min}^{(e)} = 2A$, имеем

$$P_{ср} = \tfrac{1}{4}\left[2\left(3A^2\right) + 2A^2\right] = 2A^2, \qquad (5.2.75)$$

что совпадает со средней мощностью для четырёхфазного сигнального созвездия. Следовательно, для всех практических применений вероятность ошибки двух ансамблей сигналов одинакова. Другими словами, нет преимущества двухамплитудного сигнала КАМ относительно четырёхфазной модуляции.

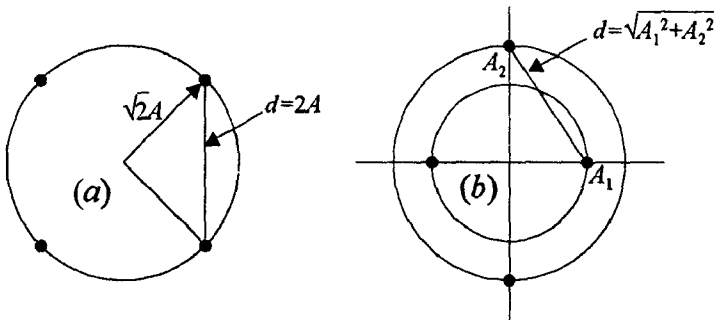


Рис. 5.2.14. Два 4-точечных сигнальных созвездия

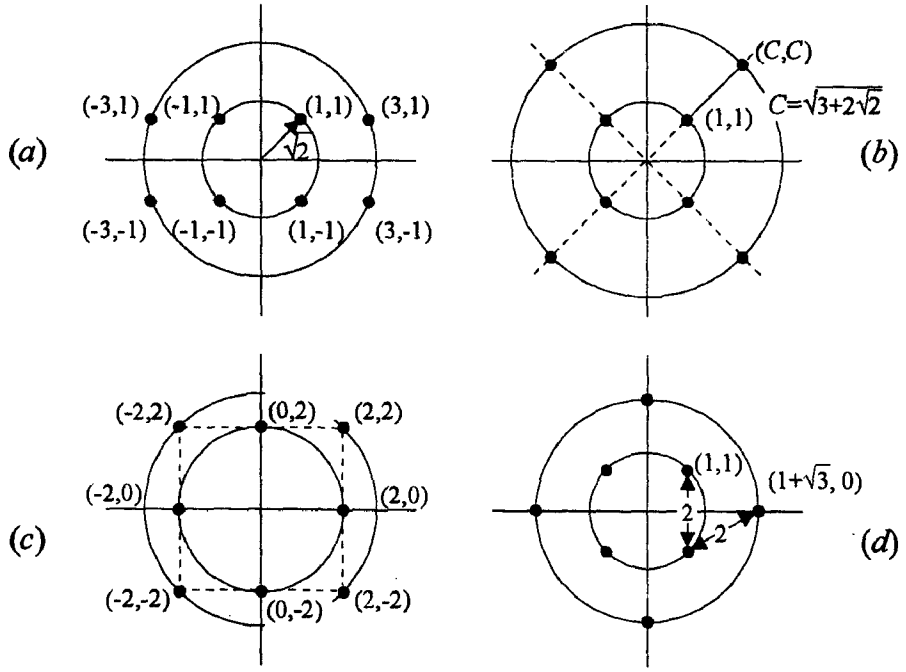


Рис. 5.2.15. Четыре 8-точечных созвездия сигналов КАМ

Далее рассмотрим восьмиуровневый ($M = 8$) сигнал КАМ. В этом случае имеются много возможных сигнальных созвездий. Рассмотрим четыре сигнальных созвездия, показанных на рис. 5.2.15. Все они характеризуются двумя амплитудами и имеют минимальные расстояния между сигнальными точками $2A$. Координаты $\left(A_{mc}, A_{ms}\right)$ для

каждой сигнальной точки, нормированные по A, даны на рисунке. Предполагая, что все сигнальные точки равновероятны, получаем для средней переданной мощности сигнала

$$P_{\text{ср}} = \frac{1}{M}\sum_{m=1}^{M}\left(A_{mc}^2 + A_{ms}^2\right) = \frac{A^2}{M}\sum_{m=1}^{M}\left(a_{mc}^2 + a_{ms}^2\right), \qquad (5.2.76)$$

где $\left(a_{mc}, a_{ms}\right)$ – координаты сигнальных точек, нормированные по A. Два сигнальных ансамбля (a) и (c) на рис. 5.2.15 содержат сигнальные точки, которые лежат на сетке прямоугольника и имеют $P_{\text{ср}} = 6A^2$. Сигнальный ансамбль (b) требует переданную среднюю мощность $P_{\text{ср}} = 6{,}8A^2$, а ансамбль (d) требует $P_{\text{ср}} = 4{,}73A^2$. Следовательно, четвёртый сигнальный ансамбль требует примерно на 1 дБ меньше мощности, чем первые два, и на 1,6 дБ меньше мощности, чем третий, для того, чтобы достичь той же вероятности ошибки. Это сигнальное созвездие известно как лучшее восьмиточечное КАМ созвездие, так как оно требует наименьшей мощности при заданном минимальном расстоянии между сигнальными точками.

Для $M \geq 16$ имеется намного больше возможностей для выбора сигнальных точек КАМ в двухмерном пространстве. Для примера мы можем выбрать круговые многоуровневые созвездия для $M = 16$, как показано на рис. 4.3.4. В этом случае сигнальные точки при заданной амплитуде поворачиваются по фазе на $\frac{1}{4}\pi$ относительно сигнальных точек соседних уровней амплитуд. Это созвездие 16 КАМ является обобщением оптимального созвездия 8 КАМ. Однако круговое созвездие 16 КАМ не является наилучшим 16-точечным созвездием КАМ в канале с АБГШ.

Прямоугольное сигнальное созвездие КАМ имеет отчётливое преимущество с точки зрения простоты генерирования, как два сигнала АМ, переданные на квадратурных по фазе несущих. Кроме того, оно легко демодулируется. Хотя оно не являются наилучшим M-позиционным сигнальным созвездием при КАМ для $M \geq 16$, средняя переданная мощность, требуемая для достижения заданного минимального расстояния, лишь ненамного больше, чем средняя мощность, требуемая при наилучшем сигнальном созвездии КАМ. Исходя из этих соображений, прямоугольное M-позиционное сигнальное созвездие КАМ наиболее часто используется на практике.

Для прямоугольных сигнальных созвездий при $M = 2^k$, где k – чётно, сигнальное созвездие КАМ эквивалентно сумме двух сигналов АМ на квадратурных несущих, причём каждый имеет $\sqrt{M} = 2^{k/2}$ сигнальных точек. Поскольку сигналы в квадратурных компонентах можно точно разделить в демодуляторе, вероятность ошибки для КАМ легко определить по вероятности ошибки АМ[1]. Конкретнее вероятность правильного решения для M-позиционной системы КАМ равна

$$P_{\text{c}} = \left(1 - P_{\sqrt{M}}\right)^2, \qquad (5.2.77)$$

где $P_{\sqrt{M}}$ – вероятность ошибки для $\sqrt{M}$-позиционной АМ с половинной средней мощностью в каждом квадратурном сигнале эквивалента КАМ. Несколько модифицируя выражение для вероятности ошибки в M-позиционной АМ, получаем

$$P_{\sqrt{M}} = 2\left(1 - \frac{1}{\sqrt{M}}\right) Q\left(\sqrt{\frac{3}{M-1}\frac{\mathscr{E}_{\text{ср}}}{N_0}}\right), \qquad (5.2.78)$$

[1] Независимая обработка квадратурных компонеит возможна только в неискажающем (однопутевом) канале (прп).

где $\mathscr{E}_{ср}/N_0$ – среднее ОСШ на символ. Следовательно, вероятность ошибки на символ для M-позиционной КАМ равна

$$P_M = 1-\left(1-P_{\sqrt{M}}\right)^2. \qquad (5.2.79)$$

Подчеркнём, что этот результат точен при $M=2^k$, когда k чётно. С другой стороны, если k нечётно, нет эквивалентной $\sqrt{M}$-позиционной системы АМ. Однако здесь нет проблемы, поскольку всегда легче определить вероятность ошибки для прямоугольного ансамбля сигналов. Если мы используем оптимальный детектор, который основывает свои решения на использовании дистанционных метрик, определяемых (5.1.49), относительно просто показать, что вероятность ошибки на символ имеет плотную верхнюю границу

$$P_M \le 1-\left[1-2Q\left(\sqrt{\frac{3\mathscr{E}_{ср}}{(M-1)N_0}}\right)\right]^2 \le 4Q\left(\sqrt{\frac{3k\mathscr{E}_{bср}}{(M-1)N_0}}\right) \qquad (5.2.80)$$

для всех $k \ge 1$, где $\dfrac{\mathscr{E}_{bср}}{N_0}$ – среднее ОСШ на бит. Кривые вероятности ошибки на символ даны на рис 5.2.16, как функции от среднего ОСШ на бит.

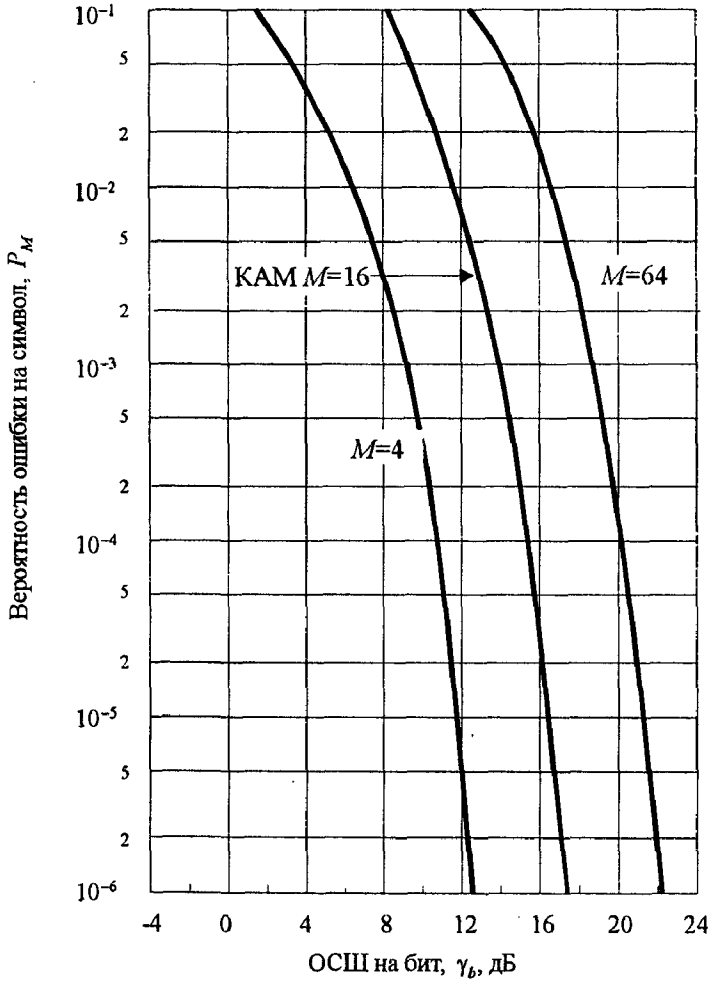


Рис. 5.2.16. Вероятность ошибки на символ для КАМ

Для непрямоугольных сигнальных созвездий КАМ можем получить верхнюю границу для вероятности ошибки, используя объединённую границу. Очевидная верхняя граница

$$P_M \le (M-1)Q\left(\sqrt{\left(d_{\min}^{(e)}\right)^2/2N_0}\right),$$

где $d_{\min}^{(e)}$ – минимальное евклидово расстояние между сигнальными точками. Эта граница может быть неплотной, когда M велико. В этом случае можем аппроксимировать P_M, заменяя $M-1$ на M_n, где M_n – наибольшее число ближайших точек, которые имеют расстояние $d_{\min}^{(e)}$ от любой точки созвездия.

Интересно сравнить характеристику качества КАМ и АМ для заданного объёма сигналов M, поскольку оба типа сигналов являются двухмерными. Напомним, что для M-позиционной ФМ вероятность ошибки на символ аппроксимируется так:

$$P_M \approx 2Q\left(\sqrt{2\gamma_s}\sin\frac{\pi}{M}\right), \tag{5.2.81}$$

где γ_s – ОСШ на символ. Для M-позиционной КАМ мы можем использовать выражение (5.2.78). Поскольку вероятность ошибки определяется аргументом Q-функция, можем сравнить аргументы Q для двух сигнальных форматов. Отношение двух обсуждаемых аргументов равно

$$\mathcal{R}_M = \frac{3/(M-1)}{2\sin^2(\pi/M)}. \tag{5.2.82}$$

Например, когда $M=4$, имеем $\mathcal{R}_M = 1$. Следовательно, 4-позиционная ФМ и 4-позиционная КАМ дают сходные характеристики качества для одинаковых ОСШ на символ. С другой стороны, когда $M>4$, находим, что $\mathcal{R}_M > 1$, так что M-позиционная КАМ даёт лучшую характеристику качества, чем M-позиционная ФМ. Таблица 5.2.1 иллюстрирует выигрыш в ОСШ системы КАМ относительно ФМ для некоторых значений M. Например, видно, что система 32 КАМ имеет выигрыш по ОСШ на 7 дБ относительно системы 32 ФМ.

Таблица 5.2.1. Выигрыш в ОСШ M-позиционной КАМ по отношению к M-позиционной ФМ

M	$10\lg\mathcal{R}_M$, дБ
8	1,65
16	4,20
32	7,02
64	9,95

5.2.10. Сравнение цифровых методов модуляции

Методы цифровой модуляции, описанные в этой главе, можно сравнить различными путями. Например, можно их сравнить на основе ОСШ, требуемого для достижения заданной вероятности ошибки. Однако такое сравнение не будет достаточно осмысленным, если не будут выполняться определённые требования, такие как фиксированная скорость передачи данных или, что эквивалентно, фиксированная полоса частот. Имея это в виду, рассмотрим требования по полосе при различных методах модуляции.

Для многофазных сигналов требуемая полоса частот – это просто полоса эквивалентно низкочастотного сигнального импульса $g(t)$, которая зависит от его подробных характеристик. Для наших целей предположим, что $g(t)$ – это импульс длительности T, а

его полоса частот W приближённо равна обратной величине T. Таким образом, $W = 1/T$, и поскольку $T = k/R = (\log_2 M)/R$, то следует

$$W = \frac{R}{\log_2 M}. \qquad (5.2.83)$$

Таким образом, по мере роста M требуемая полоса частот уменьшается при фиксированной битовой скорости R. Частотная эффективность измеряется отношением битовой скорости к полосе и равна

$$\frac{R}{W} = \log_2 M. \qquad (5.2.84)$$

Частотно-эффективный метод передачи сигналов АМ – это однополосная передача. Тогда полоса, требуемая для передачи сигнала, приблизительно равна $1/2T$, и, поскольку $T = k/R = (\log_2 M)/R$, следует, что

$$\frac{R}{W} = 2\log_2 M. \qquad (5.2.85)$$

Это в два раза лучше, чем при ФМ.

В случае КАМ имеем две ортогональные несущие, причём на каждой несущей передаётся АМ сигнал. Таким способом удваиваем скорость относительно АМ. Однако сигнал КАМ должен быть передан двумя полосами. Следовательно, КАМ и АМ имеют одинаковую частотную эффективность, если полоса относится к полосовому сигналу.

Ортогональные сигналы имеют в целом другие требования по полосе. Если $M = 2^k$ ортогональных сигнала синтезированы посредством ортогональных несущих с минимальным разносом частот $1/2T$ для ортогональности[1], то полоса частот, требуемая для передачи $k = \log_2 M$ информационных бит, равна

$$W = \frac{M}{2T} = \frac{M}{2(k/R)} = \frac{M}{2\log_2 M} R. \qquad (5.2.86)$$

В этом случае полоса увеличивается по мере увеличения M. Аналогичные соотношения можно получить для симплексных и биортогональных сигналов. В случае биортогональных сигналов требуется полоса частот, равная половине полосы для такого же количества ортогональных сигналов.

Компактное и осмысленное сравнение этих методов модуляции базируется на зависимости нормированной скорости передачи данных R/W (бит в секунду на 1 Гц полосы частот) от ОСШ на бит ($\mathscr{E}_b/N_0$), требуемого для достижения заданной вероятности ошибки. Рисунок 5.2.17 иллюстрирует зависимость R/W от ОСШ на бит для АМ, КАМ, ФМ и ортогональных сигналов для случая, когда вероятность ошибки на символ $P_M = 10^{-5}$. Видим, что в случае АМ, КАМ и ФМ рост M ведёт к росту R/W. Однако плата за достижение большей скорости передачи данных увеличивается по мере роста ОСШ на бит. Следовательно, эти методы модуляции предпочтительны для частотно-ограниченных каналов связи, когда желательно иметь $R/W > 1$ и где обеспечивается достаточно высокое ОСШ, чтобы поддержать рост M. Телефонные каналы и цифровые микроволновые радиоканалы являются примерами таких частотно-ограниченных каналов.

В противоположность этому M-позиционные ортогональные сигналы дают отношение битовой скорости к полосе канала $R/W \le 1$. При росте M величина R/W падает из-за роста

[1] Такой разнос частот является минимально необходимым для разделения сигналов при когерентном приёме (прп).

требуемой полосы частот канала. Однако ОСШ на бит, требуемое для достижения заданной вероятности ошибки (в нашем случае $P_M=10^{-5}$), уменьшается с ростом M. Следовательно, M-позиционные ортогональные сигналы предпочтительны для каналов с ограничением по мощности, которые имеют достаточно широкую полосу для размещения большого числа сигналов. В этом случае, когда $M \to \infty$, вероятность ошибки можно сделать сколь угодно малой, если обеспечено $\mathcal{E}_b/N_0 > 0{,}693$ (−1,6 дБ). Это минимальные ОСШ на бит, требуемые для достижения реализуемой передачи в пределе, когда $W \to \infty$ и соответствующее отношение битовой скорости к полосе частот $R/W \to 0$.

На рис. 5.2.17 также дан график нормированной пропускной способности частотно-ограниченного канала с АБГШ, полученный Шенноном (1948). Отношение C/W, где $C = R$ – это пропускная способность в битах/с, представляет наибольшее достижимое отношение битовой скорости к полосе в данном канале. Следовательно, оно служит верхней границей частотной эффективности для любого вида модуляции. Эта граница определяется в гл. 7 и там же детально обсуждается[1].

5.3. ОПТИМАЛЬНЫЙ ПРИЁМНИК ДЛЯ СИГНАЛОВ МНФ

Напомним, из разд. 4.3, что МНФ – это метод модуляции с памятью. Память обусловлена непрерывностью фазы передаваемой несущей от одного сигнального интервала к другому. Передаваемый МНФ сигнал можно записать в виде

$$s(t) = \sqrt{\frac{2E}{T}} \cos\left[2\pi f_c t + \phi(t;\mathbf{I})\right], \tag{5.3.1}$$

где $\phi(t;\mathbf{I})$ – фаза несущей. Отфильтрованный принимаемый сигнал $r(t)$ и шум $n(t)$ в канале представим так:

$$r(t) = s(t) + n(t), \tag{5.3.2}$$

$$n(t) = n_c(t)\cos 2\pi f_c t - n_s(t)\sin 2\pi f_c t\,. \tag{5.3.3}$$

5.3.1. Оптимальные демодуляция и детектирование для МНФ

Оптимальный приёмник для этих сигналов состоит из каскадного соединения коррелятора и детектора последовательности максимального правдоподобия, который ищет путь по решётке состояний с минимальным евклидовым расстоянием от принятого сигнала. Алгоритм Витерби позволяет эффективно осуществить этот поиск. Установим общую структуру решётки состояний для МНФ и затем опишем расчёт метрик.

Напомним, что фазу несущей для сигнала МНФ с фиксированным индексом модуляции h можно выразить так:

$$\begin{aligned}\phi(t;\mathbf{I}) &= 2\pi h \sum_{k=-\infty}^{n} I_k q(t-kT) = \\ &= \pi h \sum_{k=-\infty}^{n-L} I_k + 2\pi h \sum_{k=n-L+1}^{n} I_k q(t-kT) = \theta_n + \theta(t;\mathbf{I}), \quad nT \le t \le (n+1)T,\end{aligned} \tag{5.3.4}$$

где мы предположили, что $q(t) = 0$ для $t < 0$, $q(t) = \frac{1}{2}$ для $t \ge LT$ и

$$q(t) = \int_0^t g(\tau)d\tau\,. \tag{5.3.5}$$

Сигнальный импульс $g(t) = 0$ для $t < 0$ и $t \ge LT$. Для $L = 1$ имеем МНФ с полным откликом, а при $L > 1$, где L – положительное целое число, имеем МНФ с парциальным откликом.

---

[1] Широкие исследования зависимости между энергетической и частотной эффективностью различных систем связи выполнены А.Г. Зюко [48] (прп).

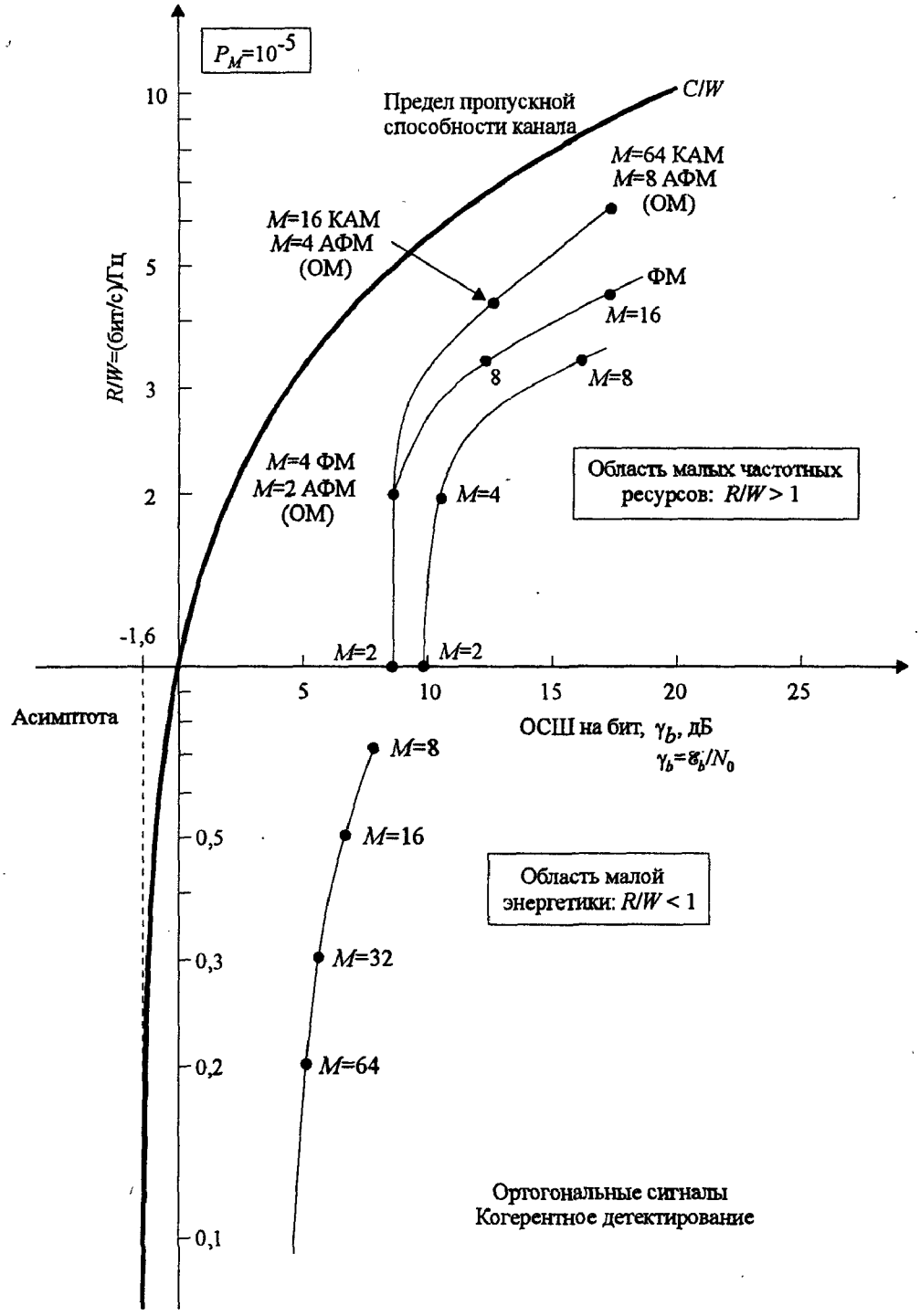


Рис. 5.2.17. Сравнение различных методов модуляции при вероятности ошибки на символ 10^{-5}

Теперь если h – рациональное число, т.е. $h=m/p$, где m и p – это взаимно простые положительные целые числа, то схему образования МНФ можно представить решёткой. В этом случае имеются p состояний фазы

$$\Theta_s=\left\{0,\frac{\pi m}{p},\frac{2\pi m}{p},\ldots,\frac{(p-1)\pi m}{p}\right\}, \qquad (5.3.6)$$

если m – чётно, и $2p$ состояний фазы

$$\Theta_s = \left\{0, \frac{\pi m}{p}, \frac{2\pi m}{p}, \ldots, \frac{(2p-1)\pi m}{p}\right\}, \qquad (5.3.7)$$

если m – нечётно. Если $L=1$, это единственное состояние решётки. С другой стороны, если $L>1$, имеем дополнительное число состояний, обусловленных парциальным откликом сигнального импульса $g(t)$., Эти дополнительные состояния можно определить, выражая $\theta(t;\mathbf{I})$ через (5.3.4):

$$\theta(t;\mathbf{I}) = 2\pi h \sum_{k=n-L+1}^{n-1} I_k q(t-kT) + 2\pi h I_n q(t-nT). \qquad (5.3.8)$$

Первое слагаемое в правой части (5.3.8) зависит от последовательности информационных символов $(I_{n-1}, I_{n-2}, \ldots, I_{n-L+1})$, которую называют *коррелированным вектором состояний*, и представляет слагаемое фазы, которое соответствует сигнальным импульсам, которые ещё не достигли финальных значений. Второе слагаемое в (5.3.8) представляют вклад фазы, обусловленный самым последним символом I_n. Таким образом, состояние сигнала МНФ (или модулятора) в точке $t=nT$ можно выразить как комбинацию *фазового состояния* и *коррелированного состояния*, обозначаемую так:

$$S_n = \{\theta_n, I_{n-1}, I_{n-2}, \ldots, I_{n-L+1}\} \qquad (5.3.9)$$

для сигнального импульса с парциальным откликом длины LT, где $L>1$. В этом случае число состояний равно

$$N_s = \begin{cases} pM^{L-1} & (m \text{ чётное}), \\ 2pM^{L-1} & (m \text{ нечётное}). \end{cases} \qquad (5.3.10)$$

Теперь предположим, что состояние модулятора в точке $t=nT$ есть S_n. Влияние нового символа в интервале $nT \le t \le (n+1)T$ сводится к изменению состояния от S_n до S_{n+1}. Следовательно, в точке $t=(n+1)T$ состояние становится

$$S_{n+1} = \{\theta_{n+1}, I_n, I_{n-1}, \ldots, I_{n-L+2}\},$$

где

$$\theta_{n+1} = \theta_n + \pi h I_{n-L+1}.$$

Пример 5.3.1. Рассмотрим схему образования двоичной МНФ с индексом модуляции $h=3/4$ и импульсом с парциальным откликом с $L=2$. Определим состояние S_n схемы МНФ и вычертим фазовое дерево и решётку состояний.

Сначала отмечаем, что имеется $2p=8$ состояний фаз, именно

$$\Theta_s = \{0, \pm\tfrac{1}{4}\pi, \pm\tfrac{1}{2}\pi, \pm\tfrac{3}{4}\pi, \pi, \}.$$

Для каждого из этих состояний фаз имеется два состояния, которые обусловлены памятью схемы МНФ. Следовательно, общее число состояний $N_s=16$, именно

$$\begin{array}{cccc} (0,1) & (0,-1) & (\pi,1) & (\pi,-1) \\ (\frac{1}{4}\pi,1) & (\frac{1}{4}\pi,-1) & (\frac{1}{2}\pi,1) & (\frac{1}{2}\pi,-1) \\ (\frac{3}{4}\pi,1) & (\frac{3}{4}\pi,-1) & (-\frac{1}{4}\pi,1) & (-\frac{1}{4}\pi,-1) \\ (-\frac{1}{2}\pi,1) & (-\frac{1}{2}\pi,-1) & (-\frac{3}{4}\pi,1) & (-\frac{3}{4}\pi,-1) \end{array}$$

Если система находится в фазовом состоянии $\theta_n = -\frac{1}{4}\pi$ и $I_{n-1}=-1$, тогда

$$\theta_{n+1} = \theta_n + \pi h I_{n-1} = -\tfrac{1}{4}\pi - \tfrac{3}{4}\pi = -\pi.$$

Решётка состояний иллюстрируется рис. 5.3.1. Путь по решётке состояний, соответствующий последовательности (1,-1,-1,-1,1,1), показан на рис. 5.3.2.

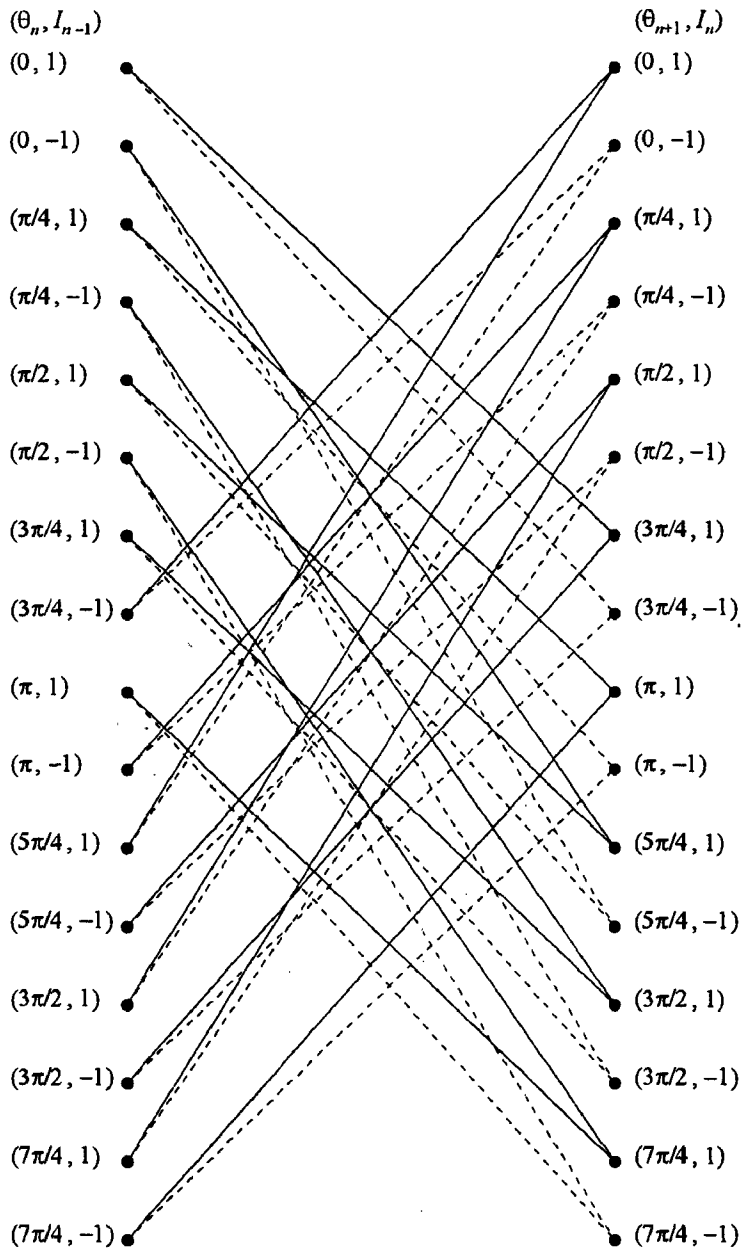


Рис. 5.3.1. Решетка состояний для МНФ с парциальным откликом (L=2) с h=3/4

Для того чтобы нарисовать фазовое дерево, мы должны знать огибающую сигнального импульса $g(t)$. Рисунок 5.3.3 иллюстрирует фазовое дерево, когда $g(t)$ является прямоугольным импульсом длительности $2T$ с начальным состоянием $(0,1)$.

Установив отображение решётки состояний для МНФ, рассмотрим расчёт метрик, формируемых алгоритмом Витерби.

Расчёт метрик. Возвращаясь к математическим основам демодулятора максимального правдоподобия, данным в разд. 5.4.1, легко видеть, что логарифм условной плотности вероятности наблюдаемого сигнала $r(t)$ при условии передачи последовательности символов $\mathbf{I}$ пропорционален метрике взаимной корреляции

$$CM_n(\mathbf{I}) = \int_{-\infty}^{(n+1)T} r(t)\cos\left[\omega_c t + \phi(t;\mathbf{I})\right]dt = CM_{n-1}(\mathbf{I}) + \int_{nT}^{(n+1)T} r(t)\cos\left[\omega_c t + \theta(t;\mathbf{I}) + \theta_n\right]dt. \quad (5.3.11)$$

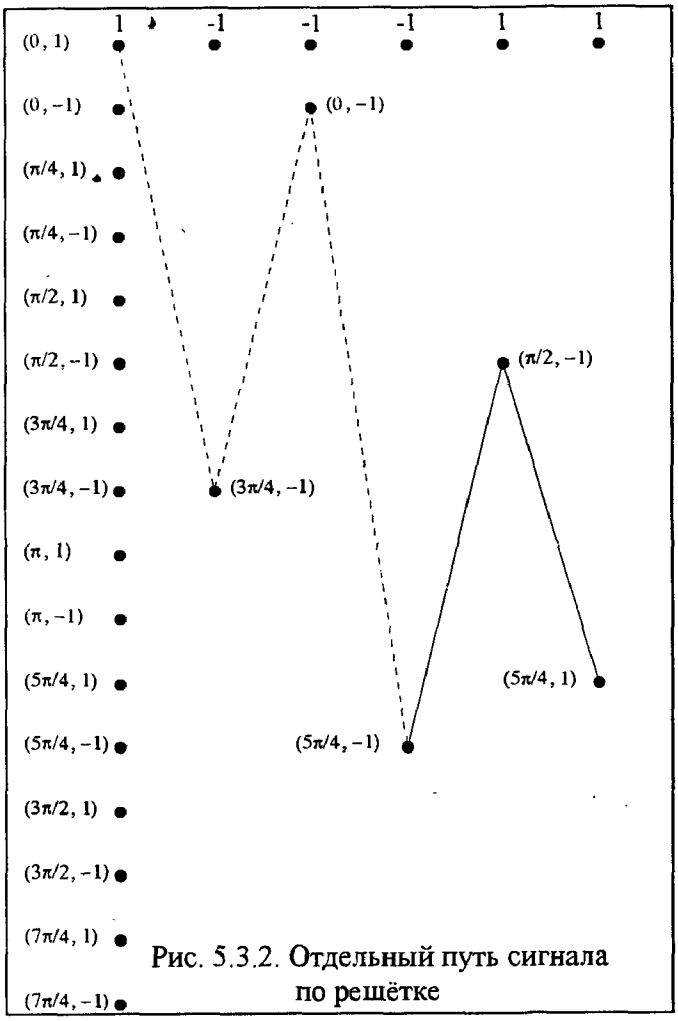


Рис. 5.3.2. Отдельный путь сигнала по решётке

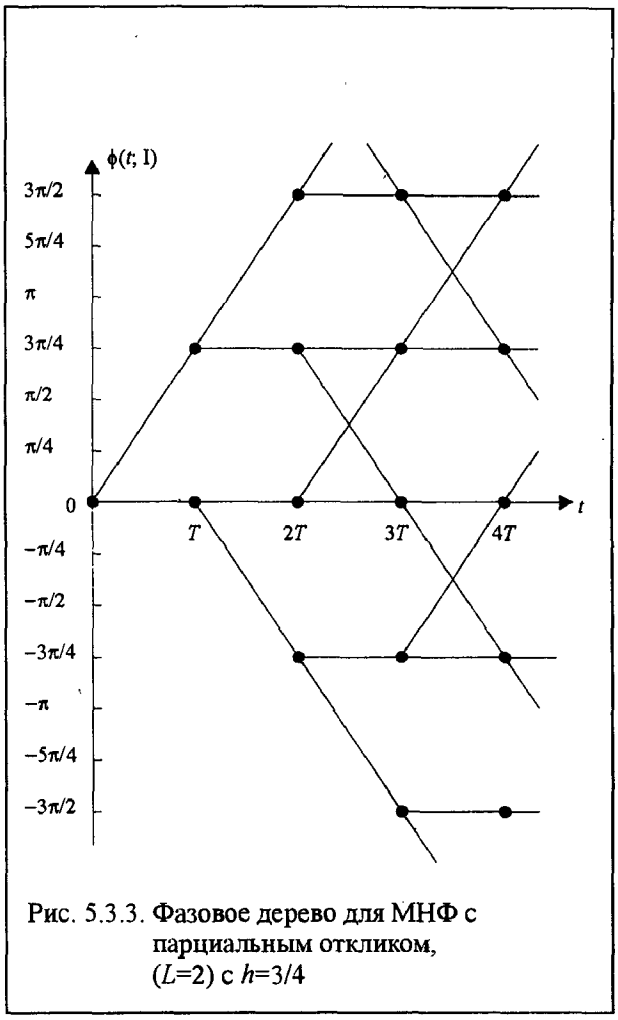


Рис. 5.3.3. Фазовое дерево для МНФ с парциальным откликом, (L=2) с h=3/4

Слагаемое $CM_{n-1}(\mathbf{I})$ представляет метрики выживших путей (последовательностей) до момента nT, а слагаемое

$$v_n(\mathbf{I};\theta_n)=\int_{nT}^{(n+1)T} r(t)\cos\left[\omega_c t+\theta(t;\mathbf{I})+\theta_n\right]dt \qquad (5.3.12)$$

представляет дополнительный прирост метрики, вносимый сигналом на интервале времени $nT\le t\le(n+1)T$. Заметим, что имеются M^L возможных последовательностей $\mathbf{I}=\left(I_n,I_{n-1},\ldots,I_{n-L+1}\right)$ символов и p (или $2p$) возможных состояний фазы $\{\theta_n\}$.

Следовательно, имеем pM^L (или $2pM^L$) различных величин $v_n(\mathbf{I};\theta_n)$, вычисляемых на каждом сигнальном интервале, и каждая величина используется для прироста метрик, соответствующих pM^{L-1} выживших последовательностей от предыдущего сигнального интервала. Общая блок-схема на рис. 5.3.4 иллюстрирует вычисления $v_n(\mathbf{I};\theta_n)$ для декодера Витерби.

Заметим, что число выживших последовательностей в каждом состоянии для процесса декодирования по Витерби равно pM^{L-1} (или $2pM^{L-1}$). Для каждой выжившей последовательности мы имеем M новых приращений $v_n(\mathbf{I};\theta_n)$, которые прибавляются к существующим метрикам, чтобы получить pM^L (или $2pM^L$) последовательностей с pM^L (или $2pM^L$) метриками. Однако их число затем снова уменьшается до pM^{L-1} (или

$2pM^{L-1}$) выживших траекторий с соответствующими метриками за счёт выбора наиболее вероятной последовательности из M последовательностей, сливающихся в каждом узле решётки и отбрасывания остальных $M-1$ последовательностей.

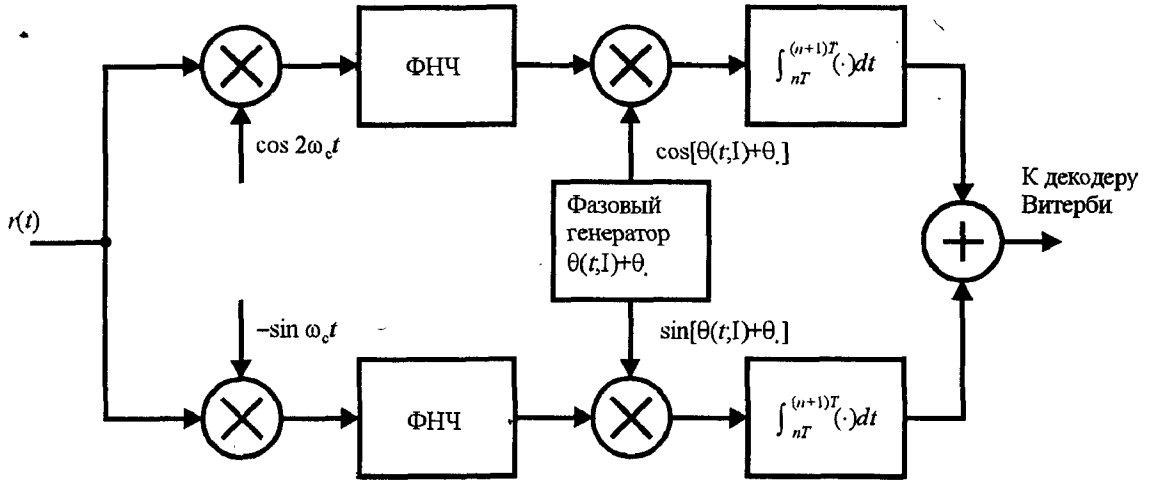


Рис. 5.3.4. Вычисление приращений метрик $\upsilon_n(I;\theta_n)$

5.3.2. Характеристики качества сигналов МНФ

При расчете характеристик качества сигналов МНФ, достигаемого при максимально правдоподобном последовательном оценивании МПО (MLSE), мы должны определить минимум евклидова расстояния путей по решётке, которые начинаются в некотором узле в момент $t=0$ и возвращаются в более позднее время в тот же узел. Как мы теперь покажем, расстояние между двумя путями на решётке относится к соответствующим сигналам.

Предположим, что имеем два сигнала $s_i(t)$ и $s_j(t)$, соответствующие двум фазовым траекториям $\phi(t;\mathbf{I}_i)$ и $\phi(t;\mathbf{I}_j)$. Последовательности $\mathbf{I}_i$ и $\mathbf{I}_j$ должны отличаться в первом символе. Тогда евклидово расстояние между двумя сигналами на интервале длиной NT, где $1/T$ – скорость передачи символов, определяется так:

$$
\begin{aligned}
d_{ij}^2 &= \int_0^{NT}\left[s_i(t)-s_j(t)\right]^2 dt = \int_0^{NT} s_i^2(t)dt + \int_0^{NT} s_j^2(t)dt - 2\int_0^{NT} s_i(t)\cdot s_j(t)dt = \\
&= 2N\mathscr{E} - 2\frac{2\mathscr{E}}{T}\int_0^{NT}\cos\left[\omega_c t+\phi(t;\mathbf{I}_j)\right]\cos\left[\omega_c t+\phi(t;\mathbf{I}_j)\right]dt = \\
&= 2N\mathscr{E} - 2\frac{2\mathscr{E}}{T}\int_0^{NT}\cos\left[\phi(t;\mathbf{I}_i)-\phi(t;\mathbf{I}_j)\right]dt = \\
&= \frac{2\mathscr{E}}{T}\int_0^{NT}\left\{1-\cos\left[\phi(t;\mathbf{I}_i)-\phi(t;\mathbf{I}_j)\right]\right\}dt .
\end{aligned}
\qquad (5.3.13)
$$

Следовательно, согласно (5.3.13) евклидово расстояние сводится к разности фаз между путями на решётке состояний.

Желательно выразить расстояние d_{ij}^2 через битовую энергию. Поскольку $\mathscr{E}=\mathscr{E}_b\log_2 M$, (5.3.13) можно выразить так:

$$
d_{ij}^2 = 2\mathscr{E}_b\delta_{ij}^2, \qquad (5.3.14)
$$

где δ_{ij}^2 определено как

$$\delta_{ij}^2 = \frac{\log_2 M}{T}\int_0^{NT}\left\{1-\cos\left[\phi\left(t;\mathbf{I}_i\right)-\phi\left(t;\mathbf{I}_j\right)\right]\right\}dt. \qquad (5.3.15)$$

Далее видим, что $\phi\left(t;\mathbf{I}_i\right)-\phi\left(t;\mathbf{I}_j\right)=\phi\left(t;\mathbf{I}_i-\mathbf{I}_j\right)$, так что, обозначив $\xi=\mathbf{I}_i-\mathbf{I}_j$, (5.3.15) можно переписать в виде

$$\delta_{ij}^2 = \frac{\log_2 M}{T}\int_0^{NT}\left\{1-\cos\phi\left(t;\xi\right)\right\}dt, \qquad (5.3.16)$$

где любой элемент из ξ может принять значения $0,\pm 2,\pm 4,\ldots\pm 2(M-1)$, кроме ξ_0, которое не может быть равным 0.

Вероятность ошибки МНФ определяется слагаемым δ_{ij}^2, соответствующим минимальному евклидову расстоянию, и ее можно выразить так:

$$P_M = K_{\delta_{\min}} Q\left(\sqrt{\frac{\mathscr{E}_b}{N_0}\delta_{\min}^2}\right), \qquad (5.3.17)$$

где

$$\delta_{\min}^2 = \lim_{N\to\infty}\min_{i,j}\delta_{ij}^2 = \lim_{N\to\infty}\min_{i,j}\left\{\frac{\log_2 M}{T}\int_0^{NT}\left[1-\cos\phi\left(t;\mathbf{I}_i-\mathbf{I}_j\right)\right]dt\right\}. \qquad (5.3.18)$$

Заметим, что для обычной двоичной ФМ без памяти $N=1$ и $\delta_{\min}^2=\delta_{12}^2=2$. Следовательно, (5.3.17) согласуется с нашим прежним результатом.

Поскольку $\delta_{\min}^2$ характеризует качество МНФ с МППО, мы можем исследовать влияние на $\delta_{\min}^2$ изменения объёма алфавита M, индекса модуляции h и длительности переданного импульса в системе МНФ с парциальным откликом.

Прежде всего рассмотрим МНФ с полным откликом $(L=1)$. Если возьмем сначала $M=2$, то заметим, что последовательности

$$\begin{aligned}\mathbf{I}_i &= +1,-1,I_2,I_3\\ \mathbf{I}_j &= -1,+1,I_2,I_3,\end{aligned} \qquad (5.3.19)$$

которые отличаются при $k=0;1$ и совпадают при $k\geq 2$, приводят к двум фазовым траекториям, которые сливаются после второго символа. Это соответствует разностной последовательности

$$\xi=\{2,-2,0,0,\ldots\}. \qquad (5.3.20)$$

Евклидово расстояние для этих последовательностей легко вычислить из (5.3.16), а затем найти верхнюю границу для $\delta_{\min}^2$. Эта верхняя граница для $M=2$ равна

$$d_B^2(h)=2\left(1-\frac{\sin 2\pi h}{2\pi h}\right),\quad M=2. \qquad (5.3.21)$$

Например, когда $h=\frac{1}{2}$, что ведет к ММС, имеем $d_B^2\left(\frac{1}{2}\right)=2$, так что $\delta_{\min}^2\left(\frac{1}{2}\right)\leq 2$.

Для $M>2$ и при полном отклике МНФ также можно легко увидеть, что фазовые траектории сливаются при $t=2T$. Следовательно, верхнюю границу δ_{min}^2 можно получить, рассматривая последовательности разностей фаз $\xi=\{\alpha,-\alpha,0,0,\ldots\}$, где

$$\alpha=\pm 2,\pm 4,\ldots,\pm 2(M-1).$$

Эти последовательности дают верхнюю границу

$$d_B^2(h)=\min_{1\leq k\leq M-1}\left\{\left(2\log_2 M\right)\left(1-\frac{\sin 2k\pi h}{2k\pi h}\right)\right\}. \qquad (5.3.22)$$

Графики зависимости $d_B^2(h)$ от h для $M = 2, 4, 8, 16$ показаны на рис. 5.3.5.

Из этих графиков очевидно, что большие выигрыши в качестве можно достичь увеличением объема алфавита M. Надо вспомнить, однако, что $\delta_{\min}^2(h) \le d_B^2(h)$. Это значит, что верхняя граница не достижима для всех значений h.

Минимальное евклидово расстояние $\delta_{\min}^2(h)$ было определено расчетом по (5.3.16) для различных сигналов МНФ Аулином и Сандбергом (1981). Например, рис. 5.3.6 иллюстрирует зависимость евклидова расстояния для двоичной ЧМНФ, как функцию индекса модуляции h при различном числе наблюдаемых символьных интервалов N $(N = 1, 2, 3, 4)$.

Показана также верхняя граница $d_B^2(h)$, определяемая (5.3.21). В частности, видим, что когда $h = \frac{1}{2}$, $\delta_{\min}^2\left(\frac{1}{2}\right) = 2$, это является тем же среднеквадратическим расстоянием, что для ФМ (двоичной или четверичной) с $N = 1$.

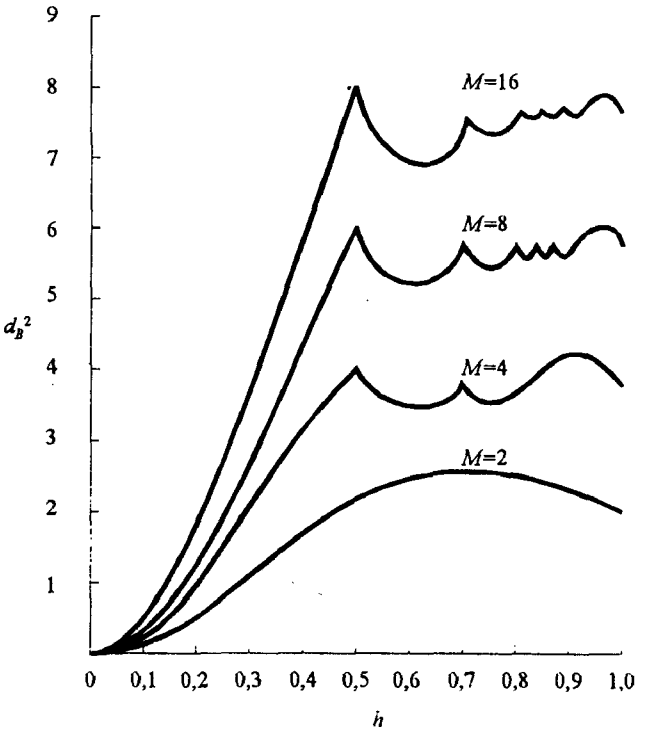


Рис. 5.3.5. Верхние границы d_B^2 как функция индекса модуляции для сигнала МНФ с полным откликом и прямоугольным импульсом [*Aulin* и *Sundberg* (1984), © 1984, *John Wiley Ltd*]

С другой стороны, требуемое число интервалов наблюдения для ММС равно $N = 2$, из чего получаем $\delta_{\min}^2\left(\frac{1}{2}\right) = 2$. Следовательно, качество ММС с МПНО сравнимо с качеством (двоичной или четверичной) ФМ, что мы видели раньше.

Из рис. 5.3.6 мы также замечаем, что оптимальный индекс модуляции для двоичной ЧМНФ равен $h = 0{,}715$, когда число интервалов наблюдения равно $N = 3$. Это дает $\delta_{\min}^2(0{,}715) = 2{,}43$ или выигрыш в 0,85 дБ относительно ММС.

Рисунок 5.3.7 иллюстрирует зависимость евклидова расстояния от h для ЧМНФ с $M = 4$ и числом интервалов наблюдения N как параметра.

Также показана (штриховой кривой, которая не достигается) верхняя граница d_B^2, рассчитанная по (5.3.22). Заметим, что $\delta_{\min}^2$ достигает верхней границы при некоторых значениях h при одинаковых N. В частности, отметим, что максимальная величина d_B^2, которая получается при $h \approx 0{,}9$, приближённо достигается при $N = 8$ наблюдаемых символьных интервалах. Действительный максимум достигается при $h = 0{,}914$ с $N = 9$. Для этого случая $\delta_{\min}^2(0{,}914) = 4{,}2$, что даёт выигрыш 3,2 дБ относительно ММС. Также отметим, что евклидово расстояние имеет минимумы при $h = \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, 1$ и других значениях. Эти значения h называют *слабыми индексами модуляции*, и их избегают. Похожие результаты возможны для больших значений M, и их можно найти в работе Аулина и Сандберга (1981) и в публикациях Андерсона и др. (1986).

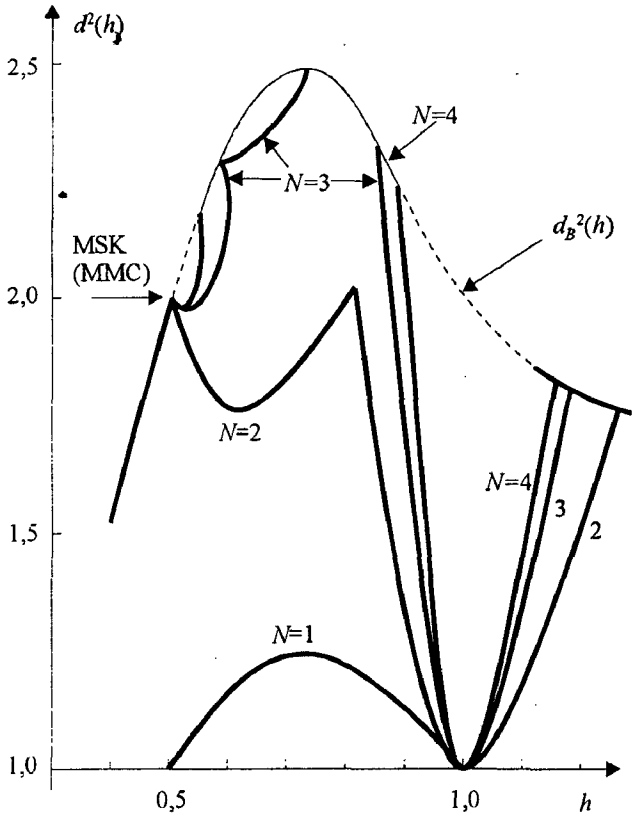


Рис. 5.3.6. Минимальное евклидово расстояние как функция индекса модуляции для двоичной ЧМНФ. Верхняя граница d_B^2. [*Aulin* и *Sundberg* (1981), © 1981, *IEEE*]

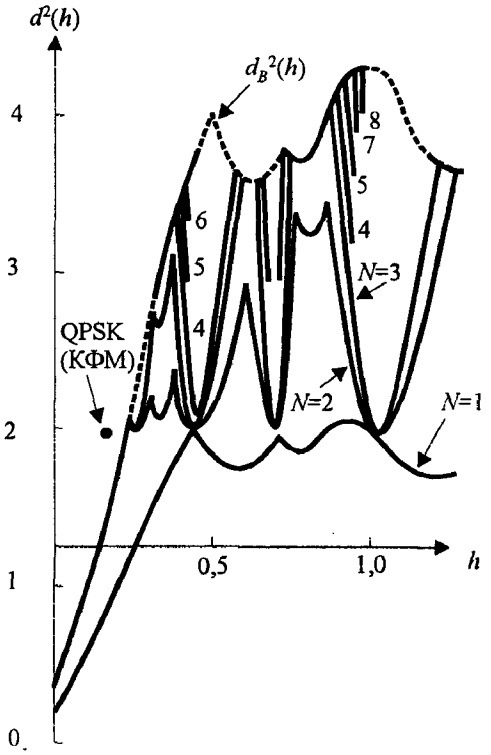


Рис. 5.3.7. Минимальное евклидово расстояние как функция индекса модуляции для четверичной ЧМНФ. Верхняя граница d_B^2. [*Aulin* и *Sundberg* (1981), © 1981, *IEEE*]

Большие выигрыши в качестве можно также достичь при МПО и для МНФ, используя сигналы с парциальным откликом. Например, граница расстояния $d_B^2(h)$ при парциальном отклике импульса приподнятого косинуса, определяемого выражением

$$g(t)=\begin{cases}\dfrac{1}{2LT}\left(1-\cos\dfrac{2\pi t}{2LT}\right) & (0\le t\le LT),\\ 0 & (\text{для других } T),\end{cases} \qquad (5.3.23)$$

показана на рис. 5.3.8 для $M=2$.

Здесь заметим, что с ростом L параметр d_B^2 также достигает больших значений. Ясно, что качество МНФ улучшается по мере увеличения корреляционной памяти L, но следует также увеличить h для того, чтобы достичь больших значений d_B^2. Поскольку больший индекс модуляции требует большей полосы частот (при фиксированном L), в то время как большая длина памяти L (при фиксированном h) требует меньшей полосы частот, то лучше сравнивать евклидово расстояние как функцию от нормированной полосы частот $2WT_b$, где W – полоса с концентрацией 99% мощности, а T_b – битовый интервал. Рисунок 5.3.9 иллюстрирует этот вид сравнения с ММС, используемой как точка отсчёта (0 дБ).

Из этого рисунка видно, что имеется выигрыш в несколько децибел при использовании сигналов с парциальным откликом и больших значений объема алфавита. Главная цена, которую нужно платить за этот выигрыш качества, – это экспоненциально растущая сложность в реализации декодера Витерби.

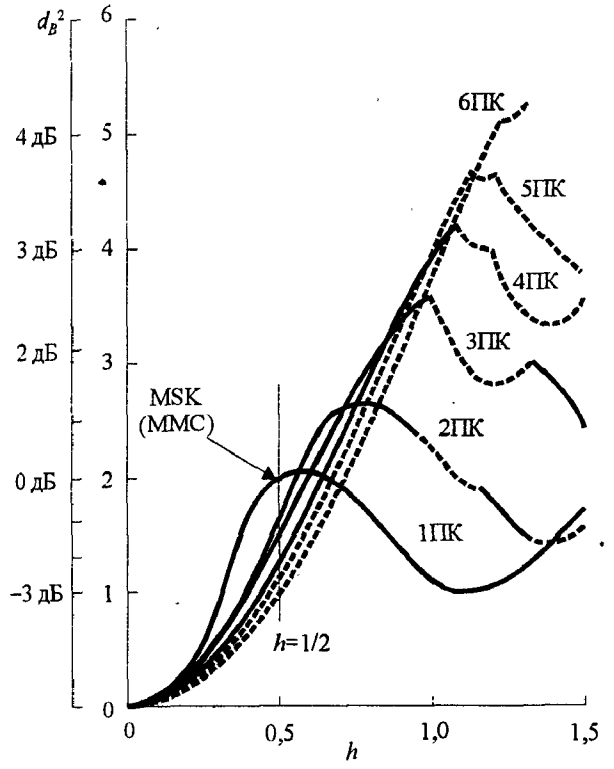


Рис. 5.3.8. Верхняя граница d_B^2 для минимального расстояния двоичной МНФ с парциальным откликом (импульс приподнятого косинуса). [*Sundberg* (1986), © 1986, *IEEE*]

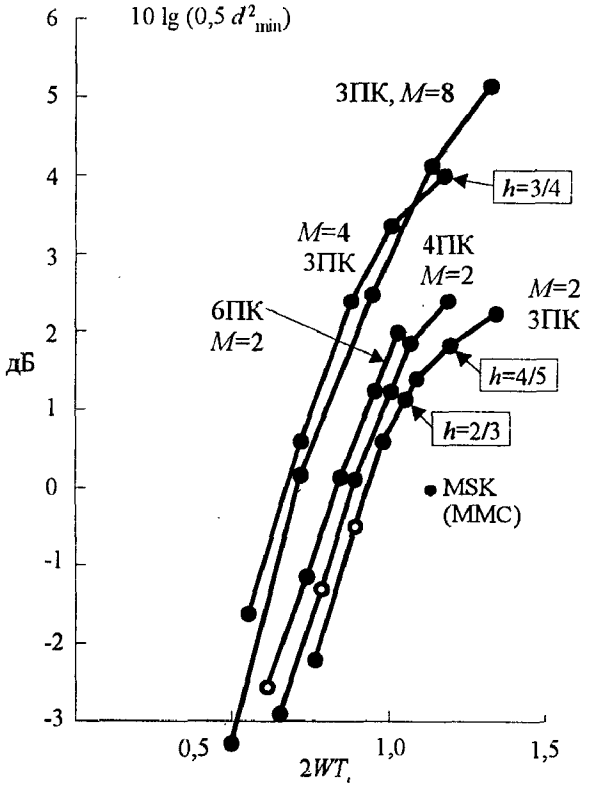


Рис. 5.3.9. Выирыш в полосе частот по мощности для сигнала МНФ с частичным откликом (импульс приподнятого косинуса – ПК). W – полоса, содержащая 99 % мощности. [*Sundberg* (1986), © 1986, *IEEE*]

Результаты качества, иллюстрируемые на рис. 5.3.9, показывают, что выигрыш относительно ММС в 3...4 дБ можно легко получить без относительного расширения полосы частот, используя импульс приподнятого косинуса и МНФ с $M = 4$ и парциальным откликом. Хотя эти результаты получены для сигнальных импульсов приподнятого косинуса, похожие выигрыши можно достичь с другими огибающими импульсов при парциальном отклике. Подчеркнем, что этот выигрыш в ОСШ достигается введением памяти при модуляции сигнала и использованием памяти при демодуляции сигнала. Кодирование здесь не вносит избыточности. Фактически код здесь встраивается в модулятор, и декодирование решётчатого типа (Витерби) использует фазовые связи в сигнале МНФ.

Дополнительный выигрыш в качестве можно достичь введением дополнительной избыточности при кодировании и увеличением размера объема алфавита как средства, при котором сохраняется фиксированная полоса частот. В частности, МНФ с решётчатым кодированием, с использованием относительно простых свёрточных кодов, широко исследуется и много результатов имеется в технической литературе. Декодер Витерби для МНФ со свёрточным кодированием сегодня используют для учёта памяти, присущей и коду, и МНФ сигналу. Выигрыш качества порядка 4...6 дБ, обусловленный кодированием ММС с сохранением полосы частот, был продемонстрирован с комбинированием сверточного кодирования и МНФ. Обильные численные результаты для кодированной МНФ даны Линделлом (1985).

МНФ со многими индексами (multi-*h*). Изменением индекса модуляции от одного сигнального интервала к другому можно увеличить минимальное евклидово расстояние

δ^2_{min} между парами фазовых траекторий и таким образом улучшить выигрыш качества относительно МНФ с фиксированным индексом h. Обычно МНФ со многими индексами h использует фиксированное число H индексов модуляции, которые меняются циклически в соседних сигнальных интервалах. Таким образом, фаза сигнала меняется кусочно-линейно.

Существенный выигрыш в ОСШ достигается использованием только небольшого количества различных значений h. Например, для МНФ с полным откликом $(L=1)$ и $H=2$ можно получить выигрыш в 3 дБ относительно двоичной или четверичной ФМ. При увеличении H до $H=4$ можно получить выигрыш в 4,5 дБ относительно ФМ. Выигрыш качества можно также увеличить с увеличением объема сигнального алфавита. Таблица 5.3.1 показывает выигрыш качества, достигаемый при $M=2$, 4 и 8 для различных значений H.

Таблица 5.3.1 Максимальные значения верхней границы d_B^2 для линейной МНФ с переменным индексом

M	H	Max d_B^2	Выигрыш относительно ММС, дБ	h_1	h_2	h_3	h_4	$\bar{h}$
2	1	2,43	0,85	0,715				0,715
2	2	4,0	3,0	0,5	0,5			0,5
2	3	4,88	3,87	0,620	0,686	0,714		0,673
2	4	5,69	4,54	0,73	0,55	0,73	0,55	0,64
4	1	4,23	3,25	0,914				0,914
4	2	6,54	5,15	0,772	0,772			0,772
4	3	7,65	5,83	0,795	0,795	0,795		0,795
8	1	6,14	4,87	0,964				0,964
8	2	7,50	5,74	0,883	0,883			0,883
8	3	8,40	6,23	0,879	0,879	0,879		0,879

Aulin и *Sundberg* (1982b)

На рис. 5.3.10 показана верхняя граница минимального евклидова расстояния для нескольких величин M и H. По оси абсцисс отложено среднее значение $\bar{h}=\frac{1}{H}\sum_i h_i$. Отметим, что основной выигрыш в качестве получается, когда H увеличивается от $H=1$ до $H=2$. Для $H>2$ дополнительный выигрыш относительно мал для малых величин $\{h_i\}$. С другой стороны, существенный выигрыш качества достигается увеличением объема алфавита M.

Результаты, показанные выше, имеют место для МНФ с полным откликом. Наверняка существует польза от МНФ со многими индексами h при парциальном отклике в попытке дальнейшего улучшения качества. Можно предвидеть, что такие схемы обеспечат дополнительный выигрыш качества, но имеющиеся численные результаты для МНФ со многими индексами h и парциальным откликом ограничены. Интересующемуся читателю рекомендуется статья Аулина и Сандберга (1982).

Многоамплитудная МНФ. Многоамплитудная МНФ (МАМНФ) является по существу схемой комбинирования амплитудной и фазовой модуляции, которая позволяет увеличить сигнальный алфавит относительно МНФ до другой размерности и таким образом достичь большей скорости передачи данных в частотно-ограниченном канале. Одновременно комбинирование АМ с МНФ приводит к эффективной по полосе частот технике модуляции. Мы уже наблюдали спектральные характеристики МАМНФ в разд. 4.3.

Характеристики качества МАМНФ были исследованы Маллиганом (1988) для некодированной и решётчато-кодированной МНФ. Особый интерес представляет результат, что решётчато-кодированная МНФ с двумя уровнями амплитуд даёт выигрыш в 3...4 дБ относительно ММС без существенного увеличения полосы частот сигнала.

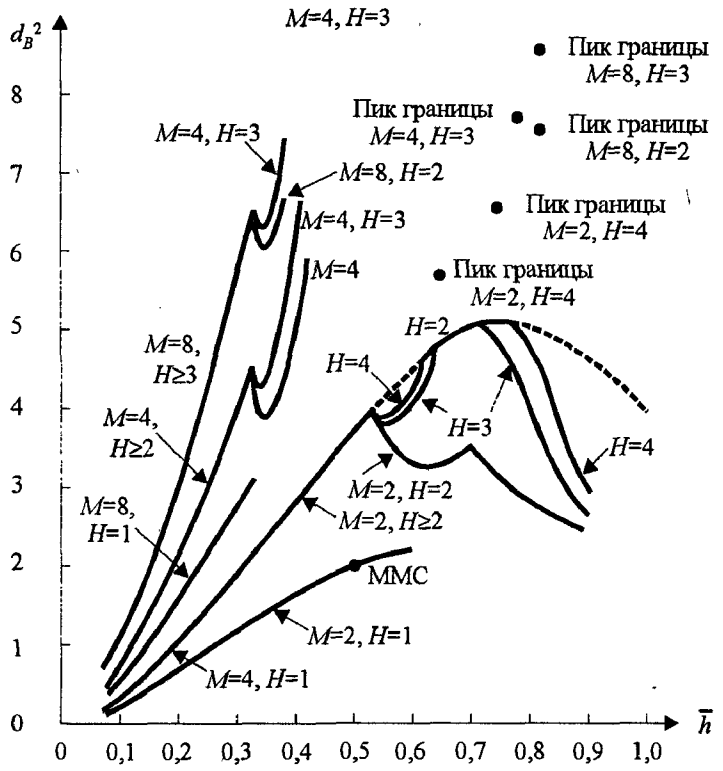


Рис. 5.3.10. Верхние границы для минимального среднеквадратического эвклидова расстояния при различных значениях M и H
[*Aulin u Sundberg* (1982*b*), © 1982, *IEEE*]

5.3.3. Посимвольные детектирование сигналов МНФ

Помимо МП детектора последовательностей, имеются другие типы детекторов, которые могут использоваться для обнаружения информационной последовательности в МНФ сигнале. В этом разделе мы рассмотрим посимвольный детектор. Один тип посимвольного детектора – это тот, который описан в разд. 5.1.15 и который использует память МНФ при формировании согласованной фильтрации или взаимной корреляции на нескольких тактовых интервалах. Однако из-за его вычислительной сложности этот рекуррентный алгоритм непосредственно не применяется для детектирования МНФ. Вместо этого были описаны два сходных субоптимальных метода посимвольного детектирования в публикациях Де Буда (1972), Осборна и Лунтца (1974) и Шонхофа (1976). Один из них функционально эквивалентен алгоритму, данному в разд. 5.1.15, а второй является субоптимальной аппроксимацией первого. Мы опишем эти два метода в контексте демодуляции сигналов ЧМНФ, для которых эти алгоритмы применяются непосредственно.

Чтобы описать эти методы, предположим, что сигнал наблюдается на заданном сигнальном интервале и на D сигнальных интервалах в будущем при решении об информационном символе, переданном на заданном сигнальном интервале. Блок-схема демодулятора, выполненного как блок взаимокорреляторов, показана на рис. 5.3.11.

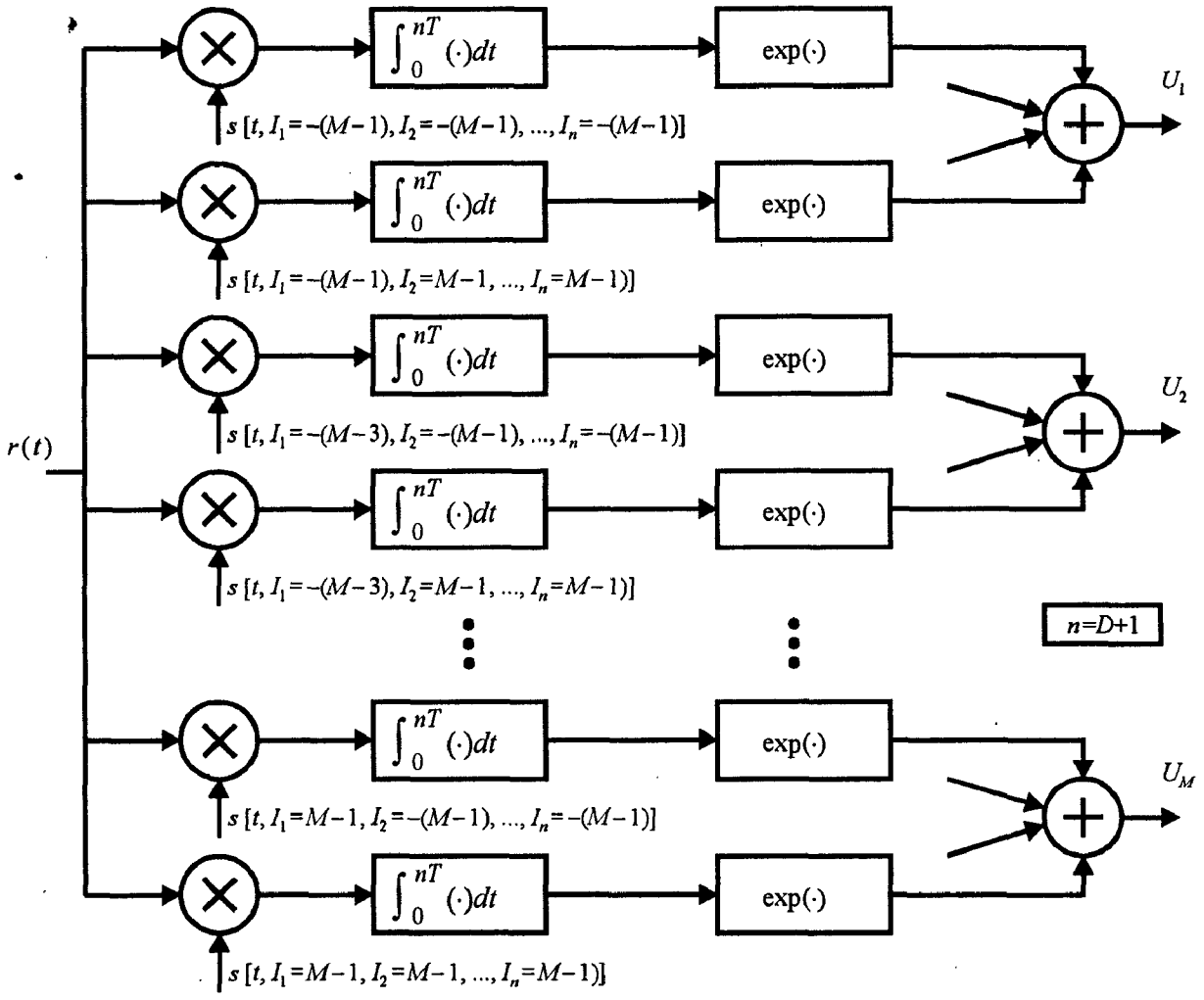


Рис. 5.3.11. Блок-схема демодулятора для детектирования ЧМНФ

Напомним, что сигнал ЧМНФ, переданный на n-м сигнальном интервале можно записать так:

$$s(t) = \mathrm{Re}\left[\upsilon(t)\mathrm{e}^{j2\pi f_c t}\right],$$

где

$$\upsilon(t) = \exp\left\{j\left[\frac{\pi h\left[t-(n-1)T\right]I_n}{T} + \pi h\sum_{k=0}^{n-1} I_k + \phi_0\right]\right\},$$

$h = 2f_d T -$ индекс модуляции, $f_d -$ максимальная девиация частоты, $\phi_0 -$ начальное значение фазы несущей.

Для декодирования символа I_1 взаимные корреляции, отмеченные на рис. 5.3.11, формируются с упомянутыми сигналами $s\left(t, I_1, I_2, \ldots, I_{1+D}\right)$ для всех возможных M^{D+1} значений символов $I_1, I_2, \ldots, I_{1+D}$, переданных на протяжении $D+1$ сигнальных интервалов. Но эти корреляции фактически образуют величины $r_1, r_2, \ldots, r_{1+D}$, которые оказываются аргументами экспонент, определяющих ФПВ

$$p\left(r_1, r_2, \ldots, r_{1+D} \middle| I_1, I_2, \ldots, I_{1+D}\right).$$

В финале суммирование по M^D возможных значений цепочек символов $I_1, I_2, \ldots, I_{1+D}$ представляет среднее от

$$p(r_1, r_2, \ldots, r_{1+D} | I_1, I_2, \ldots, I_{1+D}) P(I_1, I_2, \ldots, I_{1+D})$$

по M^D возможным значениям цепочек этих символов. M выходов демодуляторов образуют величины для решения, большее из которых выбирается для формирования демодулированных символов. Следовательно, метрики, создаваемые демодулятором, показанным на рис. 5.3.11, эквивалентны величинам для решения, даваемым (5.1.68), на них и основывается решение об I_1.

Сигналы, принятые на соседних сигнальных интервалах демодулируются аналогичным образом. Значит, демодулятор выполняет взаимную корреляцию принятого сигнала, на $D+1$ сигнальных интервалах с M^{D+1} возможными переданными сигналами и формирует данные решения, как это иллюстрируется на рис. 5.3.11. Таким образом, решение, сделанное на m-м сигнальном интервале, базируется на взаимных корреляциях, формируемых на сигнальных интервалах $m, m+1, \ldots, m+D$. Начальная фаза на интервале корреляции длительностью $(D+1)T$ считается известной. С другой стороны, алгоритм, описанный (5.1.76) и (5.1.77), включает в себя дополнительную операцию усреднения по предшествующим продетектированным символам. В этом плане демодулятор, показанный на рис. 5.3.11, отличается от рекуррентного алгоритма, описанного выше. Однако разница несущественная.

Один субоптимальный метод демодуляции, который даёт почти такое же качество, как оптимальный метод, осуществляемый схемой рис. 5.3.11, основывает свои решения на наибольшем выходе блока M^{D+1} взаимных корреляций. Таким образом, вычисление экспоненциальных функций и суммирования оказывается ненужным. Но этот метод эквивалентен выбору символа I_m, для которого максимизируется функция плотности вероятности

$$p(r_m, r_{m+1}, \ldots, r_{m+D} | I_m, I_{m+1}, \ldots, I_{m+D}).$$

Для характеристики качества детектора, показанного на рис. 5.3.11, получены верхние границы и выполнен численный расчет. Рисунок 5.3.12 иллюстрирует качество для двоичной ЧМНФ с $n = D+1$, как параметра. Индекс модуляции $h = 0{,}715$, используемый при получении этих результатов, минимизирует вероятность ошибки, как показано Шонхоффом (1976).

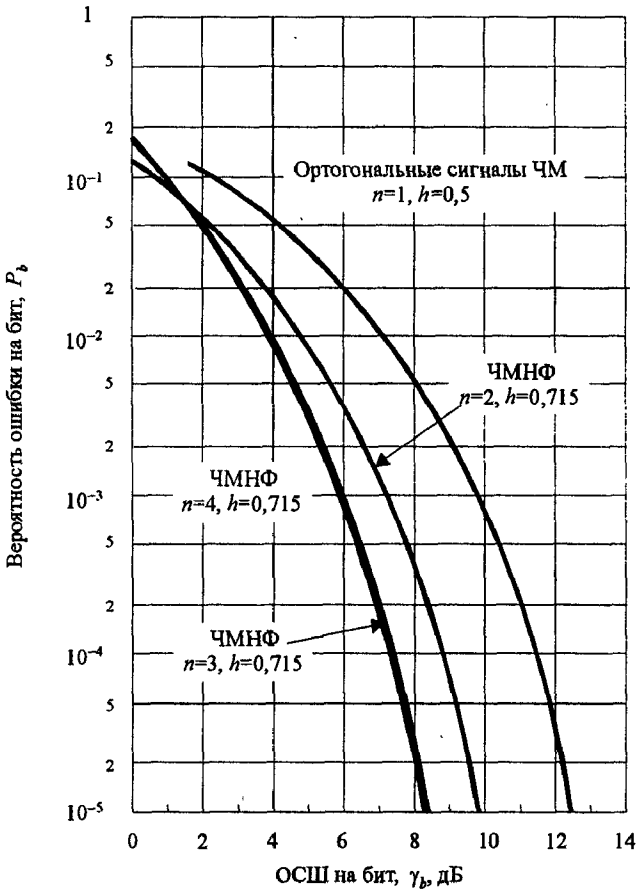


Рис. 5.3.12. Характеристики качества двоичной ЧМНФ при когерентном детектировании

Как видим, имеет место улучшение на 2,5 дБ относительно ортогональной системы ЧМ ($n = 1$) демодулятором, который вычисляет взаимную корреляцию на два сигнальных интервала. Дополнительный выигрыш приблизительно в 1,5 дБ получается путём

расширения времени корреляции на три символа. Дальнейшее расширение интервала корреляции ведет к относительно малому дополнительному выигрышу.

Похожие результаты получены посредствам большего объёма алфавита. Например, рис. 5.3.13 и 5.3.14 иллюстрируют улучшение качества четырехпозиционной и восьмипозиционной ЧМНФ соответственно. Индексы модуляции, данные на этих графиках, такие, что они минимизируют вероятность ошибки на символ.

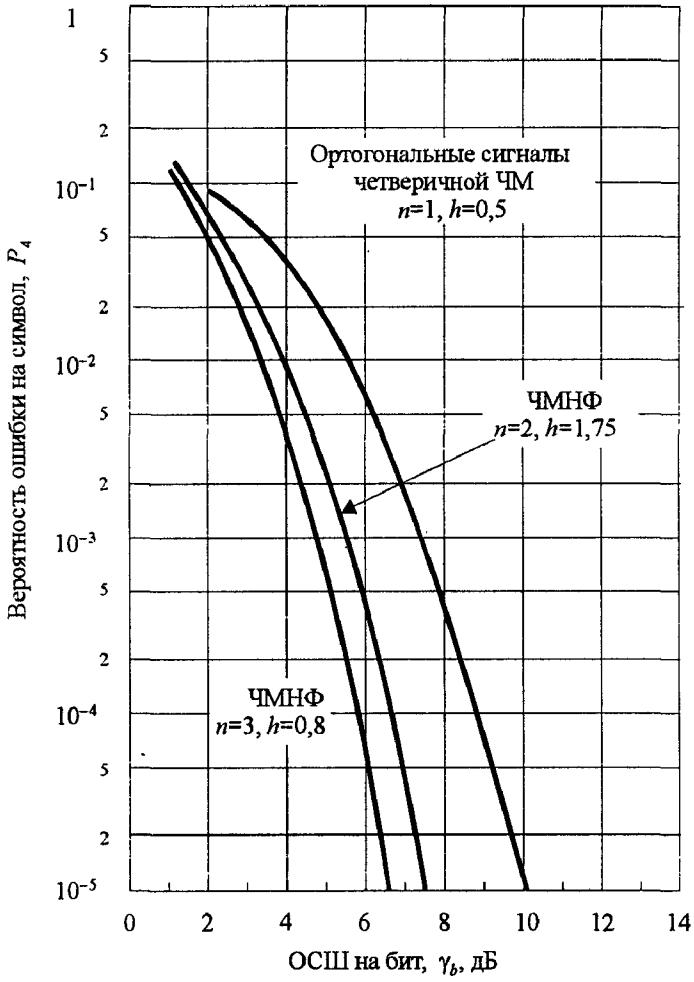


Рис. 5.3.13. Характеристики качества четверичной ЧМНФ при когерентном детектировании

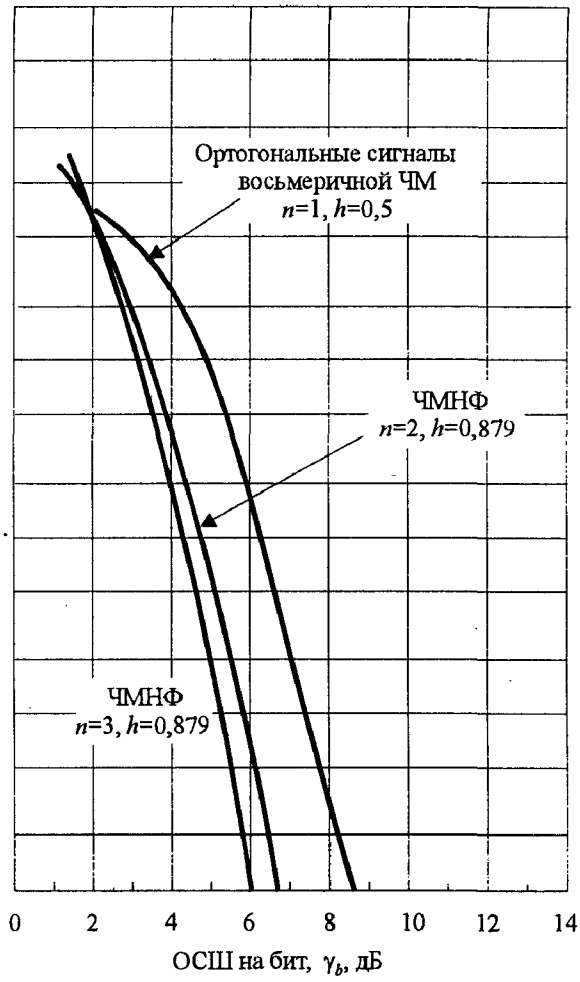


Рис. 5.3.14. Характеристики качества восьмеричной ЧМНФ при когерентном детектировании

Вместо осуществления когерентного детектирования, которое требует знания фазы несущей ϕ_0, мы можем предположить, что ϕ_0 равномерно распределена на интервале $0 \ldots 2\pi$, и выполнить усреднение по фазе при получении величин для решения. Так осуществляется когерентное интегрирование (взаимная корреляция) по $n = D+1$ сигнальным интервалам, но выход корреляторов детектируется по огибающей. Эту процедуру называют *некогерентным детектированием* ЧМНФ. В этой схеме детектирования достигается оптимизация качества путем выбора нечетного n и выполнения решения по среднему символу последовательности из n символов. Численные результаты для вероятности ошибки при некогерентном детектировании ЧМНФ похожи на результаты иллюстрированы выше для когерентного детектирования. Это значит, что выигрыш в 2...3 дБ в качестве достигается путем увеличения интервала корреляции от $n = 1$ до $n = 3$ и до $n = 5$.

5.4. ОПТИМАЛЬНЫЙ ПРИЁМНИК ДЛЯ СИГНАЛОВ СО СЛУЧАЙНОЙ ФАЗОЙ В КАНАЛЕ С АБГШ

В этом разделе рассмотрим синтез оптимального приёмника для модулированных сигналов, когда фаза несущей неизвестна в месте приёма и не делается никаких попыток оценить ее значение. Неопределенность фазы несущей принимаемого сигнала может быть обусловлено одной или большим числом причин. Первое: генераторы, которые используются в передатчике и приёмнике для генерирования сигналов несущей, обычно не имеет фазового синхронизма. Второе: время задержки при распространении сигнала от передатчика к приёмнику в общем не известно точно. Чтобы рассмотреть этот вопрос, положим, что переданный сигнал

$$s(t) = \mathrm{Re}\left[g(t)\mathrm{e}^{j2\pi f_c t}\right]$$

проходит через канал с временем запаздывания t_0, так что принимаемый сигнал

$$s(t-t_0) = \mathrm{Re}\left[g(t-t_0)\mathrm{e}^{j2\pi f_c(t-t_0)}\right] = \mathrm{Re}\left[g(t-t_0)\mathrm{e}^{-j2\pi f_c t_0}\mathrm{e}^{j2\pi f_c t}\right].$$

Фазовый сдвиг несущей, обусловленный временем задержки t_0, равен

$$\phi = -2\pi f_c t_0 .$$

Заметим, что большие изменения в фазе несущей могут возникнуть при относительно малых изменениях времени задержки. Например, если частота несущей f_c=1 МГц, неопределенность или изменения во времени распространения на 0,5 мкс вызовут неопределенность фазы π радиан. В некоторых каналах (например, радиоканалах) время распространения сигнала от передатчика к приёмнику может изменяться быстро и очевидно случайным образом, так что фаза несущей принимаемого сигнала меняется очевидно случайным образом.

При отсутствия знания фазы несущей, мы можем трактовать этот параметр сигнала как случайную величину и определять при этом оптимальный приёмник для извлечения переданной информации из принятого сигнала.

Сначала рассмотрим случай двоичных сигналов и затем рассмотрим M-позиционные сигналы.

5.4.1. Оптимальный приёмник двоичных сигналов

Рассмотрим двоичную систему связи, которую используют для передачи информации двух модулированных по несущей сигналов $s_1(t)$ и $s_2(t)$, где

$$s_m(t) = \mathrm{Re}\left[s_{lm}(t)\mathrm{e}^{j2\pi f_c t}\right], \quad m = 1, 2, \quad 0 \le t \le T, \tag{5.4.1}$$

а $s_{l,m}(t)$, $m = 1, 2$, – эквивалентные низкочастотные сигналы.

Считается, что два сигнала имеют равные энергии

$$\mathscr{E} = \int_0^T s_m^2(t)dt = \frac{1}{2}\int_0^T |s_{lm}(t)|^2 dt \tag{5.4.2}$$

и они характеризуются комплексным коэффициентом корреляции

$$\rho_{12} \equiv \rho = \frac{1}{\mathscr{E}}\int_0^T s_{l1}^*(t)s_{l2}(t)dt . \tag{5.4.3}$$

Считается, что принимаемый сигнал отличается от переданного случайной фазой и подвергается воздействию аддитивного шума

$$n(t) = \mathrm{Re}\left\{\left[n_c(t) + jn_s(t)\right]\mathrm{e}^{j2\pi f_c t}\right\} = \mathrm{Re}\left[z(t)\mathrm{e}^{j2\pi f_c t}\right]. \tag{5.4.4}$$

Таким образом, принимаемый сигнал можно выразить так:

$$r(t) = \mathrm{Re}\left\{\left[s_{lm}(t)\mathrm{e}^{j\phi} + z(t)\right]\mathrm{e}^{j2\pi f_c t}\right\}, \tag{5.4.5}$$

где

$$r_l(t) = s_{lm}(t)e^{j\phi} + z(t), \quad 0 \le t \le T, \tag{5.4.6}$$

– эквивалентный низкочастотный принимаемый сигнал. Этот принимаемый сигнал теперь проходит через демодулятор, отсчёт которого при $t = T$ подается на детектор.

Оптимальный демодулятор. В разд. 5.1.1 мы показали, что, если принимаемый сигнал коррелируется с набором ортонормированных функций $\{f_n(t)\}$, на которые натянуто пространство сигналов, выходы набора корреляторов обеспечивают набор достаточных статистик для детектора с тем, чтобы сделать решение, которое минимизирует среднюю вероятность ошибки. Мы также показали, что набор согласованных фильтров может заменить набор корреляторов.

Похожее ортонормированное разложение можно выполнить по отношению к принимаемому сигналу с неизвестной фазой несущей. Однако математически удобнее иметь дело с эквивалентным низкочастотным сигналом и выполнить сигнальные корреляторы или согласованные фильтры по отношению к эквивалентным низкочастотным сигналам.

Для конкретности: импульсная характеристика фильтра, согласованного с комплексным эквивалентным низкочастотным сигналом $s_l(t)$, $0 \le t \le T$, определяется так (см. задачу 5.6):

$$h_l(t) = s_l^*(T - t), \tag{5.4.7}$$

а выход такого фильтра в момент $t = T$ равен

$$\int_0^T |s_l(t)|^2 dt = 2\mathscr{E}, \tag{5.4.8}$$

где $\mathscr{E}$ – энергия сигнала. Аналогичный результат получается, если сигнал $s_l(t)$ коррелируется с $s_l^*(t)$, а коррелятор стробируется в момент времени $t = T$. Следовательно, оптимальный демодулятор для эквивалентного низкочастотного принимаемого сигнала $s_l(t)$ в (5.4.6) можно реализовать двумя согласованными фильтрами, работающими параллельно, один согласован с $s_{l1}(t)$, а другой с $s_{l2}(t)$. Он показан на рис. 5.4.1. Выходы согласованных фильтров или корреляторов в точках отсчёта являются двумя комплексными числами

$$r_m = r_{mc} + jr_{ms}, \quad m = 1, 2. \tag{5.4.9}$$

Предположим, что передаётся сигнал $s_1(t)$. Тогда легко показать (см. задачу 5.35), что

$$\begin{aligned} r_1 &= 2\mathscr{E}\cos\phi + n_{1c} + j\left(2\mathscr{E}\sin\phi + n_{1s}\right), \\ r_2 &= 2\mathscr{E}|\rho|\cos\left(\phi + \alpha_0\right) + n_{2c} + j\left(2\mathscr{E}|\rho|\sin\left(\phi + \alpha_0\right) + n_{2s}\right), \end{aligned} \tag{5.4.10}$$

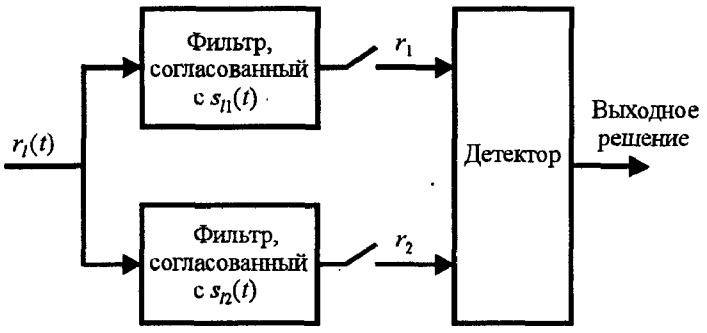


Рис. 5.4.1. Оптимальный приёмник для двоичных сигналов

где ρ – комплексный коэффициент корреляции двух сигналов $s_{l1}(t)$ и $s_{l2}(t)$, который можно выразить как $\rho = |\rho| \exp(j\alpha_0)$. Случайные величины шума n_{1c}, n_{1s}, n_{2c} и n_{2s} – совместно гауссовские с нулевыми средними и равными дисперсиями.

Оптимальный детектор. Оптимальный детектор наблюдает случайные величины $[r_{1c}\ r_{1s}\ r_{2c}\ r_{2s}] = \mathbf{r}$, где $r_1 = r_{1c} + jr_{1s}$ и $r_2 = r_{2c} + jr_{2s}$, и выносит свое решение на основе апостериорных вероятностей $P(\mathbf{s}_m|\mathbf{r})$, $m = 1,2$. Эти вероятности можно выразить так:

$$P(\mathbf{s}_m|\mathbf{r}) = \frac{p(\mathbf{r}|\mathbf{s}_m)P(\mathbf{s}_m)}{p(\mathbf{r})}, \quad m = 1,2, \tag{5.4.11}$$

и, следовательно, оптимальное правило решения можно выразить в виде

$$P(\mathbf{s}_1|\mathbf{r}) \underset{s_2}{\overset{s_1}{\gtrless}} P(\mathbf{s}_2|\mathbf{r})$$

или, что эквивалентно,

$$\frac{p(\mathbf{r}|\mathbf{s}_1)}{p(\mathbf{r}|\mathbf{s}_2)} \underset{s_2}{\overset{s_1}{\gtrless}} \frac{P(\mathbf{s}_2)}{P(\mathbf{s}_1)}. \tag{5.4.12}$$

Отношение ФПВ в левой части (5.4.12) – это *отношение правдоподобия*, которое мы обозначим так:

$$\Lambda(\mathbf{r}) = \frac{p(\mathbf{r}|\mathbf{s}_1)}{p(\mathbf{r}|\mathbf{s}_2)}. \tag{5.4.13}$$

Правая часть (5.4.12) – отношение двух априорных вероятностей, которое принимает значение 1, когда два сигнала равновероятны.

ФПВ $p(\mathbf{r}|\mathbf{s}_1)$ и $p(\mathbf{r}|\mathbf{s}_2)$ можно получить путём усреднения условной ФПВ $p(\mathbf{r}|\mathbf{s}_m, \phi)$ по случайной фазе несущей с ФПВ $p(\phi)$, т.е.

$$p(\mathbf{r}|\mathbf{s}_m) = \int_0^{2\pi} p(\mathbf{r}|\mathbf{s}_m, \phi)p(\phi)d\phi. \tag{5.4.14}$$

Мы выполним интегрирование (5.4.14) для специального случая, когда два сигнала ортогональны[1], т.е. $\rho = 0$. В этом случае выходы демодулятора

$$\begin{aligned} r_1 &= r_{1c} + jr_{1s} = 2\mathscr{E}\cos\phi + n_{1c} + j(2\mathscr{E}\sin\phi + n_{1s}), \\ r_2 &= r_{2c} + jr_{2s} = n_{2c} + jn_{2s}, \end{aligned} \tag{5.4.15}$$

где $(n_{1c},\ n_{1s},\ n_{2c},\ n_{2s})$ – взаимно некоррелированные и, следовательно, статистически независимые гауссовские случайные величины с нулевыми средними (смотри задачу 5.25). Значит, совместную ФПВ для $\mathbf{r} = [r_{1c}\ r_{1s}\ r_{2c}\ r_{2s}]$ можно выразить как произведение системных ФПВ. Следовательно,

---

[1]Речь идет об ортогональности двух комплексных эквивалентных низкочастотных сигналов $s_{l1}(t)$ и $s_{l2}(t)$, что предполагает *ортогональность в усиленном смысле* вещественных полосовых сигналов $s_1(t)$ и $s_2(t)$. Это весьма ценное для теории некогерентного приема сигналов определение введено Л.М. Финком. в 1957 г. [ ] (прп).

$$p(r_{1c}, r_{1s}|\mathbf{s}_1, \phi) = \frac{1}{2\pi\sigma^2}\exp\left[-\frac{(r_{1c} - 2\mathscr{E}\cos\phi)^2 + (r_{1s} - 2\mathscr{E}\sin\phi)^2}{2\sigma^2}\right],$$
$$p(r_{2c}, r_{2s}) = \frac{1}{2\pi\sigma^2}\exp\left[-\frac{r_{2c}^2 + r_{2s}^2}{2\sigma^2}\right], \tag{5.4.16}$$

где $\sigma^2 = 2\mathscr{E}N_0$.

Равномерное распределение для фазы несущей ϕ представляет наибольшее незнание для детектора. Она называется ФПВ *с наименьшим предпочтением* для ϕ. В этом случае $p(\phi) = 1/2\pi, \quad 0 \le \phi \le 2\pi$. Подставив $p(\phi)$ в интеграл (5.4.14), получим

$$\frac{1}{2\pi}\int_0^{2\pi} p(r_{1c}, r_{1s}|\mathbf{s}_1, \phi)d\phi =$$
$$= \frac{1}{2\pi}\exp\left(-\frac{r_{1c}^2 + r_{1s}^2 + 4\mathscr{E}^2}{2\sigma^2}\right)\frac{1}{2\pi}\int_0^{2\pi}\exp\left[\frac{2\mathscr{E}(r_{1c}\cos\phi + r_{1s}\sin\phi)}{\sigma^2}\right]d\phi. \tag{5.4.17}$$

Но

$$\frac{1}{2\pi}\int_0^{2\pi}\exp\left[\frac{2\mathscr{E}(r_{1c}\cos\phi + r_{1s}\sin\phi)}{\sigma^2}\right]d\phi = I_0\left(\frac{2\mathscr{E}\sqrt{r_{1c}^2 + r_{1s}^2}}{\sigma^2}\right), \tag{5.4.18}$$

где $I_0(x)$ – модулированная функция Бесселя нулевого порядка, определенная (2.1.120).

При выполнении интегрирования, аналогичного (5.4.17) в предположении, что передан сигнал $s_2(t)$, получим результат

$$p(r_{2c}, r_{2s}|\mathbf{s}_2) = \frac{1}{2\pi}\exp\left(-\frac{r_{2c}^2 + r_{2s}^2 + 4\mathscr{E}^2}{2\sigma^2}\right)I_0\left(\frac{2\mathscr{E}\sqrt{r_{2c}^2 + r_{2s}^2}}{\sigma^2}\right). \tag{5.4.19}$$

Если подставить эти результаты в отношение правдоподобия, определяемое (5.4.13), получим результат

$$\frac{I_0\left(2\mathscr{E}\sqrt{r_{1c}^2 + r_{1s}^2}/\sigma^2\right)}{I_0\left(2\mathscr{E}\sqrt{r_{2c}^2 + r_{2s}^2}/\sigma^2\right)} \underset{s_2}{\overset{s_1}{\gtrless}} \frac{P(\mathbf{s}_2)}{P(\mathbf{s}_1)}. \tag{5.4.20}$$

Таким образом, оптимальный детектор вычисляет две огибающие $\sqrt{r_{1c}^2 + r_{1s}^2}$ и $\sqrt{r_{2c}^2 + r_{2s}^2}$ и соответствующие значения функции Бесселя $I_0\left(2\mathscr{E}\sqrt{r_{1c}^2 + r_{1s}^2}/\sigma^2\right)$ и $I_0\left(2\mathscr{E}\sqrt{r_{2c}^2 + r_{2s}^2}/\sigma^2\right)$ для того, чтобы сформировать отношение правдоподобия. Мы видим, что эти вычисления требуют знания дисперсии шума σ^2. Затем отношение правдоподобия сравнивается с порогом $P(\mathbf{s}_2)/P(\mathbf{s}_1)$, чтобы определить, какой сигнал передан.

Существенные упрощения в реализации оптимального детектора возникают, когда оба сигнала равновероятны. В этом случае порог равен единице и с учетом монотонного изменения функции Бесселя, показанного на рис. 5.4.2, правило оптимального детектирования упрощается:

$$\sqrt{r_{1c}^2 + r_{1s}^2} \underset{s_2}{\overset{s_1}{\gtrless}} \sqrt{r_{2c}^2 + r_{2s}^2}. \tag{5.4.21}$$

Таким образом, оптимальный детектор основывает свое решение на двух огибающих $\sqrt{r_{1c}^2 + r_{1s}^2}$ и $\sqrt{r_{2c}^2 + r_{2s}^2}$, и поэтому он называется *детектором огибающей*.

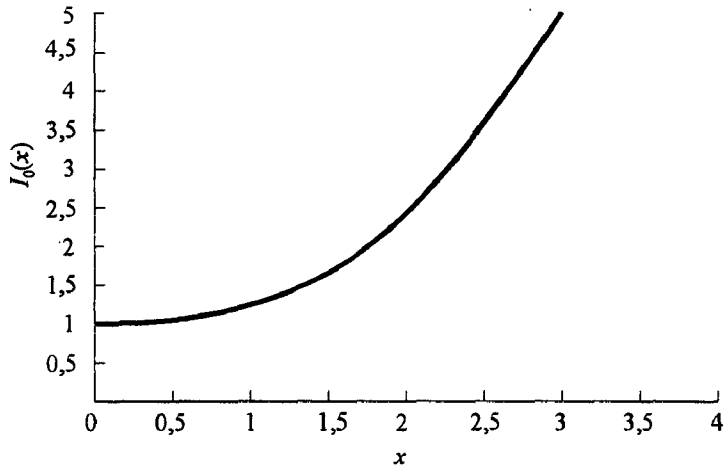


Рис. 5.4.2. График $I_0(x)$

Мы видим, что вычисление отсчётов огибающих принимаемого сигнала на выходе демодулятора делает фазу сигнала не относящейся к делу при решениях о том, какой сигнал передавался. Эквивалентно решение можно сделать на основе квадратов огибающих $r_{1c}^2 + r_{1s}^2$ и $r_{2c}^2 + r_{2s}^2$. В этом случае детектор называется *квадратичным*.

Двоичные сигналы ЧМ являются примером двоичных ортогональных сигналов. Напомним, что в двоичной ЧМ мы используем две различные частоты, скажем f_1 и $f_2 = f_1 + \Delta f$. Выбор минимального разноса частот $\Delta f = f_2 - f_1$ рассматривается ниже. Таким образом, эти сигналы можно выразить так:

$$\begin{aligned} s_1(t) &= \sqrt{2\mathscr{E}_b/T_b}\cos 2\pi f_1 t, \quad 0 \le t \le T_b, \\ s_2(t) &= \sqrt{2\mathscr{E}_b/T_b}\cos 2\pi f_2 t, \quad 0 \le t \le T_b, \end{aligned} \tag{5.4.22}$$

и их эквивалентные низкочастотные представления

$$\begin{aligned} s_{1l}(t) &= \sqrt{2\mathscr{E}_b/T_b}, \qquad 0 \le t \le T_b, \\ s_{2l}(t) &= \sqrt{2\mathscr{E}_b/T_b}\,e^{j2\pi\Delta f t}, \quad 0 \le t \le T_b. \end{aligned} \tag{5.4.23}$$

Принимаемый сигнал можно записать так:

$$r(t) = \sqrt{2\mathscr{E}_b/T_b}\cos\left(2\pi f_m t + \phi_m\right) + n(t), \tag{5.4.24}$$

где ϕ_m – фаза частоты несущей f_m. Демодуляцию вещественного сигнала можно выполнить как показано на рис. 5.4.3, посредством четырех корреляторов с базисными функциями:

$$\begin{aligned} f_{1m}(t) &= \sqrt{2/T_b}\cos\left[\left(2\pi f_1 + 2\pi m\Delta f\right)t\right], \quad m = 0, 1, \\ f_{2m}(t) &= \sqrt{2/T_b}\sin\left[\left(2\pi f_1 + 2\pi m\Delta f\right)t\right], \quad m = 0, 1. \end{aligned} \tag{5.4.25}$$

Четыре выхода корреляторов стробируются в конце каждого сигнального интервала и поступают на детектор. Если передается m-й сигнал, четыре отсчёта у детектора можно выразить так:

$$\begin{aligned} r_{kc} &= \sqrt{E_b}\left[\frac{\sin\left[2\pi(k-m)\Delta f\,T\right]}{2\pi(k-m)\Delta f\,T}\cos\phi_m - \frac{\cos\left[2\pi(k-m)\Delta f\,T\right]-1}{2\pi(k-m)\Delta f\,T}\sin\phi_m\right] + n_{kc}, \quad k,m = 1, 2, \\ r_{ks} &= \sqrt{E_b}\left[\frac{\cos\left[2\pi(k-m)\Delta f\,T\right]-1}{2\pi(k-m)\Delta f\,T}\cos\phi_m + \frac{\sin\left[2\pi(k-m)\Delta f\,T\right]}{2\pi(k-m)\Delta f\,T}\sin\phi_m\right] + n_{ks}, \end{aligned} \tag{5.4.26}$$

где n_{kc} и n_{ks} – гауссовские шумовые компоненты в выходных отсчётах.

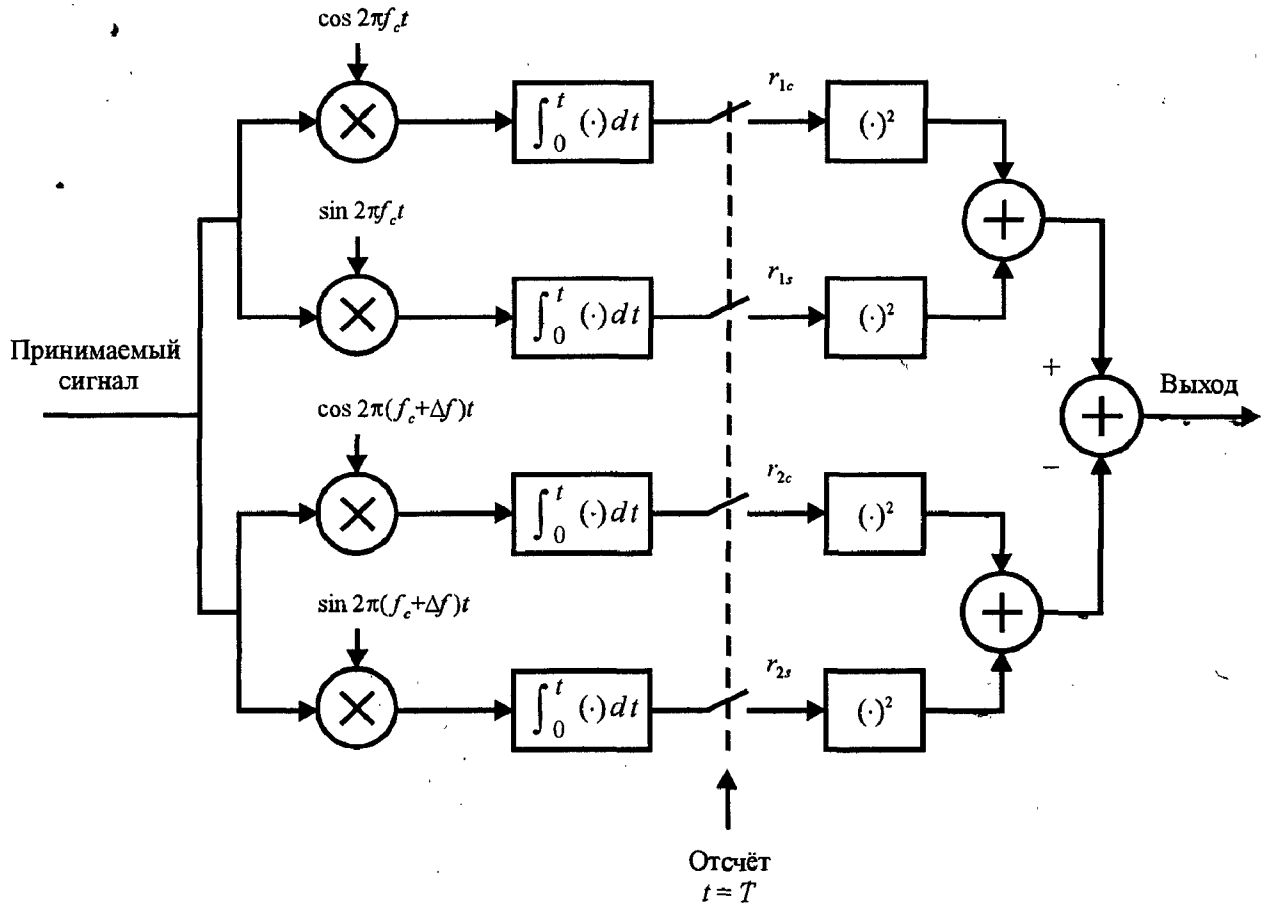


Рис. 5.4.3. Демодуляция и квадратичное детектирование двоичных сигналов ЧМ

Видим, что, когда $k = m$; отсчётные значения детектора равны

$$r_{mc} = \sqrt{\mathscr{E}_b}\cos\phi_m + n_{mc},$$
$$r_{ms} = \sqrt{\mathscr{E}_b}\sin\phi_m + n_{ms}. \qquad (5.4.27)$$

Далее видим, что, если $k \neq m$, сигнальные компоненты в отсчётах r_{kc} и r_{ks} исчезают, независимо от величин сдвигов фаз ϕ_k, обеспечивая разделение частот при $\Delta f = 1/T$. В этом случае два других выхода корреляторов состоят только из шума, т.е.

$$r_{kc} = n_{kc}, \quad r_{ks} = n_{ks}, \quad k \neq m. \qquad (5.4.28)$$

С частотным разносом $\Delta f = 1/T$ отношения (5.4.27) и (5.4.28) согласуются с прежним результатом (5.4.25) для выходов демодулятора. Следовательно, делаем заключение, что при детектировании огибающей или при квадратичном детектировании сигналов ЧМ минимальный разнос частот, требуемый для ортогональности сигналов, равен $\Delta f = 1/T$. Этот разнос вдвое больше, чем в случае когерентного детектирования.

5.4.2. Оптимальный приёмник для *M*-позиционных ортогональных сигналов

Обобщение оптимального демодулятора и детектора на случай M-позиционных ортогональных сигналов очевидно. Положим, что сигналы равновероятны и с одинаковой энергией, и представим их так:

$$s_m(t) = \mathrm{Re}\left[s_{lm}(t)e^{j2\pi f_c t}\right], \quad m = 1,2,\ldots,M, \quad 0 \le t \le T, \qquad (5.4.29)$$

где $s_{lm}(t)$ – эквивалентные низкочастотные сигналы. Оптимальный демодулятор по корреляционной схеме или с согласованными фильтрами образует M комплексных случайных величин:

$$r_m = r_{mc} + jr_{ms} = \int_0^T r_l(t) s_{lm}^*(t) dt, \quad m = 1,2,\ldots,M, \qquad (5.4.30)$$

где $r_l(t)$ – эквивалентный низкочастотный принимаемый сигнал. Тогда оптимальный детектор с учетом равномерно распределенной фазы несущей вычисляет M огибающих:

$$|r_m| = \sqrt{r_{mc}^2 + r_{ms}^2}, \quad m = 1,2,\ldots,M, \qquad (5.4.31)$$

или, что эквивалентно, квадраты огибающих $|r_m|^2$ и выбирает сигнал с наибольшей огибающей.

В частном случае M-позиционных ортогональных сигналов ЧМ оптимальный приёмник имеет структуру, показанную на рис. 5.4.4.

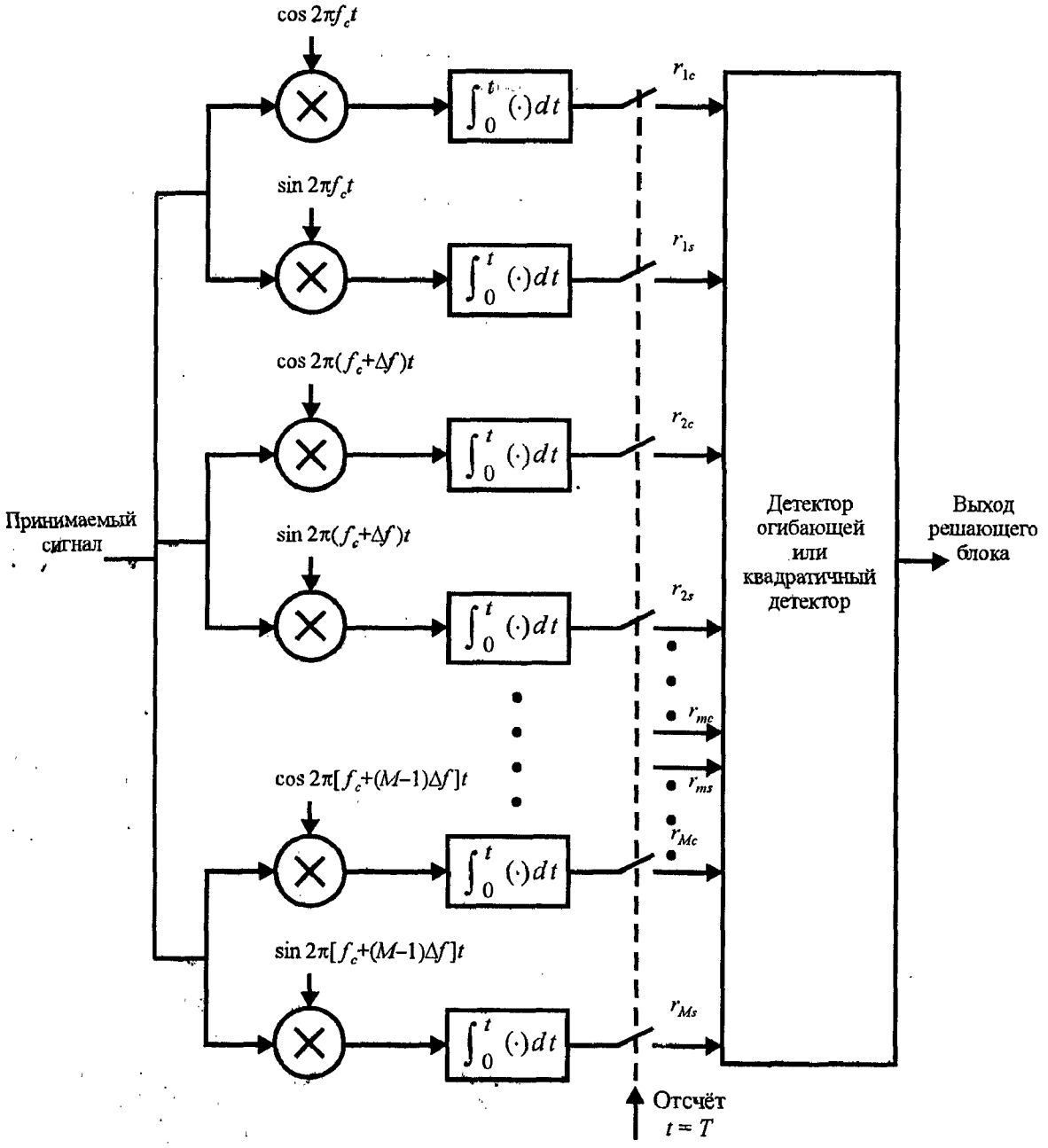


Рис. 5.4.4. Демодуляция M-ичных сигналов ЧМ при некогерентном детектировании

Здесь имеются $2M$ корреляторов: два для каждой из возможных к передаче частот. Минимальный разнос между соседними частотами для обеспечения ортогональности равен $\Delta f = 1/T$.

5.4.3. Вероятность ошибки при детектировании огибающей для M-позиционных ортогональных[1] сигналов

Рассмотрим передачу M ортогональных сигналов равной энергии по каналу с АБГШ с детектированием огибающей в приёмнике. Мы также предположим, что M сигналов априорно равновероятны и что сигнал $s_1(t)$ передается на интервал $0 \le t \le T$. В качестве M решающих метрик на входе детектора используются M огибающих:

$$|r_m| = \sqrt{r_{mc}^2 + r_{ms}^2}, \quad m = 1,2,\ldots,M,$$

где (5.4.32)

$$r_{1c} = \sqrt{\mathscr{E}_b}\cos\phi_1 + n_{1c}, \qquad (5.4.33)$$

$$r_{1s} = \sqrt{\mathscr{E}_b}\sin\phi_1 + n_{1s},$$

$$r_{mc} = n_{mc}, \quad r_{ms} = n_{ms}, \quad m = 2,3,\ldots,M. \qquad (5.4.34)$$

Компоненты аддитивного шума $\{n_{mc}\}$ и $\{n_{ms}\}$ – взаимно статистически независимые гауссовские величины с нулевым средним и одинаковой дисперсией $\sigma^2 = \frac{1}{2}N_0$. Таким образом, ФПВ для случайных величин на входе детектора равны

$$p_{r_1}(r_{1c}, r_{1s}) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{r_{1c}^2 + r_{1s}^2 + \mathscr{E}^2}{2\sigma^2}\right) I_0\left(\frac{\sqrt{\mathscr{E}_s\left(r_{2c}^2 + r_{2s}^2\right)}}{\sigma^2}\right), \qquad (5.4.35)$$

$$p_{r_m}(r_{mc}, r_{ms}) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{r_{mc}^2 + r_{ms}^2}{2\sigma^2}\right), \qquad m = 2,3,\ldots,M. \qquad (5.4.36)$$

Сделаем замену переменных в совместных ФПВ, определяемых (5.4.35) и (5.4.36). Определим нормированные величины

$$R_m = \frac{\sqrt{r_{mc}^2 + r_{ms}^2}}{\sigma}; \quad \Theta_m = \operatorname{arctg}\frac{r_{ms}}{r_{mc}}. \qquad (5.4.37)$$

Ясно, что $r_{mc} = \sigma R_m \cos\Theta_m$ и $r_{ms} = \sigma R_m \sin\Theta_m$. Якобиан этого преобразования

$$|\mathbf{J}| = \begin{vmatrix} \sigma\cos\Theta_m & \sigma\sin\Theta_m \\ -\sigma R_m \sin\Theta_m & \sigma R_m \cos\Theta_m \end{vmatrix} = \sigma^2 R_m. \qquad (5.4.38)$$

Следовательно,

$$p(R_1, \Theta_1) = \frac{R_1}{2\pi}\exp\left[-\frac{1}{2}\left(R_1^2 + \frac{2\mathscr{E}_s}{N_0}\right)\right] I_0\left(\sqrt{\frac{2\mathscr{E}_s}{N_0}}R_1\right), \qquad (5.4.39)$$

$$p(R_m, \Theta_m) = \frac{R_m}{2\pi}\exp\left(-\tfrac{1}{2}R_m^2\right), \quad m = 2,3,\ldots,M. \qquad (5.4.40)$$

В заключение при усреднении $p(R_m, \Theta_m)$ по Θ_m множитель 2π исключается из (5.4.39) и (5.4.40). Таким образом находим, что R_1 имеет распределение Райса, а R_m, $m = 2,3,\ldots,M$, имеет распределение Релея.

Вероятность правильного решения определяется вероятностью того, что $R_1 > R_2$ и $R_1 > R_3$ ... и $R_1 > R_M$. Следовательно,

---

[1] Для полосовых сигналов имеется в виду ортогональность в усиленном смысле (прп).

$$P_c = P\left(R_2 < R_1, R_3 < R_1, \ldots, R_M < R_1\right) = \\ = \int_0^\infty P\left(R_2 < R_1, R_3 < R_1, \ldots, R_M < R_1 \middle| R_1 = x\right) p_{R_1}(x)dx. \tag{5.4.41}$$

Поскольку случайные величины R_m, $m = 2, 3, \ldots, M.$, статистически независимы, совместная вероятность в (5.4.41) определяется произведением $M-1$ идентичных членов:

$$P_c = \int_0^\infty \left[P\left(R_2 < R_1 \middle| R_1 = x\right)\right]^{M-1} p_{R_1}(x)dx, \tag{5.4.42}$$

где

$$P\left(R_2 < R_1 \middle| R_1 = x\right) = \int_0^x p_{R_2}\left(r_2\right)dr_2 = 1 - \mathrm{e}^{-x^2/2}. \tag{5.4.43}$$

Степень $M-1$ от (5.4.43) можно выразить, пользуясь биномом Ньютона, так:

$$\left(1 - \mathrm{e}^{-x^2/2}\right)^{M-1} = \sum_{n=0}^{M-1} (-1)^n \binom{M-1}{n} \mathrm{e}^{-nx^2/2} \tag{5.4.44}$$

Подставив этот результат в (5.4.42) и интегрируя по x, получаем для вероятности правильного решения

$$P_c = \sum_{n=0}^{M-1} (-1)^n \binom{M-1}{n} \frac{1}{n+1} \exp\left[-\frac{n\mathscr{E}_s}{(n+1)N_0} \right], \tag{5.4.45}$$

где $\mathscr{E}_s/N_0$ – ОСШ на символ. Тогда вероятность ошибки на символ $P_M = 1 - P_c$ равна[1]

$$P_M = \sum_{n=1}^{M-1} (-1)^{n+1} \binom{M-1}{n} \frac{1}{n+1} \exp\left[-\frac{nk\mathscr{E}_b}{(n+1)N_0} \right], \tag{5.4.46}$$

где E_b/N_0 – ОСШ на бит.

Для двоичных ортогональных сигналов $(M = 2)$ (5.4.46) приводится к простому виду

$$P_2 = \tfrac{1}{2}\mathrm{e}^{-\mathscr{E}/2N_0}. \tag{5.4.47}$$

Для $M > 2$ мы можем вычислить вероятность ошибки на бит, используя соотношение

$$P_b = \frac{2^{k-1}}{2^k - 1} P_M, \tag{5.4.48}$$

которое было установлено в разд. 5.2. Рисунок 5.4.5 показывает вероятность ошибки на бит как функцию ОСШ на бит γ_b для $M = 2$, 4, 8, 16 и 32. Для случая когерентного детектирования M-позиционных ортогональных сигналов (см. разд. 5.2.2) мы видели, что для любой заданной вероятности ошибки ОСШ на бит уменьшается по мере роста M. В гл. 7 будет показано, что в пределе, когда $M \to \infty$ (или $k = \log_2 M \to \infty$), вероятность ошибки на бит P_b можно сделать сколь угодно малой при условии, что ОСШ на бит превышает предел Шеннона $-1{,}6$ дБ. Цена роста M – увеличение полосы, требуемой для передачи сигнала. Для M-позиционной ЧМ разнос частот между соседними частотами равен $\Delta f = 1/T$ для ортогональности сигналов. Полоса частот, требуемая для M сигналов, $W = M\Delta f = M/T$. Таким образом, битовая скорость $R = k/T$, где $k = \log_2 M$. Следовательно, отношение битовой скорости к полосе равно

$$\frac{R}{W} = \frac{\log_2 M}{M}. \tag{5.4.49}$$

[1] Этот результат впервые получил Л.М. Финк в 1957 г. [23] (прп).

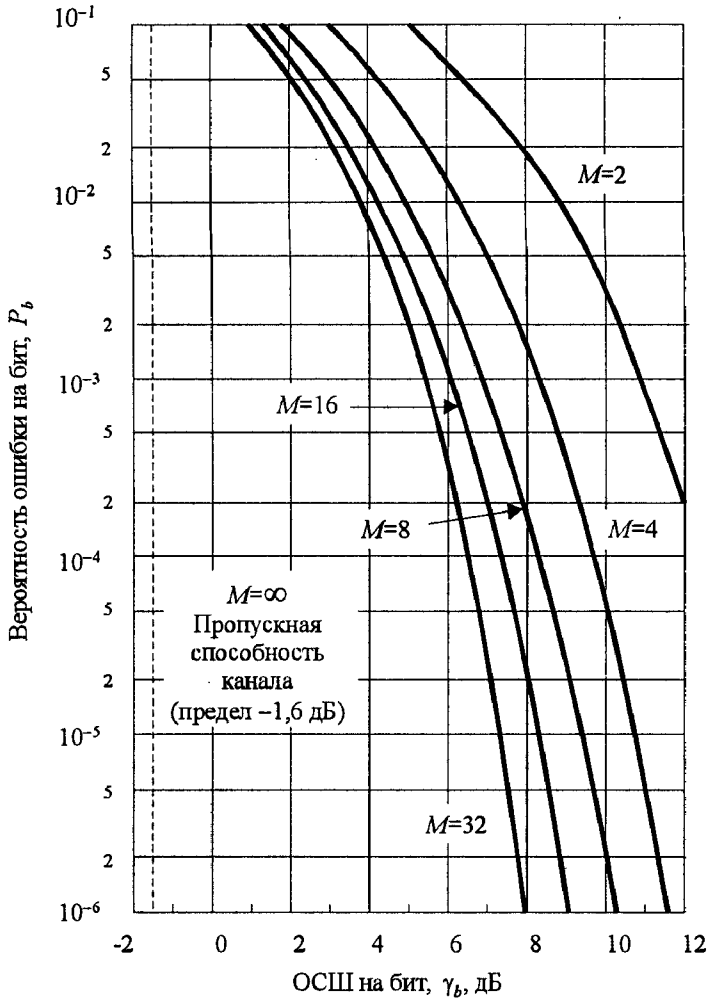


Рис. 5.4.5. Вероятность ошибки на бит для некогерентного детектирования ортогональных сигналов

5.4.4. Вероятность ошибки для коррелированных двоичных сигналов при детектировании огибающей

В этом разделе рассмотрим качество детектора огибающей для двоичных коррелированных сигналов равной энергии. Если два сигнала коррелированы, на вход детектора поступают комплексные величины, определенные (5.4.10). Предположим, что детектор выносит свои решения на основе огибающих $|r_1|$ и $|r_2|$, которые теперь коррелированы (статистически зависимы). Собственные ФПВ для $R_1 = |r_1|$ и $R_2 = |r_2|$ распределены по Райсу, и их можно выразить так:

$$p(R_m) = \begin{cases} \dfrac{R_m}{2\mathscr{E}N_0} \exp\left(-\dfrac{R_m^2 + \beta_m^2}{4\mathscr{E}N_0}\right) I_0\left(\dfrac{\beta_m R_m}{2\mathscr{E}N_0}\right) & (R_m > 0), \\ 0 & (R_m < 0), \end{cases} \tag{5.4.50}$$

$m = 1, 2$, где $\beta_1 = 2\mathscr{E}$ и $\beta_2 = 2\mathscr{E}|\rho|$ получены в предположении, что был передан сигнал $s_1(t)$.

Поскольку R_1 и R_2 статистически зависимы, как следствие неортогональности сигналов, вероятность ошибки можно найти, вычислив двойной интеграл

$$P_b = P(R_2 > R_1) = \int_0^\infty \int_{x_1}^\infty p(x_1, x_2) dx_1 dx_2, \tag{5.4.51}$$

где $p(x_1, x_2)$ – совместная ФПВ огибающих R_1 и R_2. Это исследование было впервые выполнено Хелстромом (1955), который определил совместную ФПВ R_1 и R_2 и рассчитал двойной интеграл в (5.4.51).

Альтернативный подход базируется на наблюдении, что вероятность ошибки можно выразить так:

$$P_b = P(R_2 > R_1) = P(R_2^2 > R_1^2) = P(R_2^2 - R_1^2 > 0). \qquad (5.4.52)$$

Но $R_2^2 - R_1^2$ – это частный случай общей квадратичной формы комплексных гауссовских случайных величин, рассмотренных в приложении В. В этом частном случае вероятность ошибки можно выразить в виде

$$P_b = Q_1(a,b) - \tfrac{1}{2} e^{-(a^2+b^2)/2} I_0(ab), \qquad (5.4.53)$$

где

$$a = \sqrt{\frac{E_b}{2N_0}\left(1 - \sqrt{1-|\rho|^2}\right)}, \quad b = \sqrt{\frac{E_b}{2N_0}\left(1 + \sqrt{1-|\rho|^2}\right)}, \qquad (5.4.54)$$

$Q_1(a,b)$ – это Q-функция Маркума, определенная (2.1.123), и $I_0(x)$ – модифицированная функция Бесселя нулевого порядка.

Вероятность ошибки P_b иллюстрируется на рис. 5.4.6

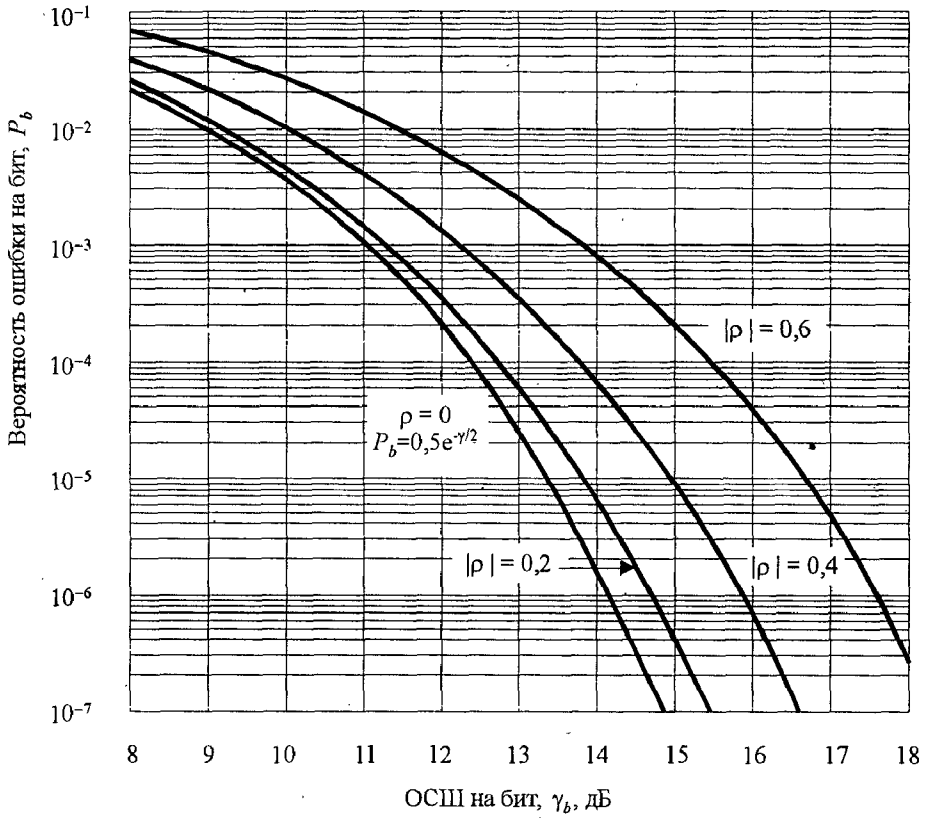


Рис. 5.4.6. Вероятность ошибки для некогерентного детектирования

для нескольких значений $|\rho|$. Вероятность P_b минимизируется, когда $\rho = 0$, т.е. когда сигналы ортогональны. В этом случае $a = 0$, $b = \sqrt{E_b / N_0}$, и (5.4.53) даёт

$$P_b = Q_1\left(0, \sqrt{\frac{\mathcal{E}_b}{N_0}}\right) - \tfrac{1}{2} e^{-\mathcal{E}_b/2N_0}. \qquad (5.4.55)$$

Из определения функции $Q_1(a,b)$ в (2.1.123) следует, что

$$Q_1\left(0,\sqrt{\frac{\mathscr{E}_b}{N_0}}\right)=e^{-\mathscr{E}_b/2N_0}.$$

Подстановка этих отношений в (5.4.55) дает желаемый результат, определенный ранее (5.4.47). С другой стороны, если $|\rho|=1$, вероятность ошибки в (5.4.53) получается, как и ожидалось, $P_b=\frac{1}{2}$.

5.5. РЕГЕНЕРАТИВНЫЕ ПОВТОРИТЕЛИ И АНАЛИЗ РЕСУРСОВ ЛИНИЙ СВЯЗИ

При передаче цифровых сигналов через канал с АБГШ мы наблюдали, что качество системы связи, измеряемое вероятностью ошибки, зависит исключительно от принимаемого ОСШ, $\mathscr{E}_b/N_0$, где $\mathscr{E}_b$ – передаваемая энергия на бит, и $\frac{1}{2}N_0$ – спектральная плотность мощности аддитивного шума. Таким образом, аддитивный шум ограничивает качество систем связи. В дополнение к аддитивному шуму другой фактор, который влияет на качество систем связи, это затухание в канале. Все физические каналы, включая проводные линии и радиоканалы, вносят потери. Следовательно, сигнал ослабляется при прохождении через канал. Для канала можно использовать простейшую математическую модель ослабления, показанную на рис. 5.5.1. Следовательно, если передан сигнал $s(t)$, то принимаемый сигнал при $0<\alpha\le 1$ равен

$$r(t)=\alpha s(t)+n(t). \qquad (5.5.1)$$

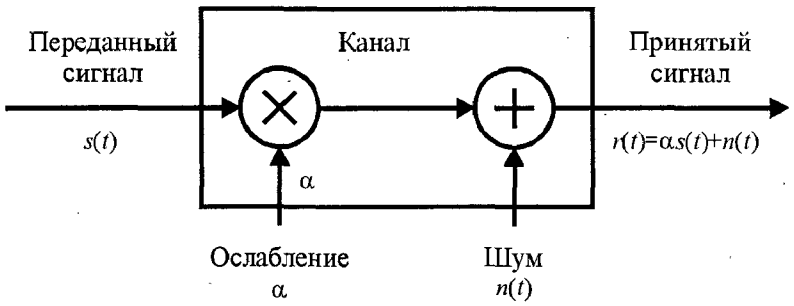


Рис. 5.5.1. Математическая модель канала с ослаблением и аддитивным шумом

Затем если энергия переданного сигнала $\mathscr{E}_b$, то энергия принятого сигнала $\alpha^2\mathscr{E}_b$. Следовательно, принимаемый сигнал имеет ОСШ $\alpha^2\mathscr{E}_b/N_0$. Таким образом, влияние ослабления сигнала сводится к уменьшению энергии принимаемого сигнала, и, следовательно, система связи делается более уязвимой к аддитивному шуму.

В аналоговых системах связи усилители, называемые повторителями, используются для периодического повышения уровня сигнала, передаваемого по каналу. Однако каждый усилитель также вносит шум в системе. В противоположность этому цифровые системы связи позволяют нам детектировать и регенерировать чистый (свободный от шума) сигнал в канале передачи. Такие устройства, называемые *регенеративными повторителями*, широко используются в проводных и волоконно-оптических каналах связи.

5.5.1. Регенеративные повторители

Любой регенеративный повторитель содержит демодулятор/детектор, который демодулирует и детектирует переданную цифровую информационную последовательность, отправленную предыдущим повторителем. С выхода детектора последовательность

отправляется на передатчик повторителя, который отображает её в сигнал, который посылается на следующий повторитель. Этот вид повторителя называется регенеративным.

Поскольку каждый повторитель регенерирует сигнал, свободный от шума, аддитивный шум не накапливается. Однако, когда ошибки появляются в детекторе повторителя, эти ошибки распространяются дальше к следующим повторителям канала. Для расчета влияний ошибок на качество всей системы предположим, что имеется двоичная АМ, так что вероятность ошибки на бит для одного участка (передающего сигнал от одного повторителя к следующему в цепочке) равна

$$P_b = Q\left(\sqrt{\frac{2\mathscr{E}_b}{N_0}}\right).$$

Поскольку ошибки возникают с малыми вероятностями, мы можем игнорировать вероятность того, что произвольный символ (бит) будет продетектирован неправильно более одного раза при передаче по каналу с K повторителями. Следовательно, число ошибок будет увеличиваться линейно с числом регенеративных повторителей в канале, и суммарную вероятность ошибки на бит можно выразить так:

$$P_b \approx KQ\left(\sqrt{\frac{2\mathscr{E}_b}{N_0}}\right). \qquad (5.5.2)$$

В противоположность этому использование Q аналоговых повторителей в канале уменьшает принимаемое значение ОСШ в K раз, поэтому вероятность ошибки на бит

$$P_b \approx Q\left(\sqrt{\frac{2\mathscr{E}_b}{KN_0}}\right). \qquad (5.5.3)$$

Ясно, что для той же вероятности ошибки использование регенеративных повторителей даёт существенную экономию передаваемой мощности по сравнению с аналоговыми повторителями. Поэтому в цифровых системах связи регенеративные повторители предпочтительны. Однако в проводных телефонных каналах, которые используются для передачи аналоговых и цифровых сигналов, обычно используются аналоговые повторители.

Пример 5.5.1. Двоичная система связи передает данные по проводному каналу длиной 1000 км. Повторители используются каждые 10 км, чтобы снять влияние канального ослабления. Определим требуемое значение $\mathscr{E}_b/N_0$ для достижения вероятности ошибки на бит $P_b = 10^{-5}$, если:

a) используются аналоговые повторители;

b) используются регенеративные повторители.

Число повторителей, используемых в системе, равно K=600. Если используются регенеративные повторители, $\mathscr{E}_b/N_0$ получим из (5.5.2):

$$10^{-5} \approx 100Q\left(\sqrt{\frac{2\mathscr{E}_b}{N_0}}\right); \qquad 10^{-7} \approx Q\left(\sqrt{\frac{2\mathscr{E}_b}{N_0}}\right),$$

что дает ОСШ=11,3 дБ. Если используются аналоговые повторители, $\mathscr{E}_b/N_0$ получим из (5.5.3):

$$10^{-5} = Q\left(\sqrt{\frac{2\mathscr{E}_b}{100N_0}}\right),$$

что дает $\mathcal{E}_b/N_0 \approx 29{,}6$ дБ. Следовательно, различие в требуемых значениях ОСШ составляет около 18,3 дБ. Это означает, что для получения одной и той же величины вероятности ошибки при переходе от регенеративных (b) к аналоговым (a) повторителям примерно в 70 раз должна возрастать мощность передатчика системы цифровой связи.

5.5.2. Анализ ресурсов линий связи

При расчёте систем радиосвязи, которые передают сигнал в зоне прямой видимости по микроволновым и спутниковым каналам, проектировщик системы должен особенно учесть размеры антенн передатчика и приёмника, передаваемую мощность и требуемое ОСШ для достижения нужного уровня качества при требуемой скорости передачи данных.

Расчет системы относительно прост, и он приводится ниже.

Начнём с передающей антенны, которая излучает изотропно в открытое пространство уровень мощности P_T, как показано на рис. 5.5.2. Плотность мощности на расстоянии d от антенны равна $P_T/4\pi d^2$ (Вт/м).

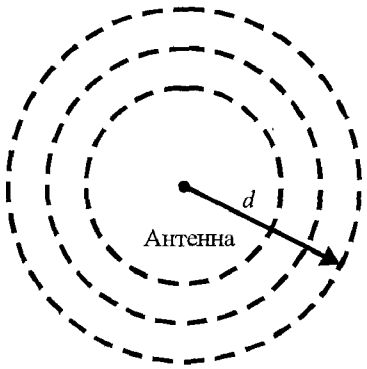


Рис. 5.5.2. Изотропно-излучающая антенна

Если передающая антенна имеет избирательность в определенном направлении, плотность мощности в этом направлении увеличивается. Коэффициент увеличения называют усилением антенны и обозначают G_T. В этом случае плотность мощности на расстоянии d равна $P_T G_T/4\pi d^2$ (Вт/м^2). Произведение $P_T G_T$ обычно называют *эффективной мощностью излучения* (ЭМИ, ERP или EIRP), которая является по существу мощностью излучения относительно изотропной антенны, для которой $G_T = 1$.

Приёмная антенна, ориентированная в направлении излучённой мощности, собирает долю мощности, которая пропорциональна площади её поперечного сечения.

Таким образом, мощность, принимаемую антенной, можно выразить так:

$$P_R = \frac{P_T G_T A_R}{4\pi d^2}, \qquad (5.5.4)$$

где A_R – эффективная площадь антенны. Из теории электрического поля получаем базовое соотношение между усилением G_R приёмной антенны и её эффективной площадью:

$$A_R = \frac{G_R \lambda^2}{4\pi}, \qquad (5.5.5)$$

где $\lambda = c/f$ – длина волны переданного сигнала, c – скорость света ($3\cdot10^8$ м/с), f – частота переданного сигнала. Если подставить (5.5.5) для A_R в (5.5.4), получим выражение для принимаемой мощности в виде

$$P_R = \frac{P_T G_T G_R}{(4\pi d/\lambda)^2}. \qquad (5.5.6)$$

Множитель

$$L_s = \left(\frac{\lambda}{4\pi d}\right)^2 \qquad (5.5.7)$$

называют *потерями в свободном пространстве*. Если при передаче сигнала встречаются другие потери, такие как потери в атмосфере, то их можно учесть путем введения дополнительного множителя потерь, скажем L_a. Таким образом, принимаемую мощность можно в итоге записать так:

$$P_R = P_T G_T G_R L_s L_a. \qquad (5.5.8)$$

Как отмечено выше, важнейшие характеристики антенны – её усиление и её эффективная площадь. Они обычно зависят от длины волны излученной мощности и физических размеров антенны. Например, параболическая антенна с диаметром D имеет эффективную площадь

$$A_R = \tfrac{1}{4}\pi D^2 \eta, \qquad (5.5.9)$$

где $\tfrac{1}{4}\pi D^2$ – физическая площадь, а η – *показатель эффективности облучения*, который находится в области $0{,}5 \le \eta \le 0{,}6$. Следовательно, усиление параболической антенны диаметром D равно

$$G_R = \eta\left(\frac{\pi D^2}{\lambda}\right)^2. \qquad (5.5.10)$$

В качестве второго примера возьмем рупорную антенну площадью A. Она имеет показатель эффективности 0,8, эффективную площадь $A_R = 0{,}8A$, и усиление антенны равно

$$G_R = \frac{10A}{\lambda^2}. \qquad (5.5.11)$$

Другой параметр, который связан с усилением антенны (направленностью) – это ширина луча, которую мы обозначим Θ_B. Он иллюстрируется графически на рис. 5.5.3.

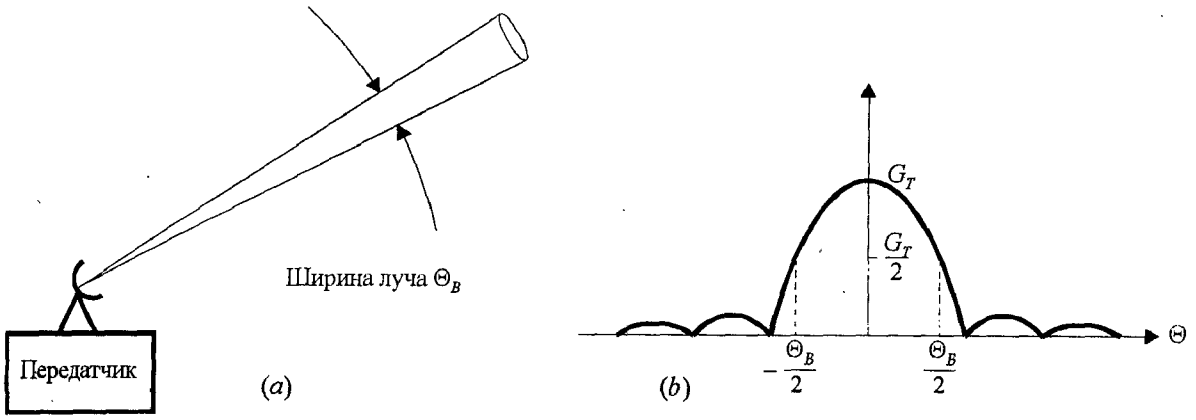


Рис. 5.5.3. Ширина луча антенны (*a*) и диаграмма направленности (*b*)

Обычно ширина луча измеряется по ширине диаграммы направленности на уровне – 3 дБ от её пика. Например, ширина луча параболической антенны на уровне – 3 дБ приближенно равна

$$\Theta_B = 70(\lambda/D)^\circ, \qquad (5.5.12)$$

так что G_T обратно пропорционально Θ_B^2. Это значит, что уменьшение ширины луча вдвое, получаемое удвоением диаметра D, увеличивает коэффициент усиления антенны примерно в 4 раза (на 6 дБ).

Основываясь на общих соотношениях для мощности принимаемого сигнала, определяемой (5.5.8), разработчик антенны может рассчитать P_R, исходя из данных усиления антенны и расстояния между передатчиком и приёмником. Такие расчеты обычно выполняются по мощности так:

$$(P_R)_{\text{дБ}} = (P_T)_{\text{дБ}} + (G_T)_{\text{дБ}} + (G_R)_{\text{дБ}} + (L_s)_{\text{дБ}} + (L_a)_{\text{дБ}}. \qquad (5.5.13)$$

Пример 5.5.2. Предположим, что имеется спутник на геостационарной орбите (36000 км над поверхностью земли), который излучает мощность W=100 Вт, т.е. 20 дБ относительно 1 Вт (20 дБВт). Передающая антенна имеет усиление 17 дБ, так что ЭМИ=37 дБВт. Также предположим, что наземная станция использует 3-метровую параболическую антенну и что линия вниз работает на частоте 4 ГГц. Коэффициент эффективности $\eta = 0{,}5$. Путем подстановки этих значений в (5.5.10) получим величину антенного усиления 39 дБ. Потери в свободном пространстве $L_s = 195{,}6$дБ. Никакие другие потери не учитываются. Следовательно, мощность принимаемого сигнала

$$(P_R)_{\text{дБ}} = 20 + 17 + 39 - 195{,}6 = -119{,}6 \text{ дБВт}$$

или, что эквивалентно,

$$P_R = 1{,}1 \cdot 10^{-12} \text{ Вт}.$$

Чтобы закончить расчёт ресурсов линии, мы должны также рассмотреть влияние аддитивного шума на приёмной стороне. Тепловой шум, который возникает в приёмнике и имеет примерно одинаковую спектральную плотность мощности вплоть до частот 10^{12} Гц, равную

$$N_0 = k_b T_0 \text{ Вт/Гц}, \qquad (5.5.14)$$

где k_b – постоянная Больцмана ($1{,}38 \cdot 10^{-23}$ Вт·с/К), а T_0 – шумовая температура в Кельвинах. Следовательно, суммарная мощность шума в полосе сигнала W равна $N_0 W$.

Качество цифровых систем связи определяется величиной E_b/N_0, требуемой для обеспечения вероятности ошибки ниже заданной величины.

Поскольку

$$\frac{\mathscr{E}_b}{N_0} = \frac{T_b P_R}{N_0} = \frac{1}{R}\frac{P_R}{N_0},$$

то следует (5.5.15)

$$\frac{P_R}{N_0} = R\left(\frac{\mathscr{E}_b}{N_0}\right)_{\text{треб}}, \qquad (5.5.16)$$

где $(\mathscr{E}_b/N_0)_{\text{треб}}$ – требуемое значение ОСШ на бит. Если имеем P_R/N_0 и требуемое ОСШ на бит, можем определить максимально возможную скорость передачи данных.

Пример 5.5.3. Для линии, рассмотренной в примере 5.5.2, мощность принимаемого сигнала $P_R = 1{,}1 \times 10^{-12}$ Вт $(-119{,}6$ дБВт$)$.

Теперь предположим, что на приёмной стороне шумовая температура 300 К, что типично для приёмника в полосе 4 ГГц. Тогда

$$N_0 = 4{,}1 \cdot 10^{-21} \text{ Вт/Гц}$$

или, что эквивалентно, $-203{,}9$ дБ Вт/Гц. Следовательно,

$$\frac{P_R}{N_0} = -119{,}6 + 203{,}9 = 84{,}3 \text{ дБ Вт/Гц}.$$

Если требуемое ОСШ на бит 10 дБ, тогда из (5.5.16) получаем допустимую скорость

$$R_{\text{дБ}} = 84{,}3 - 10 = 74{,}3 \text{ дБ}. \text{ (с расчетом на 1 бит/с).}$$

Это соответствует скорости 26,9 Мбит/с, что эквивалентно около 420 каналам ИКМ, каждый со скоростью 64 000 бит/с.

Целесообразно ввести некоторый резерв безопасности, который будем называть *резервом линии* $M_{\text{дБ}}$ в проведённых выше расчетах пропускной способности коммуникационных линий. Типично его можно выбрать $M_{\text{дБ}} = 6$ дБ. Тогда расчёт ресурса линии для её пропускной способности можно выразить в простейшем виде так:

$$\begin{aligned} R_{\text{дБ}} &= \left(\frac{P_R}{N_0}\right)_{\text{дБ Гц}} - \left(\frac{\mathscr{E}_b}{N_0}\right)_{\text{треб}} - M_{\text{дБ}} = \\ &= \left(P_T\right)_{\text{дБ Вт}} + \left(G_T\right)_{\text{дБ}} + \left(G_R\right)_{\text{дБ}} + \left(L_a\right)_{\text{дБ}} + \left(L_s\right)_{\text{дБ}} - \left(\frac{\mathscr{E}_b}{N_0}\right)_{\text{треб}} - M_{\text{дБ}} \end{aligned} \qquad (5.5.17)$$

5.6. БИБЛИОГРАФИЧЕСКИЕ ЗАМЕЧАНИЯ И ССЫЛКИ

При разработке оптимального демодулятора для сигналов, подверженных воздействию АБГШ, мы использовали подход, который был первоначально использован при разработке оптимальных приёмников радиолокационных сигналов. Например, согласованный фильтр был впервые предложен Нортом (1943) для детектирования радиолокационных сигналов, и его иногда называют фильтром Норта. Альтернативным методом для получения оптимального демодулятора и детектора является разложение Карунена–Лоэва, которое описано в классических работах Давенпорта и Рута (1968), Хелстрома (1968) и Ван Триса (1968). Их использование в теории детектирования радиолокационных сигналов описано в публикациях Келли и др. (1960). Эти методы детектирования базируются на методах проверки гипотез, разработанных статистиками, например, Нейманом и Пирсоном (1933) и Вальдом (1947).

Геометрические подходы к синтезу сигналов и детектированию, которые были представлены при рассмотрении цифровой модуляции, имеют свои корни в оригинальных работах Шеннона. Они концептуально привлекательны и теперь широко используются, начиная с книги Возенкрафта и Джекобса (1965).

Расчёт и анализ сигнальных созвездий для каналов с АБГШ привлекали значительное внимание в технической литературе. В частности, исчерпывающим является анализ качества двухмерных (КАМ) сигнальных созвездий, которые рассматривались в публикациях Кана (1960), Ханкока и Лакки (1962), Кампопиано и Глэйзера (1962), Лакки и Ханкока (1962), Сальца и др. (1971), Саймона и Смита (1973), Томаса и др. (1974), Фошини и др. (1974). Синтез сигналов, основанный на использовании многомерных сигнальных созвездий, был описан и проанализирован в работах Гершо и Лоуренса (1984).

Алгоритм Витерби был разработан его автором (1967) для целей декодирования свёрточных кодов. Его использование, как оптимального алгоритма для максимально-правдоподобного последовательного детектирования сигналов с памятью, было описано Форни (1972) и Омура (1971). Его использование для обработки сигналов с модуляцией

несущей было рассмотрено Унгербоеком (1974) и МакКеншни (1979). Он был впоследствии использован для демодуляции МНФ Аулином и Сандбергом (1981) и др.

ЗАДАЧИ

5.1. Согласованный фильтр имеет частотную характеристику

$$H(f)=\frac{1-e^{-j2\pi fT}}{j2\pi fT}.$$

a) Определите импульсную характеристику $h(t)$, соответствующую $H(f)$.

b) Определите сигналы, с которыми фильтр согласован.

5.2. Рассмотрите сигнал

$$s(t)=\begin{cases}(A/T)t\cos 2\pi f_c t & (0\le t\le T),\\ 0 & (\text{для других } T).\end{cases}$$

a) Определите импульсную характеристику фильтра, согласованного с этим сигналом

b) Определите выход согласованного фильтра при $t=T$

c) Допустим, что сигнал $s(t)$ прошёл через автокоррелятор. Определите сигнал на выходе автокоррелятора при $t=T$. Сравните ваш результат с результатом (b).

5.3. Эта задача касается характеристик сигналов ДФМ.

a) Предположим, что мы хотим передать последовательность данных

110100010110

посредством двоичной ДФМ. Пусть $s(t)=A\cos(2\pi f_c t+\theta)$ представляет переданный сигнал на некотором сигнальном интервале длительностью T. Определите фазу переданного сигнала для последовательности данных. Начните с $\theta=0$ для передачи фазы первого бита.

b) Если последовательность данных не коррелирована, определите и нарисуйте спектральную плотность мощности переданного сигнала ДФМ.

5.4. Для передачи информации двоичная цифровая система связи использует сигналы

$$s_0(t)=0, 0\le t\le T,$$
$$s_1(t)=A, 0\le t\le T.$$

Демодулятор выдает результат взаимной корреляции принимаемого сигнала $r(t)$ и $s_i(t)$ и стробирует выход коррелятора в момент $t=T$.

a) Определите оптимальный детектор для канала с АБГШ и оптимальный порог, предполагая, что сигналы равновероятны.

b) Определите вероятность ошибки как функцию ОСШ. Насколько ОСШ надо поднять по сравнению с противоположными сигналами при одинаковой вероятности ошибки.

5.5. Метрики корреляции, даваемые (5.1.44), равны

$$C(\mathbf{r},\mathbf{s}_m)=2\sum_{n=1}^{N}r_n s_{mn}-\sum_{n=1}^{N}s_{mn}^2, m=1,2,\ldots M,$$

где

$$r_n=\int_0^T r(t)f_n(t)dt,$$

$$s_{mn}=\int_0^T s_m(t)f_n(t)dt.$$

Покажите, что эти корреляционные метрики эквивалентны метрикам

$$C(\mathbf{r},\mathbf{s}_m)=2\int_0^T r(t)s_m(t)dt-\int_0^T s_m^2(t)dt.$$

5.6. Рассмотрите эквивалентный низкочастотный сигнал $s_l(t)$, $0\le t\le T$, с энергией $\mathcal{E}=\frac{1}{2}\int_0^T |s_l(t)|^2 dt$.

Предположите, что этот сигнал искажается в канале с АБГШ, который представлен эквивалентным низкочастотным шумом $z(t)$. Таким образом, наблюдаемый сигнал равен

$$r_l(t)=s_l(t)+z(t), 0\le t\le T.$$

Принимаемый сигнал проходит через фильтр, который имеет эквивалентную низкочастотную импульсную характеристику $h_l(t)$.

Определите $h_l(t)$ так, чтобы фильтр максимизировал ОСШ на своем выходе (в момент $t = T$).

5.7. Пусть $z(t) = x(t) + jy(t)$ является комплексным белым гауссовским шумовым процессом с нулевым средним и функцией автокорреляции $\phi_{zz}(t) = N_0\delta(t)$ Пусть $f_m(t), m = 1,2,...M$, является ансамблем M ортогональных эквивалентных низкочастотных сигналов, определенных на интервале $0 \le t \le T$. Найдите

$$N_{mr} = Re\left[\int_0^T z(t) f_m^*(t) dt\right], m = 1,2,...M.$$

a) Определите дисперсию для N_{mr}.

b) Докажите, что $E(N_{mr} N_{kr}) = 0$ для $k \neq m$.

5.8. Два эквивалентиых низкочастотных сигиала, показанных на рис. Р5.8, используются для передачи двоичной последовательности по каналу с АБГШ.

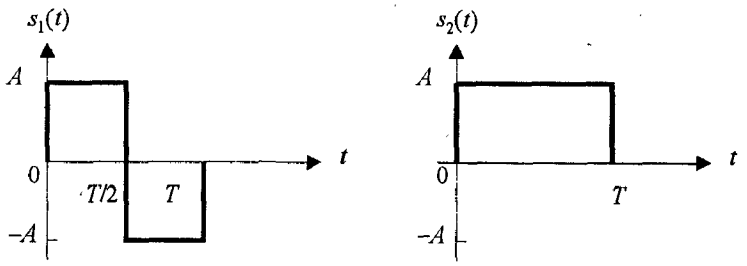


Рис. Р5.8

Принимаемый сигнал можно выразить так:

$$r_i(t) = s_i(t) + z(t), 0 \le t \le T, i = 1,2,$$

где $z(t)$ – гауссовский процесс с нулевым средним и автокорреляционной функцией

$$\phi_{zz}(\tau) = \tfrac{1}{2} E\left[z^*(t) z(t+\tau)\right] = N_0\delta(\tau).$$

a) Определите переданную энергию посредством $s_1(t)$ н $s_2(t)$ и коэффициент взаимной корреляции этих сигналов ρ_{12}.

b) Предположим, что приёмник построен как когерентный детектор с использованием двух согласованных фильтров: один согласован с $s_1(t)$, другой – с $s_2(t)$. Нарисуйте эквивалентную низкочастотную импульсную характеристику согласованного фильтра.

c) Нарисуйте свободные от шума отклики двух согласованных фильтров, если передаётся сигнал $s_2(t)$.

d) Предположим, что приёмник состоит из двух взаимных корреляторов (умножителей, за которыми следуют интеграторы), работающих параллельно. Нарисуйте выход каждого интегратора как функцию времени на интервале $0 \le t \le T$, когда передаётся сигнал $s_2(t)$.

e) Сравните рисунки, полученные в пп. (c) и (d). Объясните результат.

f) Зная характеристики сигналов, дайте выражение для вероятности ошибки для этих двоичных систем связи.

5.9. Допустим, что имеется комплексная гауссовская случайная величина $z = x + jy$, где (x, y) – статистически независимые величины с нулевыми средиими и дисперсией $E(x^2) = E(y^2) = \sigma^2$. Пусть

$$r = z + m, \text{ где } m = m_r + jm_i,$$

и определите r так:

$$r = a + jb.$$

Ясно, что $a = x + m_r$ и $b = y + m_i$. Определите следующие ФПВ:

a) $p(a,b)$,

b) $p(u,\phi)$, где $u = \sqrt{a^2 + b^2}$, а $\phi = \text{arctg}(b/a)$,

c) $p(u)$.

Замечание: в (b) удобно определить $\phi = \text{arctg}(m_i/m_r)$, так что

$$m_r = \sqrt{m_r^2 + m_i^2}\cos\theta, m_i = \sqrt{m_r^2 + m_i^2}\sin\theta.$$

Далее следует использовать соотношение

$$\frac{1}{2\pi}\int_0^{2\pi} e^{\alpha\cos(\phi-\theta)}d\phi = I_0(\alpha) = \sum_{n=0}^{\infty}\frac{\alpha^{2n}}{2^{2n}(n!)^2},$$

где $I_0(\alpha)$ – модифицированная функция Бесселя нулевого порядка.

5.10. Троичная система связи передает один из трех сигналов $s(t)$, 0 или $-s(t)$ каждые T секунд. Принимаемый сигнал равен $r_1(t)=s(t)+z(t)$, $r_1(t)=z(t)$ или $r_1(t)=-s(t)+z(t)$, где $z(t)$ – белый гауссовский шум с $E[z(t)]=0$ и $\phi_{zz}(\tau)=\frac{1}{2}E[z^*(t)z(t+\tau)]=N_0\delta(\tau)$. Оптимальный приёмник вычисляет корреляционные метрики

$$U = Re\left[\int_0^T r(t)s^*(t)dt\right]$$

и сравнивает U с порогами A и $-A$. Если $U>A$, принимается решение, что передан $s(t)$. Если $U<-A$, принимается решение в пользу $-s(t)$. Если $-A<U<A$, то принимается решение в пользу нуля.

a) Определите три условных вероятностях ошибки: P_1 при условии, что передан $s(t)$, P_2 при условии, что передан $-s(t)$ и P_3 при условии передачи 0.

b) Определите среднюю вероятность ошибки P_e как функцию от порога A, предполагая, что три сигнала априорно равновероятны.

c) Определите величину A, которая минимизирует P_e.

5.11. Два эквивалентных низкочастотных сигнала, показанных на рис. Р5.11, используются для передачи двоичной информационной последовательности. Передаваемые с одинаковой вероятностью сигналы подвергается воздействию АБГШ с нулевым средним, который имеет эквивалентное низкочастотное представление $z(t)$ с функцией корреляции

$$\phi_{zz}(\tau)=\frac{1}{2}E[z^*(t)z(t+\tau)]=N_0\delta(\tau).$$

a) Какова энергия передаваемого сигнала?

b) Какова вероятность ошибки на бит, если в приёмнике используется когерентное детектирование?

c) Какова вероятность ошибки на бит, если в приёмнике используется некогерентное детектированне?

5.12. В разд. 4.3.1 было показано, что минимальный разнос частот для ортогонализации двоичных сигналов ЧМ с когерентным детектированием равен $\Delta f = 1/2T$.

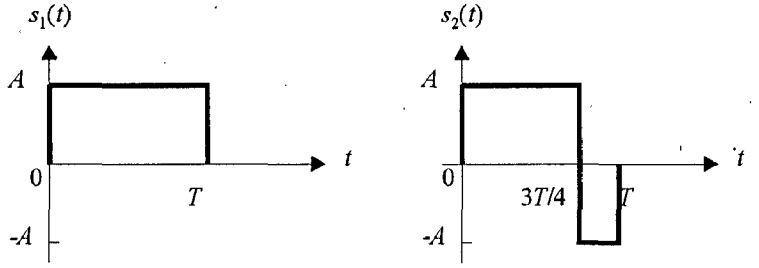


Рис. Р5.11

Однако ещё меньшие значения вероятности ошибки возможны при когерентном детектировании ЧМ, если Δf больше, чем $1/2T$. Покажите, что оптимальное значение Δf равно $0{,}715/T$ и определите вероятность ошибки для этой величины Δf.

5.13. Эквивалентные низкочастотные сигналы для трёх ансамблей сигналов показаны на рис. Р5.13.

Каждый ансамбль можно использовать для передачи одного из четырёх равновероятных сообщений через канал с АБГШ. Эквивалентный низкочастотный шум $z(t)$ имеет нулевое среднее и функцию корреляции $\phi_{zz}(\tau)=N_0\delta(\tau)$.

a) Классифицируйте сигналы в ансамблях I, II и III. Другими словами, установите класс, к которому относится каждый сигнальный ансамбль.

b) Какова средняя передаваемая энергия для каждого сигнального ансамбля?

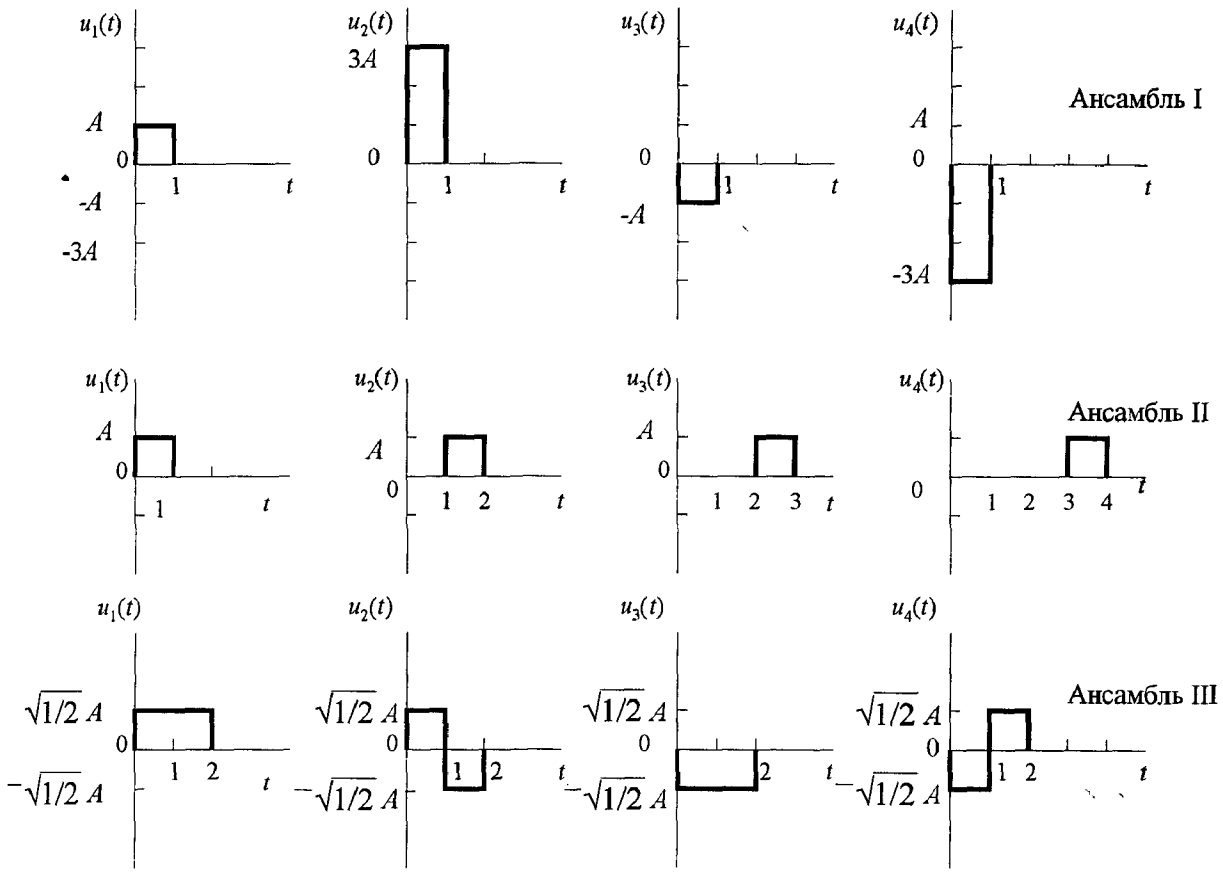


Рис. P5.13

c) Для сигналов ансамбля I определите среднюю вероятность ошибки, если сигналы детектируются когерентно.

d) Для сигналов ансамбля II дайте объединённую границу для вероятности ошибки на символ, если детектирование выполняется когерентно (i) и некогерентно (ii).

e) Возможно ли некогерентное детектирование для сигналов ансамбля III. Объясните.

f) Какой сигнальный ансамбль или сигнальные ансамбли вы выберете, если желаете достичь отношения битовой скорости к полосе частот R/W не менее 2. *Кратко* объясните ваш ответ.

5.14. Рассмотрите четырехпозиционную ($M = 4$) систему связи, которая передает каждые T секунд один из четырех равновероятных сигналов $s_1(t), -s_1(t), s_2(t), -s_2(t)$. Сигналы $s_1(t)$ и $s_2(t)$ ортогональны и с равной энергией. Аддитивный шум гауссовский белый с нулевым средним и автокорреляционной функцией $\phi_{zz}(\tau) = N_0\delta(\tau)$. Демодулятор состоит из двух фильтров, согласованных с $s_1(t)$ и $s_2(t)$, и их выходы в отсчётных точках равны U_1 и U_2.

Детектор основывает свое решение на правиле

$$U_1 > |U_2| \Rightarrow s_1(t), \quad U_1 < -|U_2| \Rightarrow -s_1(t),$$
$$U_2 > |U_1| \Rightarrow s_2(t), \quad U_2 < -|U_1| \Rightarrow -s_2(t).$$

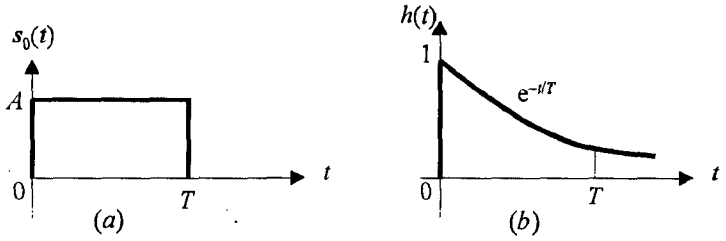


Рис. P5.15

Поскольку сигнальный ансамбль биортогонален, вероятность ошибки равна $(1-P_c)$, где P_c определяется (5.2.34). Выразите эту вероятность ошибки через одномерный интеграл и затем покажите, что вероятность ошибки на символ для биортогонального сигнального ансамбля из $M=4$ сигналов идентична той, которая определяет четырёхфазную ФМ. Подсказка: замена переменных U_1 и U_2 на $W_1=U_1+U_2$ и $W_2=U_1-U_2$ упрощает задачу.

5.15. На вход полосового фильтра поступает сигнал $s(t)$:

$$s(t)=Re\left[s_0(t)e^{j2\pi f_c t}\right],$$

где $s_0(t)$ – прямоугольный импульс, показанный на рис. P5.15 (a).

a) Определите выход $y(t)$ полосового фильтра для всех $t\geq 0$, если импульсная характеристика этого фильтра

$$g(t)=Re\left[2h(t)e^{j2\pi f_c t}\right],$$

где $h(t)$ – экспонента, показанная на рис. P5.15 (b).

b) Нарисуйте выход эквивалентного низкочастотного фильтра.

c) В каких точках возьмёте отсчёты выхода фильтра, если желаете иметь максимальный выход в точках отсчёта? Чему равны величины максимальных отсчётов?

d) Предположите, что $s(t)$ суммируется с АБГШ

$$n(t)=Re\left[z(t)e^{j2\pi f_c t}\right],$$

где $\phi_{zz}(\tau)=N_0\delta(\tau)$. В точках отсчёта, определённых в п. (c), на отсчёты сигнала теперь накладываются шумовые гауссовские слагаемые.

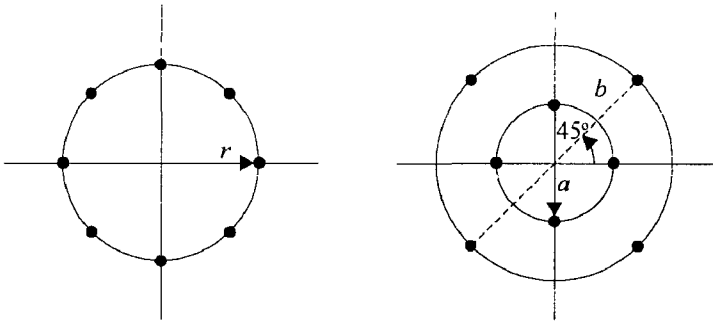
e) Каково отношение сигнал/шум γ по отсчётам выхода?

f) Определите отношение сигнал/шум, когда $h(t)$ – характеристика фильтра, согласованного с $s(t)$, и сравните эти результаты с величиной γ, полученной в (e).

5.16. Рассмотрите восьмиточечное сигнальное созвездие на рис. P5.16.

a) Ближайшие соседние точки в сигнальном созвездии 8-КАМ находятся на расстоянии A единиц. Определите радиусы a и b внутренней и внешней окружности.

b) Соседние сигнальные точки в созвездии 8-ФМ находятся на расстоянии A единиц. Определите радиус r окружности. Определите среднюю переданную мощность для двух сигнальных созвездий и сравните обе мощности.



8-ФМ 8-КАМ

Рис. P5.16

Каков относительный выигрыш по мощности одного созвездия относительно другого? (Предположите, что все сигнальные точки равновероятны.)

5.17. Рассмотрите 8-точечное сигнальное созвездие КАМ, показанное на рис. P.5.16.

a) Возможно ли сопоставить три бита данных каждой точке сигнального созвездия так, чтобы ближайшие (соседние) точки отличались бы только в одном битовом символе?

b) Определите скорость передачи символов, если желательная битовая скорость 90 Мбит/с.

5.18. Предположим, что двоичная ФМ используется для передачи информации по каналу с АБГШ со спектральной плотностью мощности $\frac{1}{2}N_0=10^{-10}$ е□/дль. Энергия передаваемого сигнала $\mathcal{E}_b=\frac{1}{2}A^2T$, где T – битовый интервал, A – амплитуда сигнала. Определите амплитуду сигнала, требуемую для достижения вероятности ошибки 10^{-6}, если скорость передачи данных равна a) 10 кбит/с; b) 100 кбит/с и c) 1 Мбит/с.

5.19. Рассмотрите детектор сигнала со входом $r=\pm A+n$, где $+A$ и $-A$ появляются с равной вероятностью, а шумовое слагаемое n характеризуется лапласовской ФПВ, показанной на рис. P5.19.

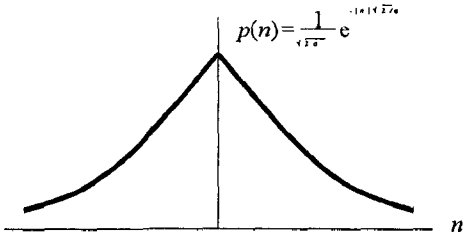
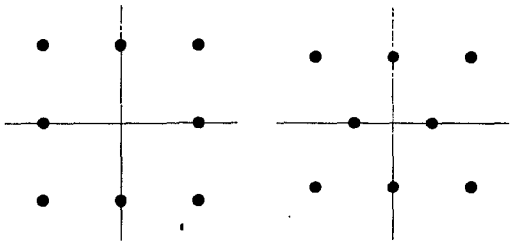


Рис. Р.5.19.

a) Определите вероятность ошибки как функцию от параметров A и σ.

b) Определите ОСШ, требуемое для достижения вероятности ошибки 10^{-5}. Каково требуемое ОСШ по сравнению с результатом для гауссовской ФПВ?

5.20. Рассмотрите два 8-точечных сигнальных созвездия КАМ, показанных на рис. Р5.20.



(*a*) (*b*)

Рис. Р.5.20

Минимальное расстояние между соседними точками равно $2A$. Определите среднюю передаваемую мощность для каждого созвездия, предполагая, что сигнальные точки равновероятны. Какое созвездие эффективнее по мощности?

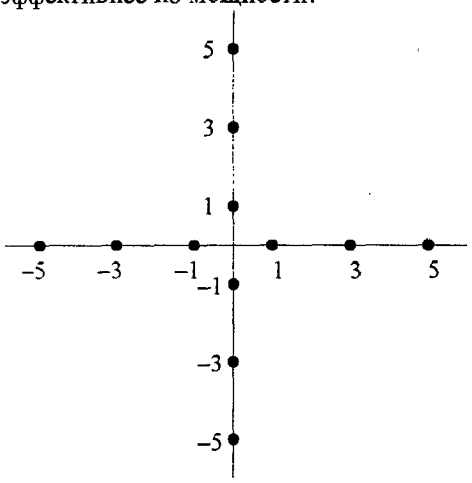


Рис. Р5.21

5.21. Для сигнального созвездия КАМ, показанного на рис. Р5.21, определите оптимальные границы решения детектора, предполагая, что ОСШ достаточно велико, так что ошибочные переходы имеют место только между соседними точками.

5.22. Постройте код Грея для сигнального созвездия 16-КАМ, показанного на рис. Р5.21.

5.23. Квадратурные несущие $cos 2\pi f_c t$ и $sin 2\pi f_c t$ используются для передачи цифровой информации по каналу с АБГШ при двух различных скоростях передачи данных 10 кбит/с и 100 кбит/с. Определите относительные амплитуды сигналов для двух несущих так, чтобы $\mathscr{E}_b/N_b$ для двух каналов были бы одинаковыми.

5.24. Три сообщения m_1, m_2 и m_3 нужно передавать по каналу с АБГШ со спектральной плотностью мощности $\frac{1}{2}N_0$. Сообщения определяются так:

$$s_1(t) = \begin{cases} 1 & (0 \le t \le T), \\ 0 & (\text{для других } t), \end{cases}$$

$$s_2(t) = -s_3(t) = \begin{cases} 1 & (0 \le t < \frac{1}{2}T), \\ -1 & (\frac{1}{2}T < t \le T), \\ 0 & (\text{для других } t). \end{cases}$$

a) Какова размерность пространства сигналов?

b) Найдите подходящий базис для пространства сигналов. (Подсказка: вы можете иайти базис без использования процедуры Грама-Шмидта.)

c) Начертите сигнальное созвездие для этой задачи.

d) Рассчитайте и постройте оптимальные области решения R_1, R_2 и R_3.

e) Какое из трех сообщений более уязвимо для помех и почему? Другими словами, чья условная вероятность $P(\text{ошибка} \mid \text{передан } m_i)$, $i = 1,2,3$, больше?

5.25. Если аддитивный шум на входе демодулятора окрашенный, фильтр, согласованный с сигналом, больше не максимизирует выходное ОСШ. В этом случае мы можем рассмотреть использование предварительного фильтра, который «обеляет» окрашенный шум. За этим предварительным фильтром следует фильтр, согласованный с профильтрованным полезным сигналом. С учетом сказанного рассмотрите схему, показанную на рис. Р5.25.

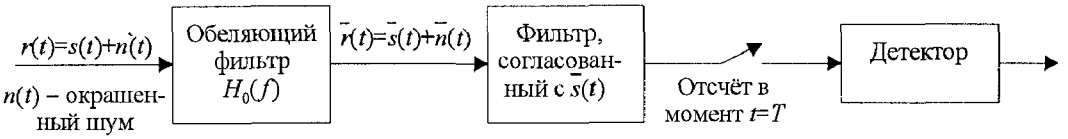


Рис. Р5.25

a) Определите частотную характеристику предварительного фильтра, который обеляет шум.

b) Определите частотную характеристику фильтра, согласованного с $\tilde{s}(t)$.

c) Рассмотрите каскадное соединение предварительного фильтра и согласованного фильтра как единый «обобщённый согласованный фильтр». Какова частотная характеристика этого фильтра?

d) Определите ОСШ на входе детектора.

5.26. Рассмотрите цифровую систему связи, которая передает информацию посредством КАМ через стандартный телефонный канал со скоростью 2400 симв./с. Аддитивный шум считается гауссовским и белым.

a) Определите $\mathscr{E}_b / N_0$, требуемое для достижения вероятности ошибки 10^{-5} при передаче 4800 бит/с.

b) Повторите (a) для скорости 9600 бит/с.

c) Повторите (a) для скорости 19200 бит/с.

d) Какие заключения вы сделаете из этих результатов?

5.27. Рассмотрите четырехфазное и восьмифазное сигнальные созвездия, показанные на рис. Р.5.27.

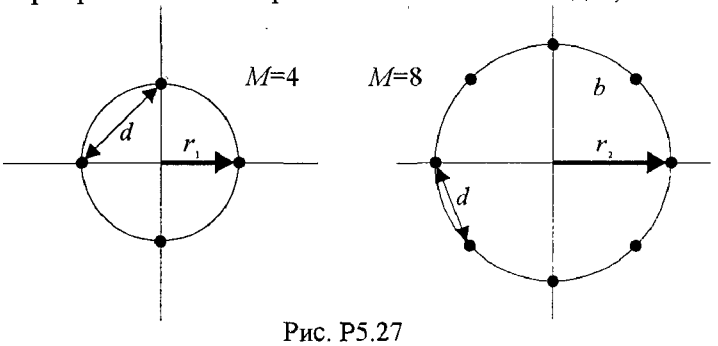


Рис. Р5.27

Определите радиусы r_1 и r_2. окружностей так, чтобы расстояние между ближайшими точками в двух созвездиях было равно d. Исходя из этого результата, определите требуемую дополнительную энергию для 8-ФМ, чтобы достичь той же вероятности ошибки, как при 4-ФМ, причем вероятность ошибки определяется ошибками при выборе соседних точек созвездия.

5.28. Цифровая информация передается модуляцией несущей по каналу с АБГШ с полосой 100 кГц, а $N_0 = 10^{-10}$ Вт/Гц. Определите максимальную скорость передачи по каналу при четырехфазной ФМ, двоичной ФМ и четырехпозиционной ортогональной ЧМ с некогерентным детектированием.

5.29. В сигнале ММС начальное состояние фазы равно 0 или π радиан. Определите финальное состояние фазы для следующих четырех входных пар данных: a) 00; b) 01; c) 10; d) 11.

5.30. Сигнал ЧМ с непрерывной фазой с $h = \frac{1}{2}$ представлен так

$$s(t) = \pm\sqrt{\frac{2\mathscr{E}_b}{T_b}}\cos\left(\frac{\pi T}{2T_b}\right)\cos 2\pi f_c t \pm \sqrt{\frac{2\mathscr{E}_b}{T_b}}\sin\left(\frac{\pi T}{2T_b}\right)\sin 2\pi f_c t, \qquad 0 \le t \le 2T_b,$$

где знаки ± зависят от передаваемых информационных битовых символов.

a) Убедитесь, что этот сигнал имеет постоянную амплитуду.

b) Нарисуйте блок-схему модулятора для синтеза сигнала.

c) Нарисуйте блок-схему демодулятора и детектора для извлечения информации.

5.31. Нарисуйте фазовое дерево, решётку состояний и диаграмму состояний для МНФ с парциальным откликом при $h = \frac{1}{2}$ и

$$g(t) = \begin{cases} 1/4T & (0 \le t \le 2T), \\ 0 & (\text{для других } t). \end{cases}$$

5.32. Определите число финальных состояний фазы на диаграмме решётки состояний для (a) двоичной ЧМНФ с полным откликом с $h = \frac{2}{3}$ или $\frac{3}{4}$ и (b) двоичной ЧМНФ с парциальным откликом с $L = 3$ и $h = \frac{2}{3}$ или $\frac{3}{4}$.

5.33. Рассмотрите биортогоиальный ансамбль с $M = 8$ сигнальными точками. Определите объединённую верхнюю границу для вероятности ошибки на символ как функцию $\mathscr{E}_b/N_0$. Сигнальные точки считаются априори равновероятными.

5.34. Рассмотрите M-позиционную цифровую систему, где $M = 2^N$, а N – размериость пространства сигналов. Предположим, что точки M сигнальных векторов лежат на вершинах гиперкуба, который центрирован относительно начала координат. Определите среднюю вероятность ошибки на символ как функцию от $\mathscr{E}_s/N_0$, где $\mathscr{E}_s$ – энергия на символ, $\frac{1}{2}N_0$ – спектральная плотность мощности АБГШ, а все сигнальные точки равновероятны.

5.35. Рассмотрите сигнал

$$s(t) = \sum_{i=1}^{n} c_i p(t' - kT_c),$$

где $p(t)$ – прямоугольный импульс единичной амплитуды и длительности T_c. Коэффициенты $\{c_i\}$ можно рассматривать как кодовый вектор $\mathbf{C} = [c_1 c_2 ... c_n]$, где элементы $c_i = \pm 1$. Покажите, что фильтр, согласованный с сигналом $s(t)$, можно реализовать как каскад из фильтра, согласованного с $p(t)$, и фильтра дискретного времени, согласованного с вектором $\mathbf{C}$. Определите величину выхода согласованного фильтра в точках отсчётов $t = nT_c$.

5.36. Речевой сигнал стробируется со скоростью 8 кГц, логарифмически сжимается и кодируется в ИКМ-формат с использованием 8 бит/отсчёт. Данные ИКМ передаются через базовый канал с АБГШ посредством M-уровневой АМ. Определите полосу частот, требуемую для передачи, когда (a) $M = 4$, (b) $M = 8$, (c) $M = 16$.

5.37. Матрица Адамара определяется как матрица, чьи элементы равны ± 1 и чьи вектор- строки попарно ортогональны. Для случая, когда n равно степени 2, $n \times n$ матрица Адамара конструируется посредством рекуррентной процедуры:

$$\mathbf{H}_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \mathbf{H}_{2n} = \begin{bmatrix} \mathbf{H}_n & \mathbf{H}_n \\ \mathbf{H}_n & -\mathbf{H}_n \end{bmatrix}.$$

a) Пусть $\mathbf{C}_i$ означает i-ю строку $n \times n$ матрицы Адамара, как она определена выше. Покажите, что сигналы, образованные так:

$$s_i(t) = \sum_{k=1}^{n} c_{ik} p(t - kT_c), i = 1,2,...n,$$

ортогональны, где $p(t)$ – произвольный импульс, определённый на временном интервале $0 \le t \le T_c$.

b) Покажите, что согласованные фильтры (или взаимные корреляторы) для n сигналов $\{s_i(t)\}$ могут быть реализованы посредством единственного фильтра (или коррелятора), согласованного с импульсом $p(t)$, за которым следуют n взаимных корреляторов, использующих кодовые слова $\{\mathbf{C}_i\}$.

5.38. Дискретная последовательность

$$r_k = \sqrt{\mathscr{E}_b} c_k + n_k, k = 1,2,...n,$$

представляет выходную последовательность отсчётов демодулятора, где $c_k = \pm 1$ – элементы одного из двух возможных кодовых слов $\mathbf{C}_1 = [11...111...1]$ и $\mathbf{C}_2 = [11...1-1-1...-1]$. Кодовое слово $\mathbf{C}_2$ имеет w элементов +1 и

$n-w$ элементов -1, где w – некоторое положительное целое число. Шумовая последовательность $\{n_k\}$ является гауссовской и белой с дисперсией σ^2.

a) Каков детектор максимального правдоподобия для двух возможных передаваемых сигналов?

b) Определите вероятность ошибки как функцию параметров (σ^2, E_b, w).

c) Какова величина w, которая минимизирует вероятность ошибки?

5.39. Получите выходы r_1 и r_2 двух корреляторов, показанных на рис. 5.4.1. Считайте, что передаётся сигнал $s_{l1}(t)$ и что

$$r_l(t) = s_{l1}(t)e^{j\phi} + z(t),$$

где $z(t) = n_c(t) + jn_s(t)$ – аддитивный гауссовский шум.

5.40. Определите ковариации и дисперсии гауссовских случайных шумовых величин $n_{1c}, n_{2c}, n_{1s}, n_{2s}$ в (5.4.15) и их совместную ФПВ.

5.41. Получите выходы согласованного фильтра, определенные (5.4.10).

5.42. В двоичной системе АМ с пассивной паузой два возможных сигнала

$$s_0(t) = 0,\ 0 \le t \le T_b,$$

$$s_1(t) = \sqrt{\frac{2E_b}{T_b}}\cos 2\pi f_c t, 0 \le t \le T_b.$$

Соответствующие принимаемые сигналы

$$r(t) = n(t),\ 0 \le t \le T_b,$$

$$r(t) = \sqrt{\frac{2E_b}{T_b}}\cos(2\pi f_c t + \phi) + n(t), 0 \le t \le T_b,$$

где ϕ – фаза несущей, а $n(t)$ – АБГШ.

a) Нарисуйте блок-схему приёмника (демодулятора и детектора), который реализует некогерентное (по огибающей) детектирование.

b) Определите ФПВ для двух участвующих в решении на выходе детектора величин, соответствующих двум возможным принятым сигналам.

c) Определите вероятность ошибки детектора.

5.43 В двухфазной ДФМ принимаемый сигнал на одном сигнальном интервале используется как опорный (эталонный) для принимаемого сигнала на следующем сигнальном интервале.

Решение принимается по величине

$$D = Re\left(V_m V_{m-1}^*\right) \underset{"0"}{\overset{"1"}{\gtrless}} 0,$$

где

$$V_k = 2\alpha e^{j(\theta_k - \phi)} + N_k$$

представляют комплексный выход фильтра, согласованного с переданным сигналом $u(t)$, N_k – комплексные гауссовские величины с нулевым средним и статистически независимыми компонентами.

a) Написав $V_k = X_k + jY_k$, покажите, что D эквивалентно

$$D = \left[\tfrac{1}{2}(X_m + X_{m-1})\right]^2 + \left[\tfrac{1}{2}(Y_m + Y_{m-1})\right]^2 - \left[\tfrac{1}{2}(X_m - X_{m-1})\right]^2 - \left[\tfrac{1}{2}(Y_m - Y_{m-1})\right]^2.$$

b) Для математического удобства предположите, что $\theta_k = \theta_{k-1}$.

Покажите, что случайные величины U_1, U_2, U_3, U_4 – статистически независимые гауссовские величины, где $U_1 = \tfrac{1}{2}(X_m + X_{m-1})$, $U_2 = \tfrac{1}{2}(Y_m + Y_{m-1})$, $U_3 = \tfrac{1}{2}(X_m - X_{m-1})$, $U_4 = \tfrac{1}{2}(Y_m - Y_{m-1})$.

c) Введите случайные величины $W_1 = U_1^2 + U_2^2$ и $W_2 = U_3^2 + U_4^2$. Тогда решающее правило детектора $D = W_1 - W_2 \underset{"0"}{\overset{"1"}{\gtrless}} 0$. Определите ФПВ для W_1 и W_2.

d) Определите вероятность ошибки P_b:

$$P_b = P(D < 0) = P(W_1 - W_2 < 0) = \int_0^\infty P(W_2 > w_1 | w_1) p(w_1) dw_1.$$

5.44. Напомним, что ММС можно представить как четырехфазную офсетную ФМ, имеющую низкочастотный эквивалент

$$\upsilon(t) = \sum_k \left[I_k u(t - 2kT_b) + jJ_k u(t - 2kT_b - T_b) \right],$$

где

$$u(t) = \begin{cases} \sin(\pi t / 2T_b) & (0 \le t \le 2T_b), \\ 0 & (\text{для других } t), \end{cases}$$

а $\{I_k\}$ и $\{J_k\}$ – последовательности информационных символов (± 1).

a) Нарисуйте блок-схему демодулятора ММС для ОКФМ

b) Рассчитайте характеристики качества для четырехфазного демодулятора в канале с АБГШ, если не принимать во внимание память модулятора.

c) Сравните качество, полученное в (b), с тем, что дает декодирование по Витерби ММС сигнала.

d) ММС сигнал также эквивалентен двоичной ЧМ. Определите качество некогерентного детектирования ММС. Сравните результат с (b) и (c).

5.45. Рассмотрите линию связи, которая для передачи двоичной информации использует $n-1$ регенеративных повторителей и оконечный приёмник. Предположите, что вероятность ошибки детектора в каждом приёмнике равна p и что ошибки у повторителей статистически независимы.

a) Покажите, что вероятность ошибки у конечного приёмника равна

$$P_n = \frac{1}{2}\left[1 - (1-2p)^n\right].$$

b) Если $p = 10^{-6}$ и $n = 100$, определите приближенное значение P_n.

5.46. Двоичная система связи состоит из линии передачи и 100 цифровых (регенеративных) повторителей. Для передачи информации используются двоичные противоположные сигналы. Если сквозная величина вероятности ошибки для системы равна 10^{-6}, определите вероятность ошибки для каждого повторителя и требуемое $\mathscr{E}_b / N_0$, чтобы достичь такое качество в канале с АБГШ.

5.47. Радиопередатчик имеет выходную мощность $P_T = 1$е□ на частоте 1 ГГц. Передающие и приёмные антенны – параболические тарелки с диаметром $D = 3$".

a) Определите усиление антенны.

b) Определите ЭМИ (эффективную мощность излучения) для передатчика.

c) Расстоянне (в свободном пространстве) между передающей н приёмной антенпами равно 20 км. Определите мощность передатчика на выходе приёмной антенны в дБм.

5.48. Система радиосвязи передает уровень мощности 0,1 Вт на 1 ГГц. Передающая и приёмная антеины параболические, каждая с диаметром 1 м. Приёмник удален от передатчика на 30 км.

a) Определите усиление передающей и приёмной антенн.

b) Определите ЭМИ переданного сигнала.

c) Определите мощность сигнала на выходе приёмной антенны.

5.49. Спутник на геостационарной орбите используется для связи с земной станцией на расстоянии 40 000 км. Спутник имеет антенну с усилением 15 дБ и передаваемую мощность 3 Вт. Земная станция использует 10-метровую параболическую антенну с эффективностью 0,6. Полоса частот $f = 10$ддль. Определите принимаемый уровень мощности на выходе приёмной антенны.

5.50. Расположенный на расстоянии 100 000 км от Земли космический аппарат дистанционного зондирования посылает данные со скоростью R бит/с. Полоса частот концентрируется у 2ддль, а мощность передатчика 10 Вт. Земная станция использует параболическую антенну с диаметром 50 м, а спутник имеет антенну с усилением 10 дБ. Шумовая температура приёмника $T_0 = 300$K.

a) Определите уровень принимаемой мощности.

b) При условии, что $\mathscr{E}_b / N_0 = 10$Льз, определите максимальную битовую скорость, с которой космический аппарат может передавать данные.

5.51. Спутник на геостационарной орбите используется как регенеративный повторитель в цифровой системе связи. Рассмотрите линию спутник–Земля, в которой антенна спутника имеет усиление 6 дБ, а стационарная антенна Земли имеет усиление 50 дБ. Линия вниз работает на центральной частоте 4 ГГц, а полоса сигнала 1 МГц. Если требуемая величина $\mathscr{E}_b / N_0$ для реализации связи равна 15 дБ, определите переданную мощность от спутника. Считайте, что $N_0 = 4{,}1 \times 10^{-21}$ Вт / Гц.

6

СИНХРОНИЗАЦИЯ НЕСУЩЕЙ И ТАКТОВАЯ СИНХРОНИЗАЦИЯ

Мы видели, что в цифровых системах выход демодулятора должен стробироваться периодически, один раз на интервале символа (такта) для того, чтобы восстановить переданную информацию. Поскольку время распространения от передатчика к приёмнику обычно неизвестно на приёме, управление этим стробированием должно осуществляться от принимаемого сигнала. Время распространения переданного сигнала влияет и на флуктуацию фазы несущей, которую также следует оценить в приемнике, если детектор когерентный. В этой главе мы рассмотрим методы обеспечения синхронизации по несущей и по тактам в приёмнике.

6.1. ОЦЕНИВАНИЕ ПАРАМЕТРОВ СИГНАЛА

Начнем с разработки математической модели для сигнала на входе приемника. Предположим, что канал задерживает переданные сигналы и искажает их посредством аддитивного гауссовского шума. Следовательно, принимаемый сигнал можно выразить так:

$$r(t) = s(t-\tau) + n(t),$$

где

$$s(t) = \mathrm{Re}\left[s_l(t)\mathrm{e}^{j2\pi f_c t}\right], \qquad (6.1.1)$$

τ – время задержки при распространении волны[1], а $s_l(t)$ – эквивалентный низкочастотный сигнал. Принимаемый сигнал можно выразить и так:

$$r(t) = \mathrm{Re}\left\{\left[s_l(t-\tau)\mathrm{e}^{j\phi} + z(t)\right]\mathrm{e}^{j2\pi f_c t}\right\}, \qquad (6.1.2)$$

где фаза несущей, обусловленная задержкой при распространении волны, равна $\phi = -2\pi f_c \tau$, $z(t)$ – низкочастотный эквивалент шума. В этой формулировке может показаться, что имеется только один сигнальный параметр, который следует оценить, именно задержка распространения τ, поскольку тогда при известных τ и f_c можно определить ϕ. Однако это не так. Прежде всего, генератор, который генерирует несущую для демодуляции в приёмнике, в общем не синхронизирован с фазой передатчика. Более того, два генератора могут медленно дрейфовать со временем, возможно, в разные стороны. Следовательно, фаза принимаемой несущей не только зависит от времени задержки τ. Более того, точность, с которой необходима синхронизация во времени для целей демодуляции принимаемого сигнала, зависит от сигнального интервала T. Обычно ошибка оценки τ должна составлять относительно малую часть интервала T. Например, ошибка в $\pm 1\%$ от T соответствует практическим применениям. Однако этот уровень точности обычно недостаточен для оценки фазы несущей, даже если ϕ зависит только от τ.

[1] Следовало бы разделить понятия задержки по огибающей (низкочастотного эквивалента) и фазовой задержки, определяющей величину ϕ (прп)

Это объясняется тем, что несущая f_c обычно велика, и, следовательно, малая ошибка в оценке τ вызовет большую ошибку в оценке фазы.

Фактически мы должны оценить оба параметра τ и ϕ для того, чтобы демодулировать, а затем, когерентно детектировать принимаемый сигнал. Следовательно, мы можем выразить принимаемый сигнал так:

$$r(t) = s(t;\phi,\tau) + n(t), \qquad (6.1.3)$$

где τ и ϕ представляют сигнальные параметры, которые должны быть оценены. Чтобы упростить обозначения, обозначим через ψ векторный параметр $\{\phi,\tau\}$, так что $s(t,\phi,\tau)$ проще обозначать $s(t;\psi)$.

Имеются два базовых подхода, которые широко используются для оценки сигнальных параметров: подход с использованием правила *максимального правдоподобия* (МП) и по правилу *максимума апостериорной вероятности* (МАВ). В правиле МАВ сигнальный векторный параметр считается случайным и характеризуется априорной плотностью вероятности $p(\psi)$. По правилу максимального правдоподобия (МП) сигнальный векторный параметр ψ трактуется как детерминированный, но неизвестный.

Формируя ортонормированное разложение $r(t)$ по N ортонормированным функциям $\{f_n(t)\}$, мы можем представить $r(t)$ вектором коэффициентов $[r_1\ r_2\ ...r_N] \equiv \mathbf{r}$. Совместную ФПВ случайных величин $[r_1\ r_2\ ...r_N]$ можно выразить как условную плотность вероятности $p(\mathbf{r}|\psi)$. Тогда МП оценка ψ – это величина, которая максимизирует $p(\mathbf{r}|\psi)$. С другой стороны, оценка МАВ – это величина ψ, которая максимизирует апостериорную функцию плотности вероятности

$$p(\psi|\mathbf{r}) = \frac{p(\mathbf{r}|\psi)p(\psi)}{p(\mathbf{r})}. \qquad (6.1.4)$$

Заметим, что, если неизвестна априорная плотность параметра ψ, мы можем предположить, что $p(\psi)$ равномерно распределена (константа) во всей области значений параметра. В этом случае величина ψ, которая максимизирует $p(\mathbf{r}|\psi)$, также максимизирует $p(\psi|\mathbf{r})$. При этом оценки МАВ и МП совпадают.

В нашей трактовке оценок параметров, даваемой ниже, рассмотрим параметры ϕ и τ как неизвестные, но детерминированные. Следовательно, для оценки используем правило МП. При МП оценке сигнальных параметров мы требуем, чтобы приёмник извлекал оценку путем наблюдения принимаемого сигнала на интервале времени $T_0 \geq T$, который называется интервалом наблюдения. Оценки, полученные на одном сигнальном интервале наблюдения, иногда называют разовыми оценками. На практике, однако, оценка формируется непрерывно путем использования замкнутых следящих систем (аналоговых или цифровых), которые непрерывно улучшают оценки. Тем не менее разовые оценки лежат в основе построения следящих систем оценивания. Кроме того, они оказываются полезными при анализе качества МП оценки, и их качество можно связать с тем, которое получается при петлевом отслеживании.

6.1.1. Функция правдоподобия

Полагая, что аддитивный шум в канале $n(t)$ является гауссовским и белым с нулевым средним, можно совместную ФПВ $p(\mathbf{r}|\psi)$ выразить так:

$$p(\mathbf{r}|\psi)=\left(\frac{1}{\sqrt{2\pi\sigma}}\right)^N \exp\left\{-\sum_{n=1}^{N}\frac{[r_n-s_n(\psi)]^2}{2\sigma^2}\right\}, \qquad (6.1.5)$$

где σ^2 – дисперсия случайной величины r_n,

$$r_n=\int_{T_0} r(t)f_n(t)dt,$$

$$s_n(\psi)=\int_{T_0} s(t;\psi)f_n(t)dt, \qquad (6.1.6)$$

а T_0 представляет интервал интегрирования в разложении $r(t)$ и $s(t;\psi)$.

Заметим, что аргумент в экспоненте можно выразить через сигналы $r(t)$ и $s(t;\psi)$, подставив (6.1.6) в (6.1.5).

Таким образом,

$$\frac{1}{2\sigma^2}\sum_{n=1}^{N}[r_n-s_n(t;\psi)]^2=\frac{1}{N_0}\int_{T_0}[r(t)-s(t;\psi)]^2dt, \qquad (6.1.7)$$

где N_0 – спектральная плотность мощности случайного процесса $n(t)$, причём доказательство оставлено как упражнение для читателя (см. задачу 6.1). Теперь максимизация $p(\mathbf{r}|\psi)$ по сигнальному параметру ψ эквивалентна максимизации функции[1] правдоподобия

$$\Lambda(\psi)=\exp\left\{-\frac{1}{N_0}\int_{T_0}[r(t)-s(t;\psi)]^2dt\right\}. \qquad (6.1.8)$$

Ниже рассмотрим оценивание параметра сигнала с точки зрения максимизации $\Lambda(\psi)$.

6.1.2. Восстановление несущей и тактовая синхронизация при демодуляции сигнала

Тактовая синхронизация требуется в любой цифровой системе связи, которая передает информацию синхронно. Восстановление несущей требуется, если сигнал детектируется когерентно.

Рисунок 6.1.1 иллюстрирует блок-схему демодулятора и детектора для двоичной ФМ (или двоичной АМ). Как видно, оценка фазы несущей $\hat{\phi}$ нужна при генерировании опорного сигнала $g(t)\cos(2\pi f_c t+\hat{\phi})$ для коррелятора. Синхронизатор тактов управляет стробирующим устройством и генератором сигнального импульса. Если сигнальный импульс прямоугольный, тогда генератор сигнального импульса можно исключить.

Блок-схема демодулятора для M-позиционной ФМ показана на рис. 6.1.2. В этом случае требуются два коррелятора (или согласованных фильтра) для корреляции принимаемых сигналов с двумя квадратурными несущими $g(t)\cos(2\pi f_c t+\hat{\phi})$ и $g(t)\sin(2\pi f_c t+\hat{\phi})$, где $\hat{\phi}$ – оценка фазы несущей. Теперь детектор фазовый, он сравнивает фазы принимаемого сигнала с возможными к передаче фазами сигнала.

---

[1] При непрерывном времени здесь и в дальнейшем следует говорить о функционале правдоподобия, а вернее об отношении функционалов правдоподобия, так как отсутствует множитель при экспоненте в (6.1.5) (прп)

Блок-схема демодулятора M-позиционной АМ показана на рис. 6.1.3. В этом случае требуется единственный коррелятор, а детектор является амплитудным детектором, который сравнивает амплитуду принимаемого сигнала с возможными к передаче амплитудами сигналов.

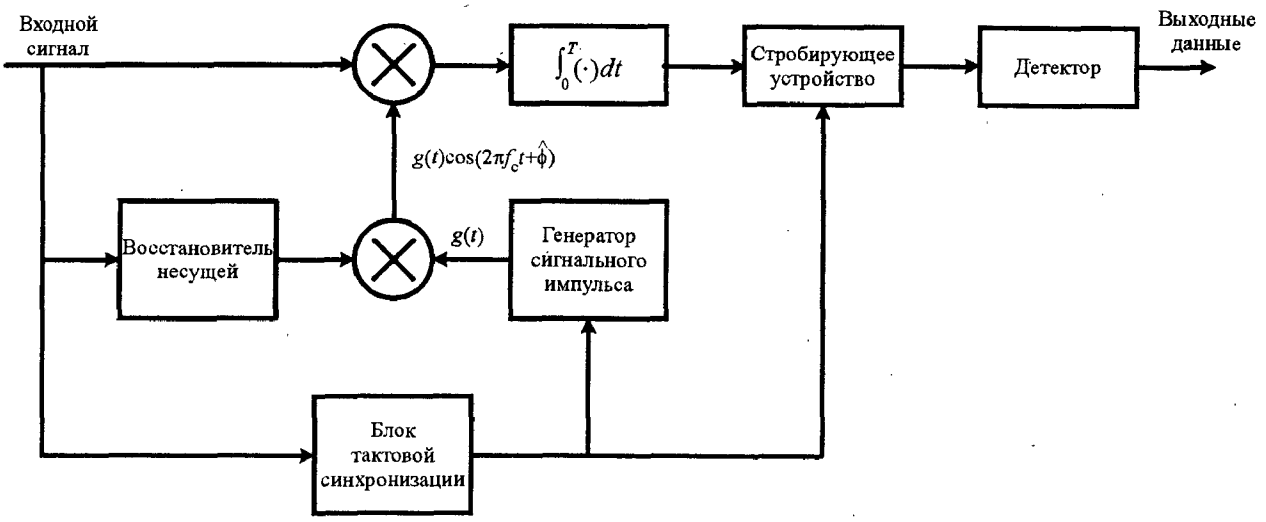


Рис. 6.1.1. Блок-схема приёмника двоичной ФМ

Заметим, что мы включили автоматическую регулировку усиления (АРУ) на входе демодулятора, чтобы скомпенсировать изменения коэффициента усиления канала, которые нарушают работу амплитудного детектора. АРУ имеет относительно большую постоянную времени, так что она не реагирует на быстрые изменения амплитуды сигнала, которые происходят от символа к символу. Вместо этого АРУ поддерживает фиксированное значение средней (сигнал+шум) мощности на своём выходе.

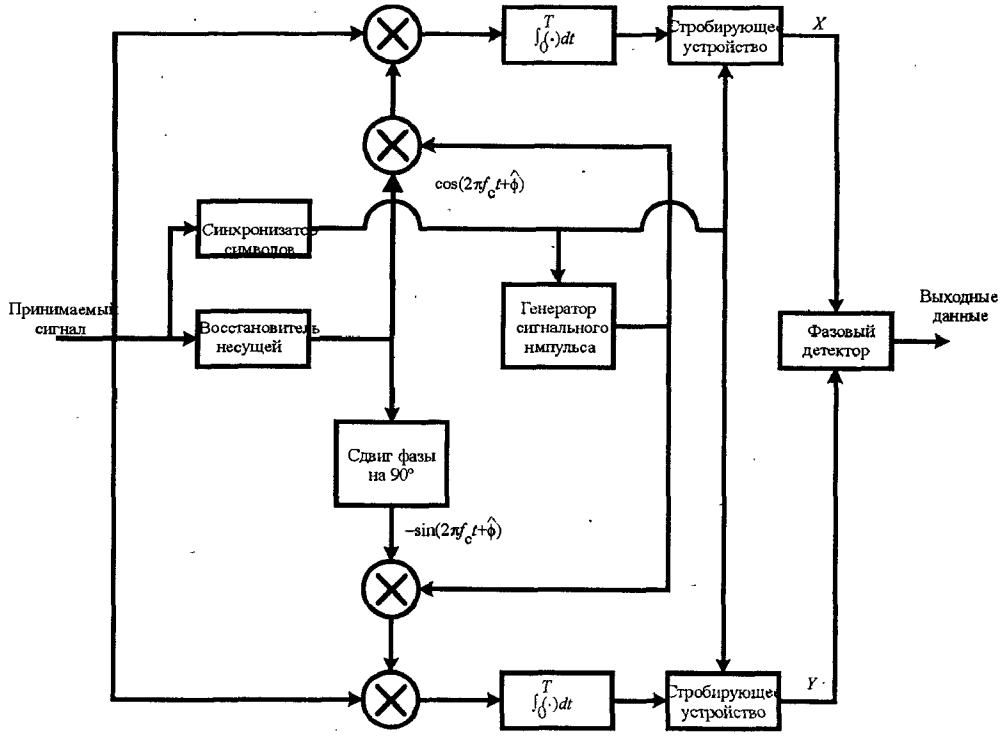


Рис. 6.1.2. Блок-схема приёмника M-позиционной ФМ

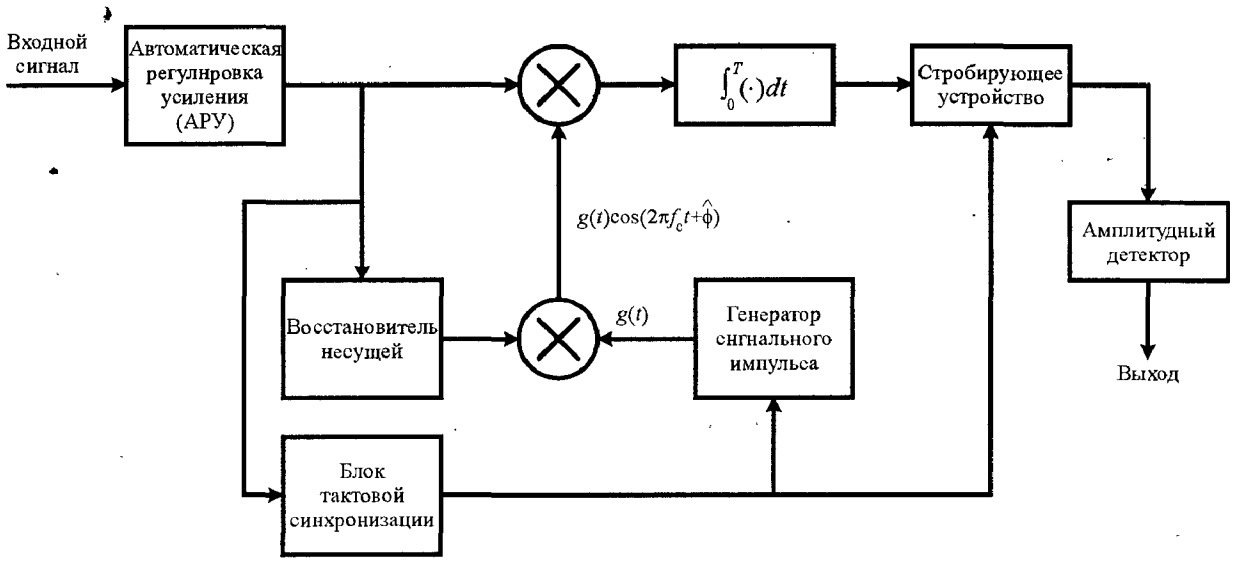


Рис. 6.1.3. Блок-схема приёмника M-позиционных сигналов АИМ

В заключение на рис. 6.1.4 проиллюстрируем блок-схему демодулятора сигналов КАМ. Как в случае АМ, требуется АРУ, чтобы поддержать постоянную среднюю мощность сигнала на входе демодулятора. Видим, что демодулятор похож на демодулятор ФМ, в обоих генерируются синфазные и квадратурные отсчеты сигнала (X,Y) для детектора. В случае КАМ детектор вычисляет евклидово расстояние между точкой принимаемого сигнала, искажённого шумом, и M возможными к передаче точками сигнала и выбирает сигнал, расположенный наиболее близко к точке принимаемого сигнала.

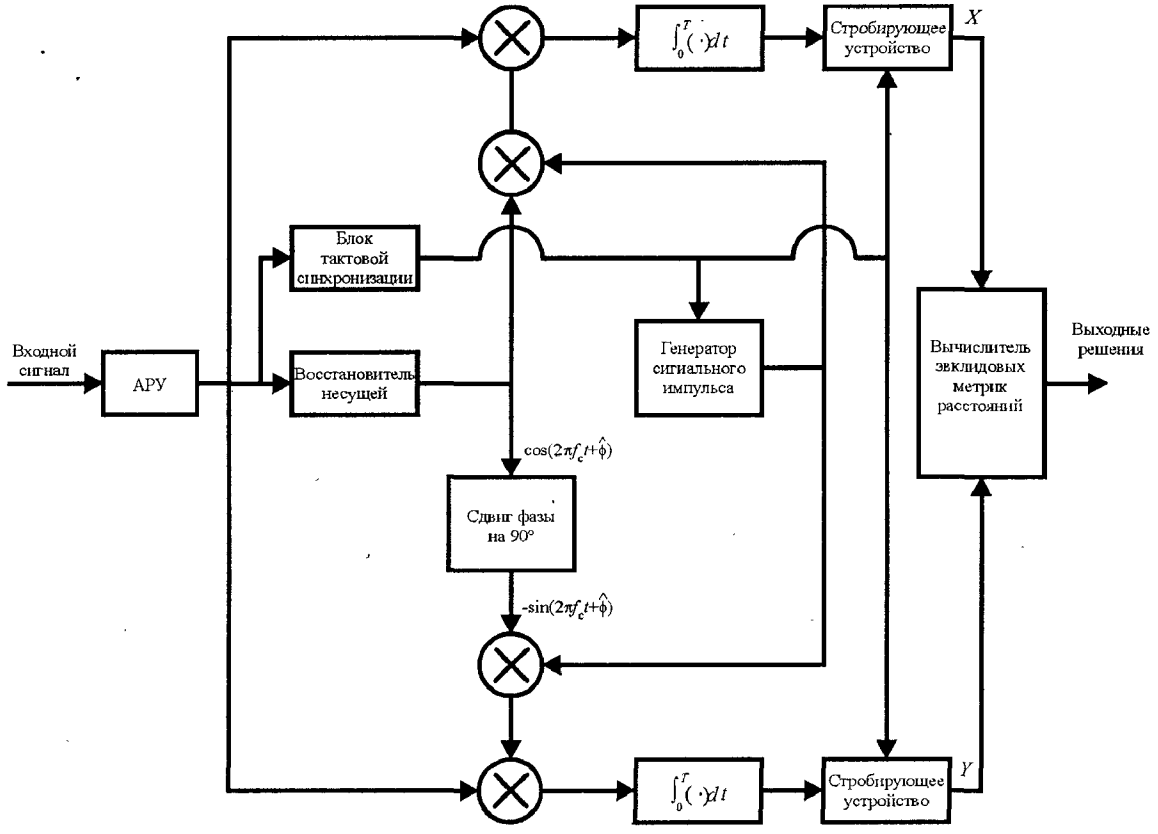


Рис. 6.1.4. Блок-схема приёмника КАМ

6.2. ОЦЕНИВАНИЕ ФАЗЫ НЕСУЩЕЙ

Имеются два базовых подхода для обеспечения синхронизации по несущей в приёмнике. Один сводится к замешиванию на передаче, обычно по частоте, специального сигнала, называемого пилот-сигналом, который можно извлечь из принимаемого сигнала и по его частоте и фазе синхронизировать местный генератор несущей. Когда по каналу передана одновременно немодулированная несущая (пилот-сигнал) и информационный сигнал, приёмник использует замкнутую петлю фазовой автоподстройки (ФАП, PLL), чтобы захватить и отслеживать несущую. ФАП рассчитывается так, чтобы иметь узкую полосу пропускания, так что она слабо подвергается воздействию частотных компонент информационного сигнала.

Второй подход, который более привлекателен на практике, сводится к непосредственной оценке фазы несущей по модулированному сигнала. Этот подход имеет отчётливое преимущество, поскольку вся мощность передатчика направлена на передачу информационного сигнала. В нашей трактовке восстановления несущей, ограничим наше внимание вторым подходом, следовательно, мы предполагаем, что сигнал передан с подавленной несущей.

Чтобы подчеркнуть важность получения точной оценки фазы, рассмотрим влияние ошибки в оценке фазы несущей на демодуляцию двухполосного с подавленной несущей (ДП/ПН – DSB/SC) сигнала. Для конкретности предположим, что имеем амплитудно-модулированный сигнал вида

$$s(t) = A(t)\cos\left(2\pi f_c t + \phi\right). \qquad (6.2.1)$$

После умножения (в демодуляторе) сигнала $s(t)$ на опорную несущую

$$c(t) = \cos\left(2\pi f_c t + \hat{\phi}\right) \qquad (6.2.2)$$

получим

$$c(t)s(t) = \tfrac{1}{2}A(t)\cos\left(\phi - \hat{\phi}\right) + \tfrac{1}{2}A(t)\cos\left[4\pi f_c t + \phi + \hat{\phi}\right].$$

Компонента удвоенной частоты может быть устранена путём пропускания произведения $c(t)s(t)$ через ФНЧ. Такая фильтрация даёт информационный сигнал

$$y(t) = \tfrac{1}{2}A(t)\cos\left(\phi - \hat{\phi}\right). \qquad (6.2.3)$$

Заметим, что воздействие фазовой ошибки $\phi - \hat{\phi}$ ведёт к изменению уровня сигнала по напряжению в $\cos\left(\phi - \hat{\phi}\right)$ раз, а по мощности в $\cos^2\left(\phi - \hat{\phi}\right)$ раз.

Следовательно, фазовая ошибка в 10° ведёт к потере в мощности сигнала на 0,13 дБ, а фазовая ошибка на 30° ведёт к потере в мощности сигнала на 1,25 дБ, когда речь идёт о модулированном по амплитуде сигнале. Влияние фазовой ошибки при КАМ и многопозиционной ФМ оказывается более серьёзным. Сигнал КАМ и многопозиционной ФМ можно представить так:

$$s(t) = A(t)\cos\left(2\pi f_c t + \phi\right) - B(t)\sin\left(2\pi f_c t + \phi\right). \qquad (6.2.4)$$

Этот сигнал демодулируется посредством двух квадратурных несущих

$$c_c(t) = \cos\left(2\pi f_c t + \hat{\phi}\right),$$

$$c_s(t) = -\sin\left(2\pi f_c t + \hat{\phi}\right). \qquad (6.2.5)$$

Умножение $s(t)$ и $c_c(t)$ даёт после ФНЧ синфазную компоненту

$$y_c(t) = \tfrac{1}{2} A(t)\cos\left(\phi - \hat{\phi}\right) - \tfrac{1}{2} B(t)\sin\left(\phi - \hat{\phi}\right). \qquad (6.2.6)$$

Аналогично умножение $s(t)$ на $c_s(t)$ даёт после ФНЧ квадратурную компоненту

$$y_s(t) = \tfrac{1}{2} B(t)\cos\left(\phi - \hat{\phi}\right) + \tfrac{1}{2} A(t)\sin\left(\phi - \hat{\phi}\right). \qquad (6.2.7)$$

Выражения (6.2.6) и (6.2.7) ясно указывают на то, что ошибка фазы при детектировании КАМ и M-позиционной ФМ имеет более тяжёлые последствия, чем при детектировании АМ сигналов. Здесь имеет место не только уменьшение мощности желательной сигнальный компоненты в $\cos^2\left(\phi - \hat{\phi}\right)$ раз, но также взаимная интерференция между синфазной и квадратурной компонентой. Поскольку средние уровни мощности $A(t)$ и $B(t)$ равны, малая ошибка в фазе вызовет большое ухудшение качества. Следовательно, фазовая точность, требуемая при КАМ и многопозиционной ФМ, более высокая, чем при ДП/ПН АМ.

6.2.1. МП оценка фазы несущей

Сначала рассмотрим максимально правдоподобную оценку фазы несущей. Для простоты предположим, что задержка τ известна, и, в частности, предположим $\tau = 0$. Максимизации подлежит функция правдоподобия, определяемая (6.1.8). Если ϕ подставить в ψ для этой функции, получим

$$\begin{aligned}\Lambda(\phi) &= \exp\left\{-\frac{1}{N_0}\int_{T_0}\left[r(t) - s(t;\phi)\right]^2 dt\right\} = \\ &= \exp\left\{-\frac{1}{N_0}\int_{T_0} r^2(t)dt + \frac{2}{N_0}\int_{T_0} r(t)s(t;\phi)dt - \frac{1}{N_0}\int_{T_0} s^2(t;\phi)dt\right\}.\end{aligned} \qquad (6.2.8)$$

Заметим, что первое слагаемое справа не содержит параметра ϕ.

Третье слагаемое, которое содержит интеграл от $s^2(t;\phi)$, является постоянной, равной энергии сигнала на интервале T_0 для любых значений ϕ. Только второе слагаемое, которое включает в себя взаимную корреляцию принимаемого сигнала $r(t)$ с сигналом $s(t;\phi)$, зависит от ϕ. Следовательно, функцию правдоподобия $\Lambda(\phi)$ можно выразить так:

$$\Lambda(\phi) = C\exp\left[\frac{2}{N_0}\int_{T_0} r(t)s(t;\phi)dt\right], \qquad (6.2.9)$$

где C – константа, не зависящая от ϕ.

МП оценка $\hat{\phi}_{\text{МП}}$ – это величина ϕ, которая максимизирует $\Lambda(\phi)$ в (6.2.9). Эквивалентно $\hat{\phi}_{\text{МП}}$ максимизирует также $\ln\Lambda(\phi)$, т.е. логарифм функции правдоподобия

$$\Lambda_L(\phi) = \frac{2}{N_0}\int_{T_0} r(t)s(t;\phi)dt. \qquad (6.2.10)$$

Заметим, что при определении $\Lambda_L(\phi)$ мы отбросили слагаемое $\ln C$.

Пример 6.2.1. Как пример оптимизации при оценке фазы несущей рассмотрим передачу по каналу немодулированной несущей $A\cos 2\pi f_c t$. Принимаемый сигнал

$$r(t) = A\cos\left(2\pi f_c t + \phi\right) + n(t),$$

где ϕ – неизвестная фаза. Мы ищем величину ϕ, скажем $\hat{\phi}_{\text{МП}}$, которая максимизирует

$$\Lambda_L(\phi) = \frac{2A}{N_0}\int_{T_0} r(t)\cos(2\pi f_c t + \phi)dt .$$

Необходимое условие для максимума

$$\frac{d\Lambda_L(\phi)}{d\phi} = 0 .$$

Это условие даёт

$$\int_{T_0} r(t)\sin(2\pi f_c t + \hat{\phi}_{\text{МП}})dt = 0 \qquad (6.2.11)$$

или, что эквивалентно,

$$\hat{\phi}_{\text{МП}} = -\text{arc tg}\left[\int_{T_0} r(t)\sin 2\pi f_c t dt \Big/ \int_{T_0} r(t)\cos 2\pi f_c t dt\right]. \qquad (6.2.12)$$

Видим, что условие оптимальности (6.2.11) подразумевает использование петли для извлечения оценки, что показано на рис. 6.2.1, где ГУН – генератор, управляемый напряжением, выдающий гармоническое колебание с такой частотой и фазой, которые соответствуют уравнению (6.2.11).

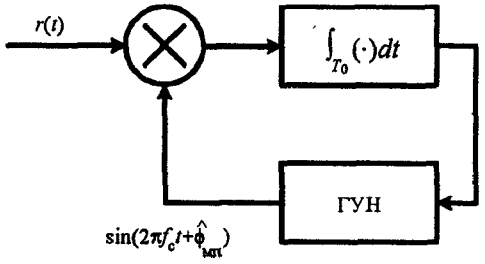


Рис. 6.2.1. Цепь ФАП для получения МП оценки фазы немодулированной несущей

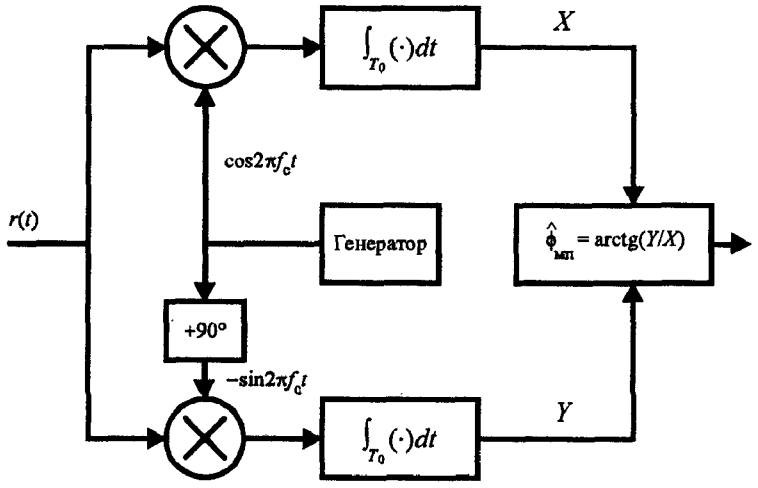


Рис. 6.2.2. МП оценка фазы немодулированной несущей на одном тактовом интервале (разовая оценка)

Петлевой фильтр состоит из интегратора, полоса пропускания которого пропорциональна обратной величине интервала интегрирования T_0. С другой стороны, (6.2.12) предполагает не петлевую реализацию, которая использует квадратурные несущие

для взаимной корреляции с $r(t)$. МП оценка $\hat{\phi}_{МП}$ определяется арктангенсом отношения этих двух выходов корреляторов, как показано на рис. 6.2.2. Заметим, что эта схема оценки выдает $\hat{\phi}_{МП}$ с неоднозначностью в π рад.

Этот пример ясно демонстрирует, что ФАП обеспечивает МП оценку фазы немодулированной несущей.

6.2.2. Модели замкнутой ФАП

ФАП в своей основе состоит из умножителя, петлевого фильтра и генератора, управляемого напряжением (ГУН), как показано на рис. 6.2.3. Если предположить, что входом цепи ФАП является косинусоида $\cos(2\pi f_c t + \phi)$, а выходом ГУН является $\sin(2\pi f_c t + \hat{\phi})$, где $\hat{\phi}$ представляет оценку ϕ, то произведение этих двух сигналов

$$e(t) = \cos(2\pi f_c t + \phi)\sin(2\pi f_c t + \hat{\phi}) = \tfrac{1}{2}\sin(\hat{\phi} - \phi) + \tfrac{1}{2}\sin(4\pi f_c t + \phi + \hat{\phi}). \qquad (6.2.13)$$

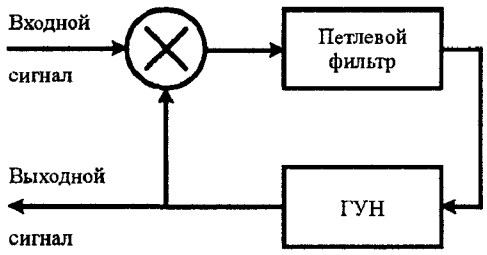


Рис. 6.2.3. Базовые элементы замкнутой петли автоподстройки фазы (ФАП)

Петлевой фильтр является низкочастотным фильтром, который пропускает только низкочастотную составляющую $\frac{1}{2}\sin(\hat{\phi} - \phi)$ и устраняет компоненту с удвоенной частотой несущей $2f_c$. Этот фильтр обычно выбирается так, чтобы иметь относительно простую передаточную функцию

$$G(s) = \frac{1 + \tau_2 s}{1 + \tau_1 s}, \qquad (6.2.14)$$

где τ_1 и τ_2 – расчётные параметры ($\tau_1 >> \tau_2$), которые управляют полосой петли. Фильтр более высокого порядка, содержащий дополнительные полюсы, можно использовать, если необходимо получить лучшую характеристику петли.

Выход петлевого фильтра обеспечивает управляющее напряжение $\upsilon(t)$ для ГУН. ГУН является в принципе генератором гармонического сигнала с мгновенной фазой, определяемой так:

$$2\pi f_c t + \hat{\phi}(t) = 2\pi f_c t + K\int_{-\infty}^{t} \upsilon(t)\, d\tau, \qquad (6.2.15)$$

где K – постоянная величина с размерностью радиан/вольт. Следовательно,

$$\hat{\phi}(t) = K\int_{-\infty}^{t} \upsilon(t)\, d\tau. \qquad (6.2.16)$$

Пренебрегая слагаемым с удвоенной частотой, которая получается от умножения входного сигнала с выходом ГУН, можем свести схему ФАП к эквивалентной замкнутой петлевой модели, показанной на рис. 6.2.4. Синусоидальная зависимость от разности фазы

$\phi - \hat{\phi}$ делает эту систему нелинейной и, как следствие, затрудняет анализ её качества в присутствии шума. Но всё же она поддается математическому анализу для некоторых простых петлевых фильтров.

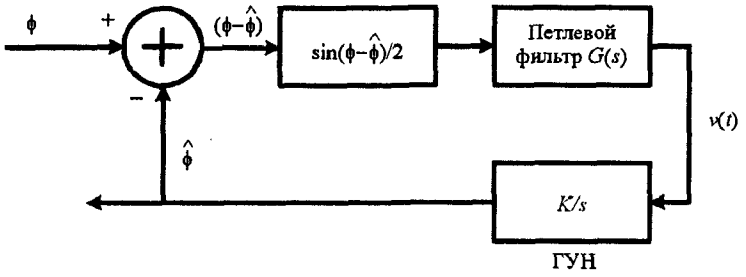


Рис. 6.2.4. Модель замкнутой петли ФАП

При нормальной работе, когда петля отслеживает фазу пришедшей несущей, фазовая ошибка $\phi - \hat{\phi}$ мала, и, следовательно,

$$\sin\left(\hat{\phi} - \phi\right) \approx \hat{\phi} - \phi . \qquad (6.2.17)$$

С этой аппроксимацией схема с ФАП на рис. 6.2.4 получается линейной и характеризуется передаточной функцией для замкнутой петли

$$H(s) = \frac{KG(s)/s}{1 + KG(s)/s}, \qquad (6.2.18)$$

где множитель $\frac{1}{2}$ включён в параметр усиления K. Подставив из (6.2.14) $G(s)$ в (6.2.18), получим

$$H(s) = \frac{1 + \tau_2 s}{1 + \left(\tau_2 + 1/K\right)s + \left(\tau_1/K\right)s^2} . \qquad (6.2.19)$$

Следовательно, замкнутая петлевая система для линеаризованной ФАП является системой второго порядка, если $G(s)$ определяется (6.2.14). Параметр τ_2 управляет положением нуля, в то время как K и τ_1 используются для управления положением полюсов замкнутой системы. Принято выразить знаменатель $H(s)$ в стандартной форме

$$D(s) = s^2 + 2\zeta\omega_n s + \omega_n^2, \qquad (6.2.20)$$

где ζ называют петлевым множителем затухания, а ω_n является резонансной частотой петли. Через параметры петли $\omega_n = \sqrt{K/\tau_1}$ и $\zeta = (\tau_2 + 1/K)/2\omega_n$ передаточная функция замкнутой петли выражается следующим образом:

$$H(s) = \frac{\left(2\zeta\omega_n - \omega_n^2/K\right)s + \omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} . \qquad (6.2.21)$$

Односторонняя эквивалентная шумовая полоса петли равна

$$B_{\text{экв}} = \frac{\tau_2^2\left(1/\tau_2^2 + K/\tau_1\right)}{4\left(\tau_2 + 1/K\right)} = \frac{1 + \left(\tau_2\omega_n\right)^2}{8\zeta\omega_n} . \qquad (6.2.22)$$

Амплитудно-частотные характеристики $20\log|H(\omega)|$ как функции нормированной частоты ω/ω_n иллюстрируются рис. 6.2.5 с множителем затухания ζ как параметром и при $\tau_1 >> 1$. Заметим, что $\zeta = 1$ ведет к критическому затуханию характеристики петли. $\zeta < 1$ ведет к поддемпинговой характеристике, а $\zeta > 1$ ведет к наддемпинговой характеристике.

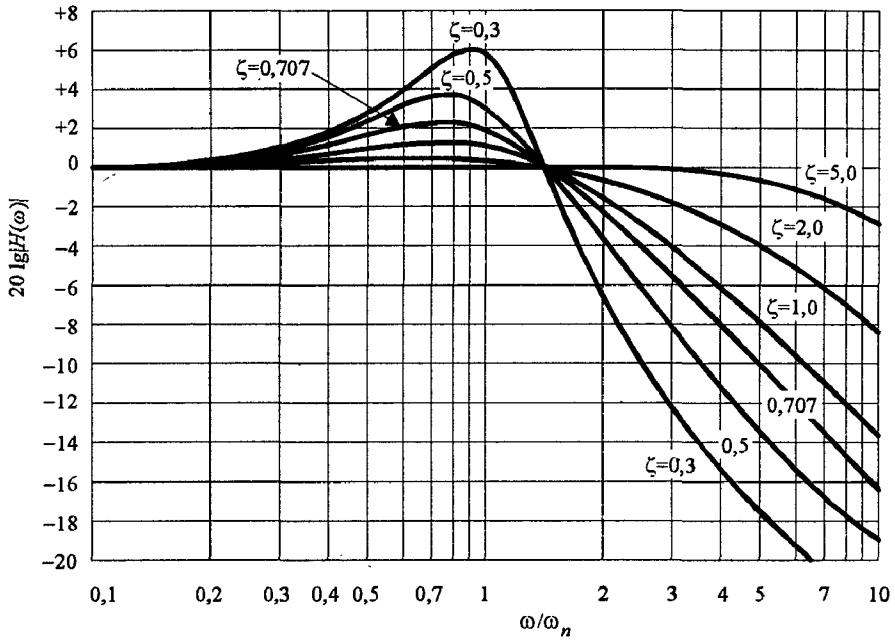


Рис. 6.2.5. Амплитудно-частотные характеристики петли второго порядка [Phaselock Techniques, 2-е издание, F.M.Gardner, © 1979]

На практике выбор полосы для ФАП включает компромисс между скоростью отслеживания и остаточным шумом в оценке фазы, что является темой, рассматриваемой ниже. С одной стороны, желательно выбрать полосу петли достаточно широкой, чтобы отслеживать любые изменения во времени фазы принимаемой несущей. С другой стороны, широкополосная ФАП позволяет шуму в большей степени попасть в петлю, что ухудшает оценку фазы. Ниже мы оценим влияние шума на качество оценки фазы.

6.2.3. Влияние аддитивного шума на оценку фазы

Чтобы рассчитать влияние шума на оценку фазы несущей, предположим, что шум на входе ФАП узкополосный. Для этого анализа мы предположим, что ФАП отслеживает синусоидальный сигнал вида

$$s(t) = A_c \cos\bigl(2\pi f_c t + \phi(t)\bigr), \tag{6.2.23}$$

который искажается узкополосным аддитивным шумом

$$n(t) = x(t)\cos 2\pi f_c t - y(t)\sin 2\pi f_c t\,. \tag{6.2.24}$$

Синфазная и квадратурная компоненты шума предполагаются статистически независимыми стационарными гауссовскими процессами с (двухсторонней) спектральной плотностью мощности $\frac{1}{2}N_0$ Вт/Гц. Используя простые тригонометрические соотношения, шум (6.2.24) можно выразить так

$$n(t) = n_c(t)\cos\bigl[2\pi f_c t + \phi(t)\bigr] - n_s(t)\sin\bigl[2\pi f_c t + \phi(t)\bigr], \tag{6.2.25}$$

где

$$\begin{aligned} n_c(t) &= x(t)\cos\phi(t) + y(t)\sin\phi(t)\,, \\ n_s(t) &= -x(t)\sin\phi(t) + y(t)\cos\phi(t)\,. \end{aligned} \tag{6.2.26}$$

Заметим, что

$$n_c(t) + jn_s(t) = \bigl[x(t) + jy(t)\bigr]e^{-j\phi(t)},$$

так что квадратурные компоненты $n_c(t)$ и $n_s(t)$ имеют точно такие же статистические характеристики, как $x(t)$ и $y(t)$.

Если $s(t)+n(t)$ умножается на выход ГУН, а слагаемым с удвоенной частотой несущей можно пренебречь, на вход петлевого фильтра действует зашумлённый сигнал

$$e(t) = A_c \sin\Delta\phi + n_c(t)\sin\Delta\phi - n_s(t)\cos\Delta\phi = A_c \sin\Delta\phi + n_1(t), \qquad (6.2.27)$$

где, по определению, $\Delta\phi = \phi - \hat{\phi}$ - фазовая ошибка. Таким образом, мы имеем эквивалентную модель для ФАП с аддитивным шумом, как показано на рис. 6.2.6.

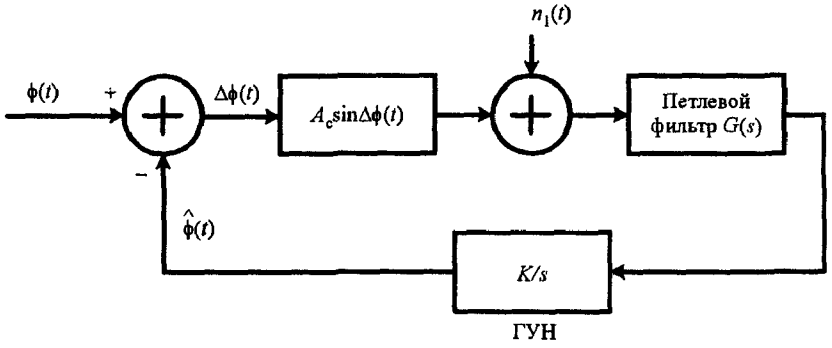


Рис. 6.2.6. Эквивалентная модель замкнутой петли ФАП с аддитивиым шумом

Если мощность $P_c = \frac{1}{2}A_c^2$ приходящего сигнала намного больше, чем мощность шума, мы можем линеаризовать ФАП и, таким образом, легко определить влияние аддитивного шума на качество оценки $\hat{\phi}$. При этих условиях модель линеаризованной ФАП с аддитивным шумом иллюстрируется рис. 6.2.7.

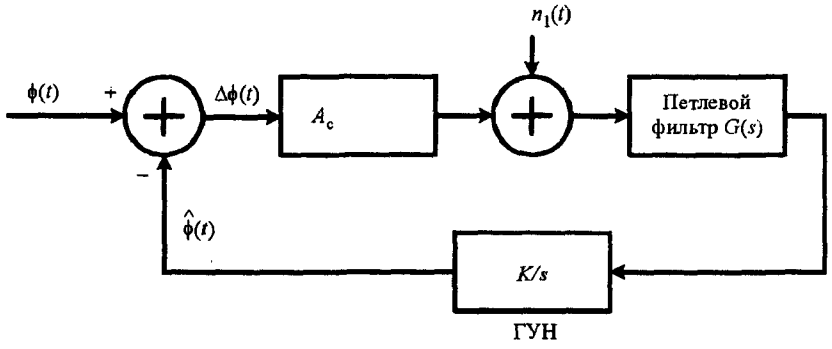


Рис. 6.2.7. Линеаризованная модель замкнутой петли ФАП с аддитивным шумом

Заметим, что параметр усиления A_c можно нормировать к 1, выполнив умножение шумовых слагаемых на $1/A_c$; тогда шумовая компонента становится равной

$$n_2(t) = \frac{n_c(t)}{A_c}\sin\Delta\phi - \frac{n_s(t)}{A_c}\cos\Delta\phi. \qquad (6.2.28)$$

Поскольку шум $n_2(t)$ на входе петли является аддитивным, дисперсия фазовой ошибки $\Delta\phi$, которая является также дисперсией фазы выхода ГУН, равна

$$\sigma_\phi^2 = \frac{N_0 B_{экв}}{A_c^2}, \tag{6.2.29}$$

где $B_{экв}$ – односторонняя эквивалентная шумовая полоса петли, определяемая (6.2.22). Заметим, что σ_ϕ^2 – просто отношение суммарной мощности шума в полосе ФАП к A_c^2. Следовательно,

$$\sigma_\phi^2 = 1/\gamma_L, \tag{6.2.30}$$

где γ_L определено как отношение сигнал/шум

$$\text{ОСШ} \equiv \gamma_L = \frac{A_c^2}{N_0 B_{экв}}. \tag{6.2.31}$$

Выражение для дисперсии σ_ϕ^2 ошибки фазы на входе ГУН относится к случаю, когда ОСШ достаточно велико, так что приемлема линейная модель ФАП. Точный анализ нелинейной модели ФАП поддается математической обработке, когда $G(s) = 1$, что относится к петле первого порядка. В этом случае можно получить ФПВ для фазовой ошибки (см. Витерби, 1966), и она имеет вид

$$p(\Delta\phi) = \frac{\exp(\gamma_L \cos\Delta\phi)}{2\pi I_0(\gamma_L)}, \tag{6.2.32}$$

где γ_L – ОСШ, даваемое (6.2.31) с $B_{экв}$, которая является соответствующей шумовой полосой петли первого порядка, а $I_0(\cdot)$ – модифицированная функция Бесселя нулевого порядка.

Из выражения для $p(\Delta\phi)$ можно получить величину дисперсии для фазовой ошибки для петли первого порядка. График дан на рис 6.2.8 как функция от $1/\gamma_L$. Для сравнения показан результат, полученный для линеаризованной модели ФАП. Заметим, что дисперсия для линейной модели тесно примыкает к точной дисперсии для $\gamma_L > 3$. Следовательно, линейная модель подходит для практических целей.

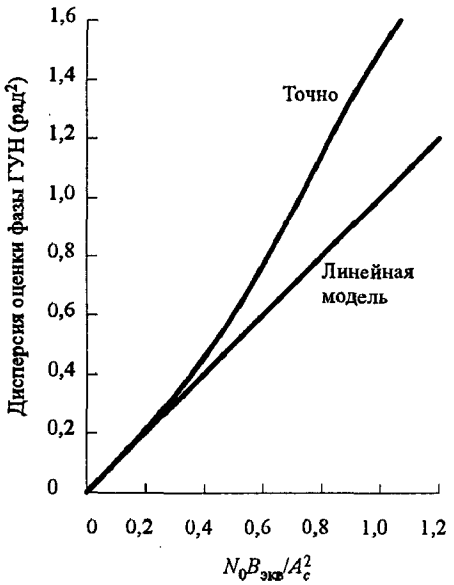


Рис. 6.2.8. Сравнение дисперсии фазы на выходе ГУН при точной и приближённой (линеаризованной) модели ФАП 1 порядка. [А.Дж.Витерби, Принципы когерентной связи; © 1966 McGraw-Hill Book Company]

Приближённый анализ статистических характеристик фазовой ошибки для нелинейной ФАП также проведён. Особую важность имеет переходный процесс ФАП на начальной стадии. Другая важная проблема – поведение петли при низких ОСШ. Известно, например, что, когда ОСШ на входе ФАП понижается ниже определенной величины, наблюдается быстрое ухудшение качества ФАП. Начинаются срывы синхронизма, приводящие к импульсному шуму, проявляющемуся как щелчки, который приводит к потере качества ФАП. Результаты по этим вопросам можно найти в книге Витерби (1966), Линдсея (1972), Линдсея и Саймона (1979), Гарднера (1979) и в некоторых статьях Гупта (1975), Линдсея и Чай (1981).

До сих пор мы рассматривали оценку фазы несущей, когда несущая не модулирована. Ниже мы рассмотрим восстановление фазы несущей, когда несущая несёт информацию

6.2.4. Петли, управляемые решениями

Когда сигнал $s(t;\phi)$ несёт на себе информационную последовательность $\{I_n\}$, встаёт проблема максимизации (6.2.9) или (6.2.10). В этом случае мы можем принять один из двух подходов: или мы предположим, что $\{I_n\}$ известно на приёме, или мы будем трактовать $\{I_n\}$ как случайную последовательность и выполним усреднение по её статистике.

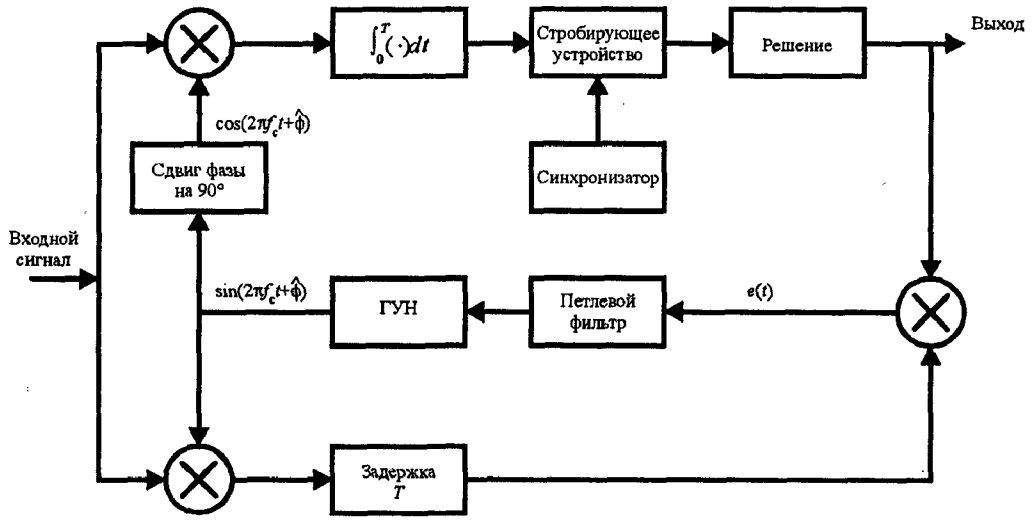


Рис. 6.2.9. Восстановление несущей в системе ФАП с обратной связью по решению

При оценивании параметра в условиях управления решениями мы считаем, что информационная последовательность $\{I_n\}$ на интервале наблюдения оценена и в отсутствие ошибок демодуляции $\tilde{I}_n = I_n$, где $\tilde{I}_n$ означает продетектированное информационное значение одного символа I_n В этом случае $s(t;\phi)$ в целом известна, за исключением фазы несущей. Оценивание фазы, управляемое решениями (об информационных символах), было впервые описано Прокисом и др. (1964).

Для конкретности, рассмотрим оценивание фазы, управляемое решениями (ОУР, DDE), при линейной модуляции, когда принимаемый низкочастотный эквивалентный сигнал можно выразить так:

$$r(t) = e^{-j\phi}\sum_n I_n g(t-nT) + z(t) = s_l(t)e^{-j\phi} + z(t), \qquad (6.2.33)$$

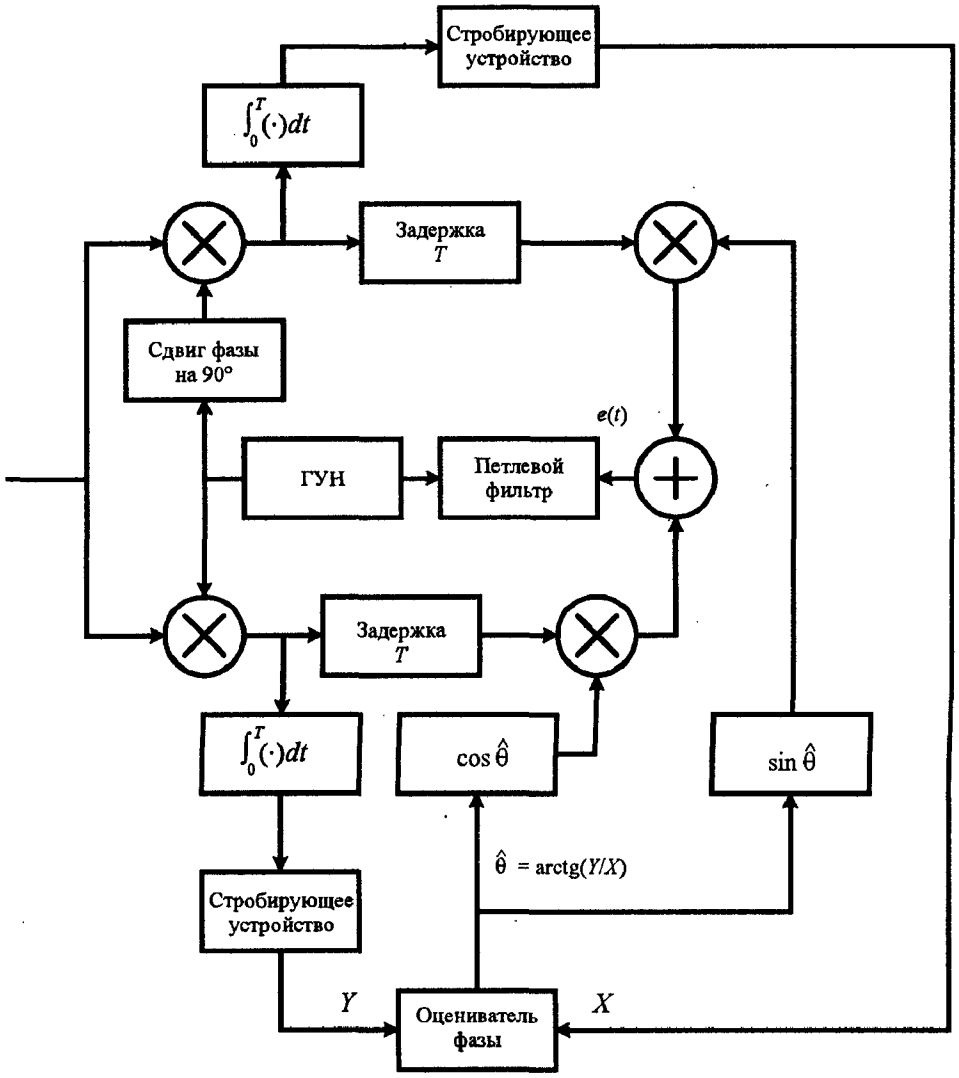


Рис. 6.2.10. Восстановление несущей в M-позиционной системе ФМ с использованием обратной связи по решеиию в системе ФАП

где $s_l(t)$ – известный сигнал, если последовательность $\{I_n\}$ считать известной. Функция правдоподобия и соответствующий логарифм функции правдоподобия равны

$$\Lambda(\phi) = C\exp\left\{\mathrm{Re}\left[\frac{1}{N_0}\int_{T_0} r(t)s_l^*(t)e^{j\phi}dt\right]\right\}, \qquad (6.2.34)$$

$$\Lambda_L(\phi) = \mathrm{Re}\left\{\left[\frac{1}{N_0}\int_{T_0} r(t)s_l^*(t)dt\right]e^{j\phi}\right\}. \qquad (6.2.35)$$

Подставив выражение для $s_l(t)$ в (6.2.35) и предположив, что интервал наблюдения $T_0 = KT$, где K – положительное целое, получим .

$$\Lambda_L(\phi) = \mathrm{Re}\left\{e^{j\phi}\frac{1}{N_0}\sum_{n=0}^{K-1} I_n^*\int_{nT}^{(n+1)T} r(t)g^*(t-nT)dt\right\} = \mathrm{Re}\left\{e^{j\phi}\frac{1}{N_0}\sum_{n=0}^{K-1} I_n^* y_n\right\}, \qquad (6.2.36)$$

где, по определению,

$$y_n = \int_{nT}^{(n+1)T} r(t)g^*(t-nT)\,dt\,. \qquad (6.2.37)$$

Заметим, что y_n – это выход согласованного фильтра на n-м сигнальном интервале. МП оценку для ϕ легко найти из (6.2.36), дифференцируя логарифм функции правдоподобия

$$\Lambda_L(\phi) = \mathrm{Re}\left(\frac{1}{N_0}\sum_{n=0}^{K-1} I_n^* y_n\right)\cos\phi - \mathrm{Im}\left(\frac{1}{N_0}\sum_{n=0}^{K-1} I_n^* y_n\right)\sin\phi$$

по ϕ и приравнивая результат нулю. Таким образом, получаем

$$\hat{\phi}_{\text{МП}} = -\mathrm{arctg}\left[\mathrm{Im}\left(\frac{1}{N_0}\sum_{n=0}^{K-1} I_n^* y_n\right) \Big/ \mathrm{Re}\left(\frac{1}{N_0}\sum_{n=0}^{K-1} I_n^* y_n\right)\right]. \qquad (6.2.38)$$

Мы назвали $\hat{\phi}_{\text{МП}}$ в (6.2.38) оценкой фазы несущей, управляемой решениями (ОУР) (или с обратной связью по решению). Легко показать (задача 6.10), что среднее значение $\hat{\phi}_{\text{МП}}$ равно ϕ, так что оценка несмещенная. Далее можно показать, что ФПВ для $\hat{\phi}_{\text{МП}}$ можно получить (задача 6.11), используя процедуру, описанную в разд. 5.2.7.

ФАП с обратной связью по решению, предназначенная для двухполосного сигнала вида $A\cos(2\pi f_c t + \phi)$, показана на рис. 6.2.9. Принимаемый сигнал умножается на квадратурные несущие $c_c(t)$ и $c_s(t)$, как это определено (6.2.5), создаваемые ГУН. Сигнал произведения

$$r(t)\cos\left(2\pi f_c t + \hat{\phi}\right) = \tfrac{1}{2}\left[A(t) + n_c(t)\right]\cos\Delta\phi - \tfrac{1}{2}n_s(t)\sin\Delta\phi +$$

$$+ \text{слагаемое с двойной частотой} \qquad (6.2.39)$$

используется для восстановления информации, имеющейся в $A(t)$. Детектор выносит решение о принимаемом символе каждые T секунд. Таким образом, в отсутствие ошибок решения реконструируется сигнал $A(t)$, свободный от шума. Этот реконструированный сигнал используется для перемножения с результатом второго квадратурного умножителя, который задерживается на T секунд, чтобы дать время демодулятору вынести решение. Таким образом, входом на петлевой фильтр при отсутствии ошибок решения является сигнал ошибки

$$e(t) = \tfrac{1}{2}A(t)\left\{\left[A(t) + n_c(t)\right]\sin\Delta\phi - n_s(t)\cos\Delta\phi\right\} +$$

$$+ \text{слагаемое с двойной частотой} =$$

$$= \tfrac{1}{2}A^2(t)\sin\Delta\phi + \tfrac{1}{2}A(t)\left[n_c(t)\sin\Delta\phi - n_s(t)\cos\Delta\phi\right] +$$

$$+ \text{слагаемое с двойной частотой.} \qquad (6.2.40)$$

Петлевой фильтр низкочастотный, и, следовательно, он подавляет слагаемые с удвоенной частотой в $e(t)$. Желательная компонента $A^2(t)\sin\Delta\phi$ содержит фазовую ошибку для управления петлёй.

Для случая M-позиционной системы ФМ, ФАП с обратной связью по решению (ФАП с ОСР) имеет конфигурацию, показанную на рис. 6.2.10. Принимаемый сигнал демодулируется, чтобы получить оценку фазы

$$\hat{\theta}_m = \frac{2\pi}{M}(m-1),$$

которая при отсутствии ошибки решения является фазой переданного сигнала. Два выхода квадратурных умножителей задерживаются на длительность символа T и умножаются на $\cos\theta_m$ и $\sin\theta_m$, чтобы получить

$$r(t)\cos\left(2\pi f_c t+\hat{\phi}\right)\sin\theta_m=\tfrac{1}{2}\left[A\cos\theta_m+n_c(t)\right]\sin\theta_m\cos\left(\phi-\hat{\phi}\right)-$$
$$-\tfrac{1}{2}\left[A\sin\theta_m+n_s(t)\right]\sin\theta_m\sin\left(\phi-\hat{\phi}\right)+$$

+ слагаемое с двойной частотой; (6.2.41)

$$r(t)\sin\left(2\pi f_c t+\hat{\phi}\right)\cos\theta_m=-\tfrac{1}{2}\left[A\cos\theta_m+n_c(t)\right]\cos\theta_m\sin\left(\phi-\hat{\phi}\right)-$$
$$-\tfrac{1}{2}\left[A\sin\theta_m+n_s(t)\right]\cos\theta_m\cos\left(\phi-\hat{\phi}\right)+$$

+ слагаемое с двойной частотой.

Два сигнал складываются, чтобы генерировать сигнал ошибки

$$e(t)=-\tfrac{1}{2}A\sin\left(\phi-\hat{\phi}\right)+\tfrac{1}{2}n_c(t)\sin\left(\phi-\hat{\phi}-\theta_m\right)+\tfrac{1}{2}n_s(t)\cos\left(\phi-\hat{\phi}-\theta_m\right)+$$

+ слагаемое с двойной частотой. (6.2.42)

Этот сигнал ошибки является входом петлевого фильтра, который обеспечивает сигнал управления для блока ГУН.

Мы видим, что две квадратурные шумовые компоненты в (6.2.42) возникают как аддитивные слагаемые. Здесь нет слагаемых, определяемых произведением двух шумовых компонент, как в устройстве с нелинейной характеристикой M-й степени, описанном в следующем разделе. Следовательно, здесь нет дополнительной потери мощности, которая связана с ФАП и ОСР.

Эта M-фазная отслеживающая петля имеет фазовую неоднозначность $360°/M$, которая заставляет использовать дифференциальное кодирование информационной последовательности до передачи и дифференциальное декодирование принимаемой последовательности после демодуляции для восстановления информации.

МП оценка, определяемая (6.2.38), также используется для КАМ. МП оценка для ОКФМ также получается (задача 6.12) путём максимизации функции правдоподобия в (6.2.35) с сигналом $s_l(t)$, определённым так:

$$s_l(t)=\sum_n I_n g(t-nT)+j\sum_n J_n g\left(t-nT-\frac{1}{2}T\right), \qquad (6.2.43)$$

где $I_n=\pm1$ и $J_n=\pm1$.

В заключение мы хотим также упомянуть, что восстановление фазы несущей для сигналов НФМ можно осуществить схемой, управляемой решениями, используя ФАП. Посредством оптимального демодулятора для сигналов НФМ, который был описан в разд. 5.3, можем генерировать сигнал ошибки, который фильтруется в петлевом фильтре, чей выход управляет ФАП.

6.2.5. Петли, не управляемые решениями

Вместо использования схемы, управляемой решениями для получения оценки фазы, можно трактовать данные как случайные величины и просто усреднить $\Lambda(\phi)$ по этим случайным величинам до её максимизации. Чтобы выполнить такое усреднение, можно использовать или действительную функцию распределения вероятностей данных, если она известна, или можно предположить некоторое распределение вероятностей, которое является подходящим приближением для правильного распределения. Следующие примеры демонстрируют первый подход.

Пример 6.2.2. Предположим, что сигнал двоичной линейной модуляции $s(t)$ является вещественным. Тогда на сигнальном интервале мы можем написать

$$s(t) = A\cos 2\pi f_c t, \quad 0 \le t \le T,$$

где $A = \pm 1$ с равными вероятностями. Ясно, что ФПВ для A равна

$$p(A) = \tfrac{1}{2}\delta(A-1) + \tfrac{1}{2}\delta(A+1).$$

Теперь функция правдоподобия $\Lambda(\phi)$, определяемая (6.2.9), является условной при заданном значении A, и её следует усреднять по этим двум значениям. Таким образом,

$$\overline{\Lambda}(\phi) = \int_{-x}^{x} \Lambda(\phi) p(A) dA = \frac{1}{2}\exp\left[\frac{2}{N_0}\int_0^T r(t)\cos(2\pi f_c t + \phi)dt\right] +$$

$$+\frac{1}{2}\exp\left[-\frac{2}{N_0}\int_0^T r(t)\cos(2\pi f_c t + \phi)dt\right] = \mathrm{ch}\left[\frac{2}{N_0}\int_0^T r(t)\cos(2\pi f_c t + \phi)dt\right],$$

а соответствующий логарифм функции правдоподобия

$$\overline{\Lambda}_L(\phi) = \ln \mathrm{ch}\left[\frac{2}{N_0}\int_0^T r(t)\cos(2\pi f_c t + \phi)dt\right]. \qquad (6.2.44)$$

Если продифференцируем $\overline{\Lambda}_L(\phi)$ и приравняем производную нулю, получим МП оценку для фазы, не управляемую решениями (оценивание, не управляемое решениями – ОНУР, NDDE). К сожалению, функциональное отношение в (6.2.44) существенно нелинейно, и, следовательно, точное решение трудно получить. С другой стороны, возможна аппроксимация. В частности,

$$\ln \mathrm{ch}\, x = \begin{cases} x^2/2 & (|x| << 1), \\ |x| & (|x| >> 1). \end{cases} \qquad (6.2.45)$$

С этой аппроксимацией решение для ϕ получается в трактуемом виде.

В этом примере мы усреднили по двум возможным значениям амплитуды информационных символов. Если информационные символы M-позиционные, а M велико, операция усреднения содержит нелинейные функции высокого порядка от параметра, который оценивается. В этом случае мы можем упростить проблему, предположив, что амплитуды информационных символов являются непрерывными случайными величинами. Например, мы можем предположить, что они подчиняются гауссовскому распределению с нулевым средним. Следующие примеры иллюстрируют эту аппроксимацию и результирующую форму для усредненной функции правдоподобия.

Пример 6.2.3. Рассмотрим тот же сигнал, что в примере 6.2.2, но теперь предположим, что амплитуда A является гауссовской, с нулевым средним и единичной дисперсией, т.е.

$$p(A) = \frac{1}{\sqrt{2\pi}} e^{-A^2/2}.$$

Если мы усредним $\Lambda(\phi)$ по заданной ФПВ A, получим для усреднённой функции правдоподобия

$$\overline{\Lambda}(\phi) = C\exp\left\{\left[\frac{2}{N_0}\int_0^T r(t)\cos(2\pi f_c t + \phi)dt\right]^2\right\}, \qquad (6.2.46)$$

и соответствующий логарифм усреднённой функции правдоподобия

$$\overline{\Lambda}_L(\phi) = \left[\frac{2}{N_0}\int_0^T r(t)\cos(2\pi f_c t + \phi)dt\right]^2. \qquad (6.2.47)$$

Теперь мы можем получить МП оценку для ϕ путём дифференцирования $\overline{\Lambda}_L(\phi)$ и приравнивания результата нулю.

Интересно отметить, что логарифм усреднённой функции правдоподобия является квадратичным при гауссовском предположении, а также то, что он имеет квадратичную аппроксимацию, определённую (6.2.45), для малых значений взаимной корреляции $r(t)$ и $s(t;\phi)$. Другими словами, если взаимная корреляция на одном интервале мала, гауссовское предположение для распределения амплитуд информационных символов даёт хорошую аппроксимацию для логарифма усреднённой функции правдоподобия .

С точки зрения этих результатов мы можем использовать гауссовскую аппроксимацию на все символы на интервале $T_0 = KT$. Конкретнее, предположим, что K информационных символов статистически независимы и одинаково распределены. При усреднении функции правдоподобия $\Lambda(\phi)$ по гауссовской ФПВ на каждом из K символов на интервале $T_0 = KT$ получаем результат

$$\overline{\Lambda}(\phi) = C\exp\left\{\sum_{n=0}^{K-1}\left[\frac{2}{N_0}\int_{nT}^{(n+1)T} r(t)\cos(2\pi f_c t + \phi)dt\right]^2\right\}. \qquad (6.2.48)$$

Если мы возьмём логарифм от (6.2.48), продифференцируем его и приравняем результат нулю, получим условие для МП оценки в виде

$$\sum_{n=0}^{K-1}\int_{nT}^{(n+1)T} r(t)\cos\left(2\pi f_c t + \hat{\phi}\right)dt\int_{nT}^{(n+1)T} r(t)\sin\left(2\pi f_c t + \hat{\phi}\right)dt = 0. \qquad (6.2.49)$$

Хотя это уравнение можно преобразовывать и дальше, уже его настоящая форма предполагает схему петлевого отслеживания, показанного на рис. 6.2.11. Эта петля похожа на петлю Костаса, которая будет описана ниже. Заметим, что произведение двух сигналов от интеграторов устраняет знак несущей, обусловленной информационными символами.

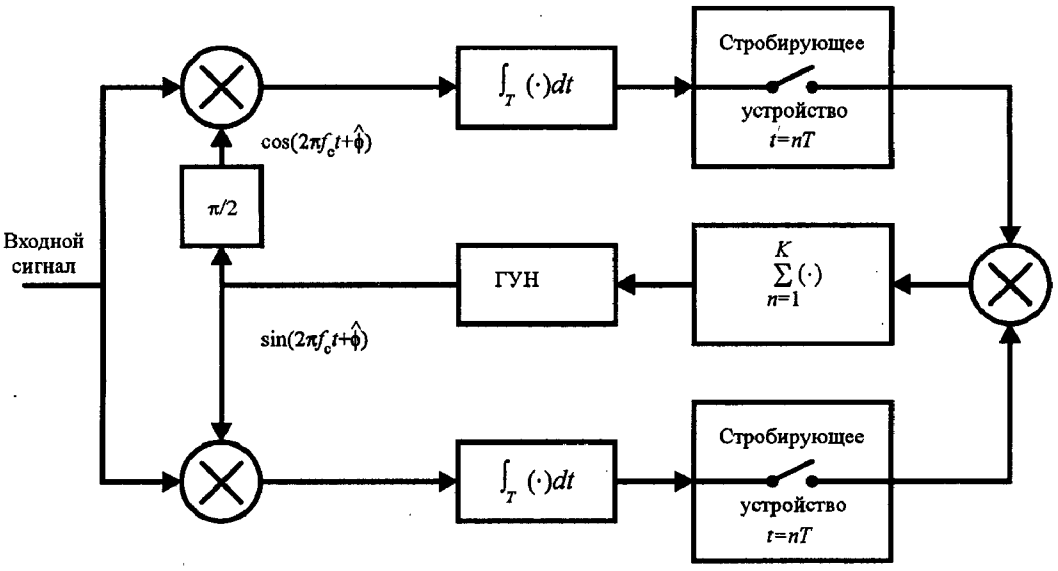


Рис. 6.2.11. ФАП, не использующая решения детектора для оценивания фазы АМ сигналов

Сумматор играет роль петлевого фильтра. В петлевой схеме отслеживания сумматор можно реализовать или как цифровой фильтр со скользящим окном (сумматор), или как низкочастотный цифровой фильтр с экспоненциальным взвешиванием последних данных.

Подобным образом можно осуществить МП оценку фазы, не управляемую решениями, для КАМ и многопозиционной ФМ.

Исходная операция сводится к усреднению функции правдоподобия (6.2.91) по

статистике параметров данных. Здесь снова мы можем использовать гауссовскую аппроксимацию (двухмерное гауссовское распределение для комплексных информационных символов) или усреднение по информационной последовательности.

Квадратичная петля. Квадратичная петля – это петля, не управляемая решениями, которая широко используется на практике для установления фазы несущей в двухполосной системе с подавленной несущей, такой как АМ. Чтобы описать её работу, рассмотрим проблему оценивания фазы несущей сигнала цифровой АМ в виде

$$s(t) = A(t)\cos\left(2\pi f_c t + \phi\right), \qquad (6.2.50)$$

где $A(t)$ несёт цифровую информацию. Заметим, что $E[s(t)] = E[A(t)] = 0$, когда сигнальные уровни распределены симметрично относительно нуля. Следовательно, усреднённое значение $s(t)$ не может дать ни одной фазокогерентной частотной компоненты, исключая несущую. Один из методов восстановления несущей от принимаемого сигнала сводится к его квадратированию и, следовательно, к генерированию частотной компоненты $2f_c$, которую можно использовать для образования фазозамкнутой петли (ФЗП), настроенной на частоту $2f_c$. Этот метод иллюстрируется блок-схемой, показанной на рис. 6.2.12.

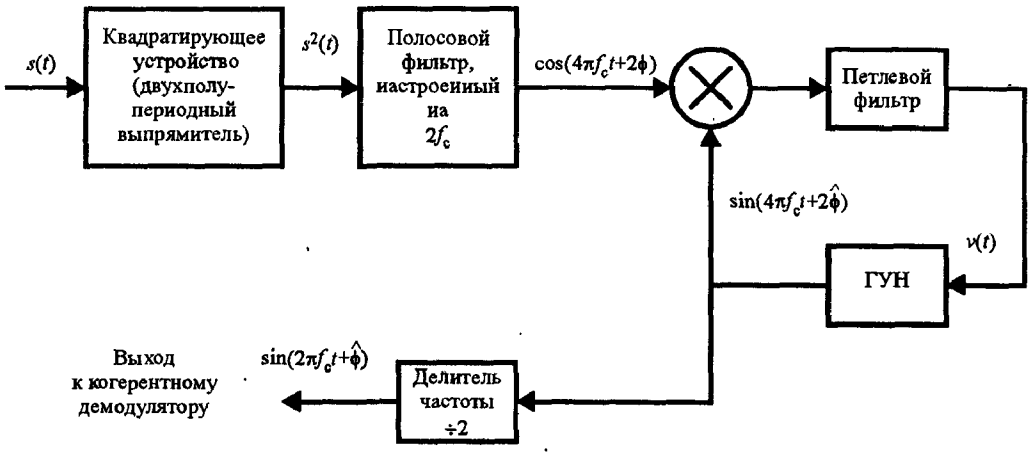


Рис. 6.2.12. Восстановлеиие несущей с использованием квадратирующего устройства

Выход квадратичного устройства равен

$$s^2(t) = A^2(t)\cos^2\left(2\pi f_c t + \phi\right) = \tfrac{1}{2}A^2(t) + \tfrac{1}{2}A^2(t)\cos\left(4\pi f_c t + 2\phi\right). \qquad (6.2.51)$$

Поскольку модулированный сигнал является циклостационарным случайным процессом, математическое ожидание от $s^2(t)$ равно

$$E\left[s^2(t)\right] = \tfrac{1}{2}E\left[A^2(t)\right] + \tfrac{1}{2}E\left[A^2(t)\right]\cos\left(4\pi f_c t + 2\phi\right). \qquad (6.2.52)$$

Следовательно, имеется мощность на частоте $2f_c$.

Когда выход квадратирующего устройства проходит через полосовой фильтр, настроенный на удвоенную частоту в (6.2.51), среднее значение на выходе фильтра – это синусоида с частотой $2f_c$, фазой 2ϕ и амплитудой $\frac{1}{2}E\left[A^2(t)\right]H\left(2f_c\right)$, где $H\left(2f_c\right)$ – усиление фильтра на частоте $f = 2f_c$. Таким образом, квадратирующее устройство образует периодическую компоненту от входного сигнала $s(t)$. По существу, квадратирование уничтожает знак информации, содержащейся в $A(t)$ и таким образом приводит к фазо-

когерентной частотной компоненте на удвоенной частоте несущей. Фильтруемая компонента на частоте $2f_c$ затем используется для управления ФЗП.

Операция квадратирования ведет к обогащению шума, что увеличивает уровень шумовой мощности на входе ФАП и ведёт к увеличению дисперсии фазовой ошибки.

Чтобы разобраться с этим вопросом, допустим, что на вход квадратирующего звена поступает сигнал $s(t)+n(t)$, где $s(t)$ определено (6.2.50), а $n(t)$ представляет полосовой аддитивный гауссовский шумовой процесс. При квадратировании $s(t)+n(t)$ получаем

$$y(t) = s^2(t) + 2s(t)n(t) + n^2(t), \qquad (6.2.53)$$

где $s^2(t)$ – желательная сигнальная компонента, а две остальные компоненты – это слагаемые сигнал × шум и шум × шум. Вычислив автокорреляционную функцию и спектральную плотность мощности этих двух шумовых компонент, можно *легко показать, что обе компоненты имеют спектральную плотность мощности в частотной полосе, сосредоточенной вблизи* $2f_c$. Следовательно, полосовой фильтр с полосой $B_{\text{ПФ}}$, центрированной на частоте $2f_c$, который создает желательные синусоидальные компоненты сигнала, управляющие ФЗП, также пропускает шум, обусловленный двумя слагаемыми. Поскольку полоса петли рассчитывается так, чтобы быть существенно меньшей, чем полоса $B_{\text{ПФ}}$ полосового фильтра, суммарный спектр шума на входе ФЗП можно аппроксимировать константой на частотах внутри полосы петли. Такая аппроксимация позволяет нам получить простое выражение для дисперсии фазовой ошибки

$$\sigma_\phi^2 = 1/\gamma_L S_L, \qquad (6.2.54)$$

где S_L названа квадратичными потерями и определяется так:

$$S_L = \left(1 + \frac{B_{\text{ПФ}}/2B_{\text{экв}}}{\gamma_L}\right)^{-1}. \qquad (6.2.55)$$

Поскольку $S_L < 1$, S_L^{-1} определяет увеличение дисперсии фазовой ошибки, вызванной дополнительным шумом (слагаемым шум × шум), обусловленным квадратированием. Заметим для примера, что, когда $\gamma_L = B_{\text{ПФ}}/2B_{\text{экв}}$, потери составляют 3 дБ.

В заключение заметим, что выход ГУН в схеме с квадратированием необходимо делить по частоте на 2, чтобы генерировать синхронную несущую для демодуляции сигнала. Надо подчеркнуть, что выход делителя частоты характеризуется неоднозначностью фазы на 180° относительно фазы принимаемого сигнала. Из этих соображений двоичные данные следует дифференциально кодировать до модуляции и дифференциально декодировать в приёмнике.

Петля Костаса. Другой метод восстановления фазированной несущей для двухполосной системы с подавленной несущей иллюстрируется рис. 6.2.13. Эта схема была предложена Костасом (1956) и называется *петлей Костаса*.

Принимаемый сигнал умножается на $\cos\left(2\pi f_c t + \hat{\phi}\right)$ и $\sin\left(2\pi f_c t + \hat{\phi}\right)$, которые являются выходами ГУН. Получаем два произведения:

$$y_c(t) = \left[s(t)+n(t)\right]\cos\left(2\pi f_c t + \hat{\phi}\right) = \tfrac{1}{2}\left[A(t)+n_c(t)\right]\cos\Delta\phi + \tfrac{1}{2}n_s(t)\sin\Delta\phi +$$

+ слагаемое с двойной частотой, (6.2.56)

$$y_s(t) = \left[s(t)+n(t)\right]\sin\left(2\pi f_c t + \hat{\phi}\right) = \tfrac{1}{2}\left[A(t)+n_c(t)\right]\sin\Delta\phi + \tfrac{1}{2}n_s(t)\cos\Delta\phi +$$

+ слагаемое с двойной частотой,

где фазовая ошибка $\Delta\phi = \hat{\phi} - \phi$. Слагаемое с удвоенной частотой устраняется фильтрами низких частот, расположенными после умножителей.

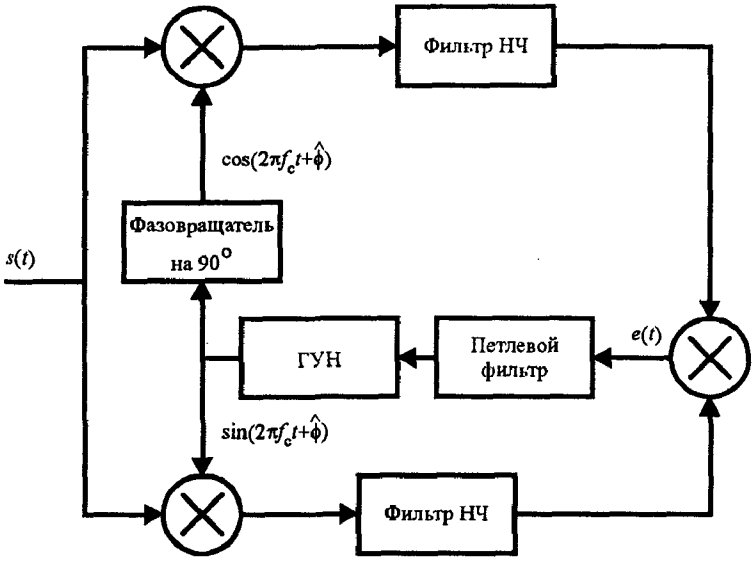


Рис. 6.2.13. Блок-схема петли Костаса

Сигнал ошибки генерируется путем умножения двух выходов фильтров низких частот. Таким образом,

$$e(t) = \tfrac{1}{8}\left\{\left[A(t)+n_c(t)\right]^2 - n_s^2(t)\right\}\sin(2\Delta\phi) - \tfrac{1}{4}n_s(t)\left[A(t)+n_c(t)\right]\cos(2\Delta\phi). \qquad (6.2.57)$$

Этот сигнал ошибки фильтруется петлевым фильтром, выход которого управляет ГУН. Читатель может убедиться в похожести петли Костаса и петли ФЗП, показанной на рис. 6.2.11.

Заметим, что сигнал ошибки на входе петлевого фильтра состоит из желательного слагаемого $A^2(t)\sin 2(\hat{\phi}-\phi)$ и слагаемых, которые содержат сигнал × шум и шум × шум. Эти слагаемые похожи на две шумовые слагаемые на входе ФЗП при использовании метода квадратирования. Действительно, если петлевой фильтр в петле Костаса идентичен тому, который используется в квадратичной петле, две эти петли эквивалентны. С учетом этого условия ФПВ ошибки фазы и качество этих двух схем ФАП идентичны.

Интересно заметить, что оптимальный ФНЧ для подавления слагаемых с двойной частотой в петле Костаса – это фильтр, согласованный с сигнальным импульсом информационной последовательности. Если согласованные фильтры используются как ФНЧ, их выходы можно стробировать с битовой скоростью в конце каждого сигнального интервала, а отсчёты в дискретных точках времени могут быть использованы для управления в петле. Использование согласованного фильтра ведет к меньшему шуму в петле.

В заключение заметим, что, как и в квадратичном ФЗП, выход ГУН в петле Костаса даёт неоднозначность фазы на 180°, что делает необходимым предварительное дифференциальное кодирование на передаче и дифференциальное декодирование после детектора.

Оценка несущей в системах с многопозиционными сигналами. Когда цифровая информация передается посредством M-позиционной модуляции фазы несущей, методы, описанные выше, можно обобщить, чтобы получить хорошую сфазированную несущую в

демодуляторе. Принимаемый M-фазный сигнал, исключая аддитивный шум, можно выразить так:

$$s(t) = A\cos\left[2\pi f_c t + \phi + \frac{2\pi}{M}(m-1)\right],\ m = 1, 2, \ldots, M, \qquad (6.2.58)$$

где $2\pi(m-1)/M$ представляет информационную компоненту фазы сигнала. Проблема *восстановления несущей сводится к устранению информационной компоненты фазы и, как* следствие, получению немодулированной несущей $\cos(2\pi f_c t + \phi)$. Один из методов, при помощи которого это можно сделать, иллюстрируется на рис. 6.2.14, который представляет обобщение петли с квадратированием. Сигнал проходит через устройство возведения в M-ю степень, которое генерирует определённое число гармоник f_c. Полосовой фильтр выбирает гармонику $\cos(2\pi M f_c t + M\phi)$ для управления ФЗП. Слагаемое информационной компоненты фазы сигнала

$$\frac{2\pi}{M}(m-1)M = 2\pi(m-1) \equiv 0 \pmod{2\pi}, \quad m = 1, 2, \ldots, M$$

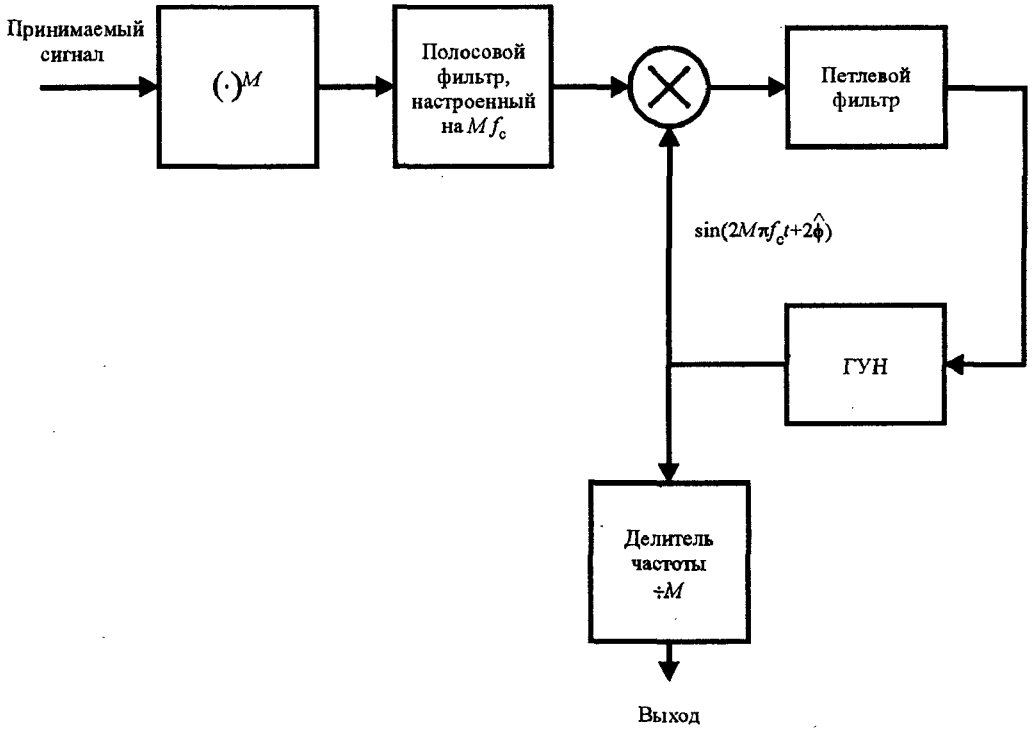


Рис. 6.2.14. Восстановление несущей с использованием устройства возведения в M-ю степень для M-позиционной ФМ

Таким образом, информация устранена. Выход ГУН – это $\sin(2\pi M f_c t + M\hat{\phi})$; этот выход делится по частоте на M для получения $\sin(2\pi f_c t + \hat{\phi})$ и сдвигается на 90° для получения $\cos(2\pi f_c t + \hat{\phi})$. Эта компонента затем подаётся на демодулятор. Нетрудно показать, что имеется неоднозначность в этих упомянутых синусоидах на $360°/M$, которую можно преодолеть дифференциальным кодированием данных на передаче и дифференциальным декодированием после демодуляции на приёме.

Как в случае квадратичной ФЗП, полиномиальная ФЗП работает в присутствии шума, возрастающего после прохождения через устройство возведения в M-ю степень, которое даёт на выходе сигнал $y(t) = [s(t) + n(t)]^M$.

Дисперсию фазовой ошибки в ФЗП, обусловленной аддитивным шумом, можно выразить в простой форме:

$$\sigma_{\hat{\phi}}^2 = \frac{S_{\text{МП}}^{-1}}{\gamma_L}, \qquad (6.2.59)$$

где γ_L – петлевое ОСШ, а $S_{\text{МП}}^{-1}$ – потери из-за возведения в M-ю степень. $S_{\text{МП}}$ рассчитали Линдсей и Саймон (1979) для $M = 4$ и 8.

Другой метод восстановления несущей в M-фазной ФМ базируется на обобщении петли Костаса. Этот метод требует умножения принимаемого сигнала на несущую с фазовым сдвигом вида

$$\sin\left[2\pi f_c t + \hat{\phi} + \frac{\pi}{M}(k-1)\right], \quad k = 1, 2, \ldots, M,$$

низкочастотной фильтрации слагаемых произведения и затем перемножения выходов низкочастотных фильтров для генерирования сигнала ошибки. Сигнал ошибки возбуждает петлевой фильтр, который создаёт сигнал управления для ГУН. Этот метод относительно сложен для применения и, как следствие, обычно не используется на практике.

Сравнение петель, управляемых и не управляемых решениями. Заметим, что фазозамкнутая петля с ОСР (ФЗПОСР, DFPLL) отличается от петли Костаса только методом очищения $A(t)$ с целью устранения модуляции. В петле Костаса каждый из двух квадратурных сигналов, используемых для очищения $A(t)$, поражается шумом. С другой стороны, квадратичная петля похожа на петлю Костела шумовыми компонентами, влияющими на оценку $\hat{\phi}$. Следовательно, ФЗПОСР предпочтительнее по качеству по отношению как к петле Костаса, так и к квадратичной петле, обеспечивая работу демодулятора при вероятности ошибки ниже 10^{-2}, причём редкие ошибки решения несущественно влияют на $\hat{\phi}$. Численные сравнения дисперсии фазовых ошибок в петле Костаса относительно тех, которые имеют место в ФЗПОСР, были выполнены Линдсеем и Саймоном (1973), и они показывает, что дисперсия в ФЗПОСР в 4...10 раз меньше для отношения сигнал/шум на бит около 0 дБ.

6.3. ОЦЕНИВАНИЕ ПРИ СИНХРОНИЗАЦИИ СИМВОЛОВ

В цифровой системе связи выход демодулятора должен периодически стробироваться со скоростью передачи символов в точно выбранных моментах $t_{ni} = mT + \tau$, где T – символьный интервал, а τ – номинальное время задержки, которое определяется временем распространения сигнала от передатчика к приёмнику. Чтобы выполнять это периодическое стробирование, требуется таймерный сигнал в приемнике. Процесс получения такого таймерного сигнала в приёмнике обычно называется *тактовой синхронизацией* (*синхронизацией символов*) или *восстановлением отсчётов времени*.

Восстановление отсчётов времени – одна из наиболее ответственных функций, которые выполняются в приёмнике синхронных цифровых систем связи. Отметим, что приёмник должен знать не только частоту $(1/T)$, с которой стробируются выходы согласованных фильтров или корректоров, но также тот момент времени, в который необходимо взять отсчёты внутри каждого символьного интервала. Положение отсчётного момента внутри символьного интервала длительностью T называют *фазой синхронизации*.

Тактовую синхронизацию можно выполнить несколькими путями. В некоторых системах связи таймеры передатчика и приёмника синхронизированы по образцовому таймеру, который обеспечивает высокостабильную сетку времени. В этом случае,

приёмник должен оценить и скомпенсировать только относительное время задержки между переданным и принятыми сигналом. Такой метод может использоваться в системах радиосвязи, которые работают в очень низкочастотном диапазоне (ниже 30 КГц), где сигналы точного времени посылаются эталонной радиостанцией.

Другой метод для достижения синхронизации символов заключается в одновременной передаче передатчиком таймерной частоты, равной или кратной $1/T$, вместе с информационным сигналом. Приёмник может просто использовать узкополосный фильтр, настроенный на переданную таймерную частоту, и таким образом извлечь таймерный сигнал для стробирования. Этот подход прост для применения. Здесь имеется, однако, несколько неудобств. Первое заключается в том, что передатчик должен отдать некоторую часть возможной мощности для передачи таймерного сигнала. Другое заключается в том, что некоторая малая часть доступной в канале полосы частот должна быть занята для передачи таймерного сигнала. Несмотря на эти неудобства, этот метод часто используется в системах телефонной связи, которые используют большие полосы частот для передачи сигналов многих пользователей. В таком случае затраты на передачу таймерного сигнала делятся на многих пользователей. Посредством такого деления потери в мощности передатчика и полосе частот уменьшаются пропорционально числу пользователей.

Таймерный сигнал можно также извлечь из принимаемого информационного сигнала. Имеется определённое число различных методов, которые можно использовать в приёмнике, чтобы достичь самосинхронизации. В этом разделе мы обсудим как метод оценивания, управляемого решениями, так и метод оценивания, не управляемого решениями.

6.3.1. Максимально правдоподобная оценка параметра задержки τ[1]

Начнём с получения МП оценки времени задержки τ. Если сигнал является базовым АМ, он может быть представлен так:

$$r(t) = s(t;\tau) + n(t), \tag{6.3.1}$$

где

$$s(t;\tau) = \sum_n I_n g(t - nT - \tau). \tag{6.3.2}$$

Как в случае МП оценки фазы, мы различаем два типа оценивателей времени: оцениватель времени, управляемый решениями и не управляемый решениями. В первых информационные символы на выходе демодулятора рассматриваются как известная переданная последовательность. В этом случае логарифм функции правдоподобия имеет вид

$$\Lambda_L(\tau) = C_L \int_{T_0} r(t)\, s(t;\tau) dt. \tag{6.3.3}$$

Подставив (6.3.2) в (6.3.3), получим

$$\Lambda_L(\tau) = C_L \sum_n I_n \int_{T_0} r(t)\, g(t - nT - \tau)\, dt = C_L \sum_n I_n y_n(\tau), \tag{6.3.4}$$

где $y_n(\tau)$ определено так:

$$y_n(\tau) = \int_{T_0} r(t)\, g(t - nT - \tau)\, dt. \tag{6.3.5}$$

Необходимое условие, при котором τ является МП оценкой, имеет вид

---

[1] Автор излагает проблемы синхронизации символов, ограничиваясь лишь вопросом оценивания единственного параметра – задержки τ. Не рассматриваются проблемы цикловой синхронизации, проблемы синхронизации при временной дискретизации сигнала и вопросы аномальных ошибок при синхронизации [ ] (прп)

$$\frac{d\Lambda_L(\tau)}{d\tau} = \sum_n I_n \frac{d}{d\tau}\int_{T_0} r(t)\, g(t - nT - \tau)\, dt = \sum_n I_n \frac{d}{d\tau}\left[y_n(\tau)\right] = 0. \qquad (6.3.6)$$

Результат (6.3.6) подсказывает реализацию посредством отслеживающей петли, показанной на рис. 6.3.1. Необходимо отметить, что суммирование в петле обеспечивается петлевым фильтром, полоса которого управляется длиной скользящего окна при суммировании. Выход петлевого фильтра подаётся на таймер, управляемый напряжением (ТУН), или на генератор, управляемый напряжением, который управляет интервалом стробирования на входе петли. Поскольку детектируемая информационная последовательность $\{I_n\}$ используется при оценке τ, оценка оказывается управляемой решениями.

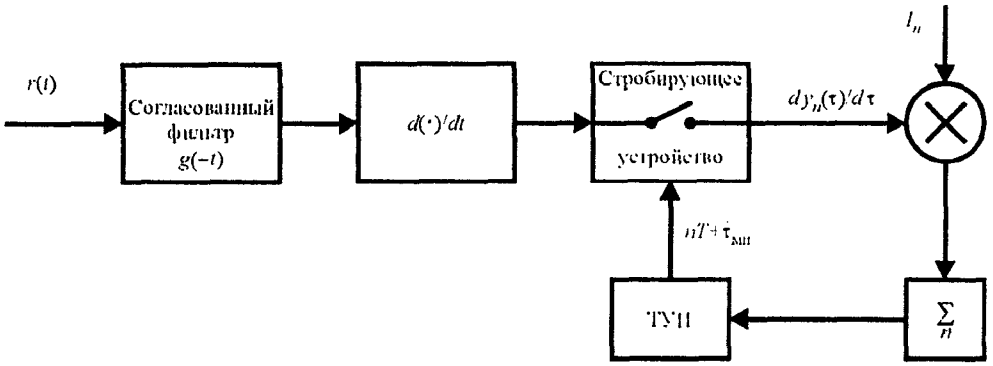


Рис. 6.3.1. МП оценивание параметра задержки, управляемое решениями, для базового сигнала АМ

Описанную выше техника для МП оценивания параметра задержки базового сигнала АМ можно распространить на форматы модулированных сигналов, таких как КАМ и ФМ непосредственно, если иметь дело с эквивалентной низкочастотной формой сигналов. Таким образом, проблема МП оценки символьной синхронизации для несущих сигналов очень похожа на формулировку проблемы для базовых сигналов АМ .

6.3.2 Оценивание параметра задержки, не управляемое решениями

Оценку параметра задержки, не управляемую решениями, можно получить путём усреднения отношения правдоподобия $\Lambda(\tau)$ с учётом ФПВ информационных символов для получения $\overline{\Lambda}(\tau)$. Затем $\overline{\Lambda}(\tau)$ или $\ln\overline{\Lambda}(\tau) = \overline{\Lambda}_L(\tau)$ дифференцируется по τ для получения условия для МП оценки $\hat{\tau}_{\text{МП}}$.

В случае двоичного (базового) АМ, где $I_n = \pm 1$ с равной вероятностью, усреднение по данным дает результат

$$\overline{\Lambda}_L(\tau) = \sum_n \ln\operatorname{ch} C y_n(\tau), \qquad (6.3.7)$$

как раз такой, как в случае оценивания фазы. Поскольку $\ln\operatorname{ch} x \approx \frac{1}{2}x^2$ для малых x, квадратичная аппроксимация

$$\overline{\Lambda}_L(\tau) = \tfrac{1}{2}C^2 \sum_n y_n^2(\tau) \qquad (6.3.8)$$

предназначается для низких отношений сигнал/шум. Для многоуровневой АМ мы можем аппроксимировать статистику информационных символов $\{I_n\}$ гауссовской ФПВ с нулевым средним и единичной дисперсией. Когда мы усредняем $\Lambda(\tau)$ по гауссовской ФПВ, то получаем $\ln\overline{\Lambda}(\tau)$ идентично $\overline{\Lambda}_L(\tau)$ в (6.3.8). Следовательно, оценку τ можно

получить дифференцированием (6.3.8). Результат является аппроксимацией для МП оценки времени задержки без управления решениями. Производная от (6.3.8) приводит к результату

$$\frac{d}{d\tau}\sum_n y_n^2(\tau) = 2\sum_n y_n(\tau)\frac{dy_n(\tau)}{d\tau} = 0, \qquad (6.3.9)$$

где $y_n(\tau)$ определено (6.3.5).

Реализация отслеживающей петли, основанная на вычислении производной $\overline{\Lambda}_L(\tau)$ согласно (6.3.9), показана на рис. 6.3.2.

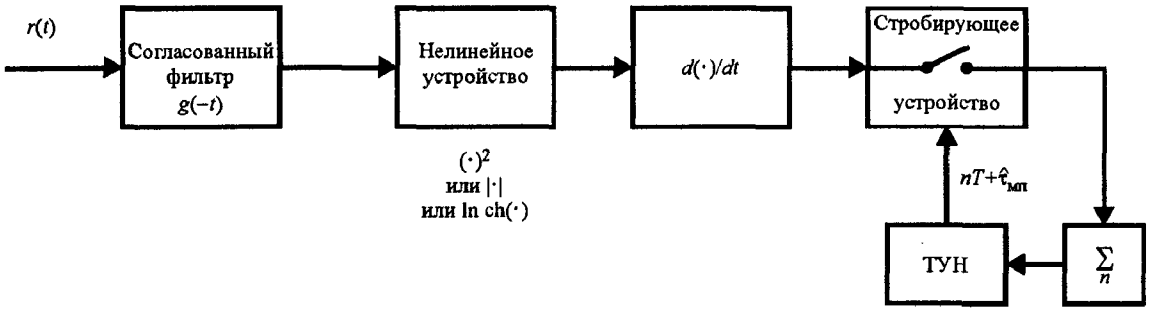


Рис. 6.3.2. МП оценивание времени задержки для базового сигнала АМ, не управляемое решениями

Альтернативно реализация отслеживающей петли, основанная на (6.3.9), иллюстрируется на рис. 6.3.3. В обоих структурах мы видим, что суммирование служит петлевым фильтром, который управляет ТУН. Интересно отметить сходство таймерной петли на рис. 6.3.3 и петли Костаса для оценивания фазы.

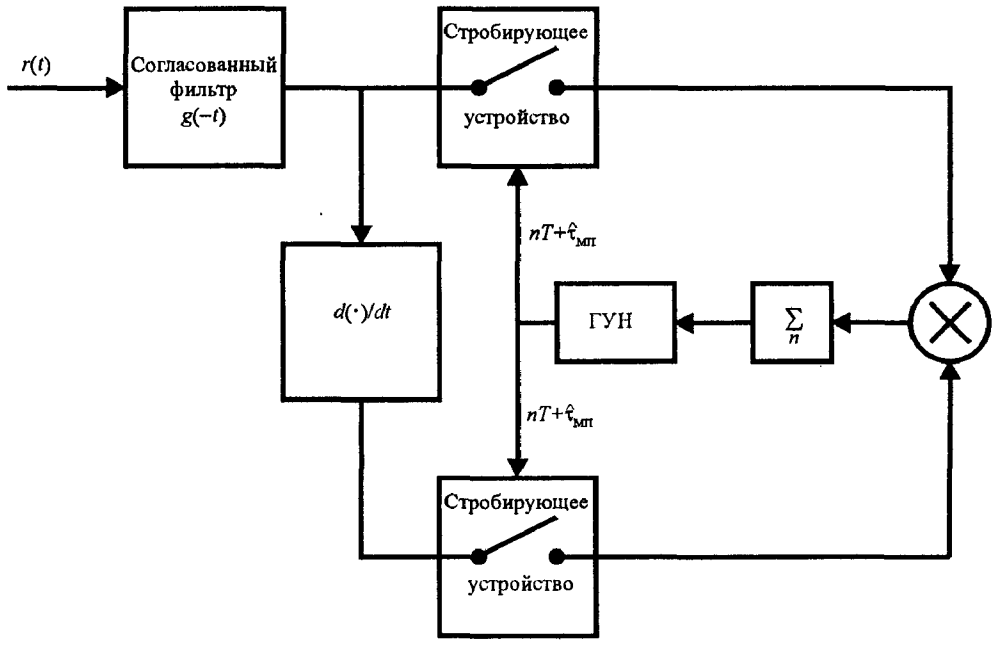


Рис. 6.3.3. Оценивание временного сдвига без обратной связи по решению для АМ в базовой полосе частот

Синхронизаторы с окнами на задержку-опережение. Другой оцениватель времени задержки, не управляемый решениями, использует симметричные свойства сигнала на

выходе согласованного фильтра или коррелятора. Чтобы описать этот метод, рассмотрим прямоугольный импульс $s(t)$, $0 \le t \le T$, показанный на рис. 6.3.4 (a). Выход фильтра, согласованного с $s(t)$, получает свое максимальное значение в точке $t = T$, как показано на рис. 6.3.4 (b). Таким образом, выход согласованного фильтра является временной функцией корреляции импульса $s(t)$. Конечно, это положение справедливо для произвольной огибающей импульса, так что подход, который мы опишем, применим в общем к произвольному сигнальному импульсу. Ясно, что хорошая точка для взятия отсчёта на выходе согласованного фильтра для получения максимального выхода – это $t = T$, т.е. точка на пике корреляционной функции.

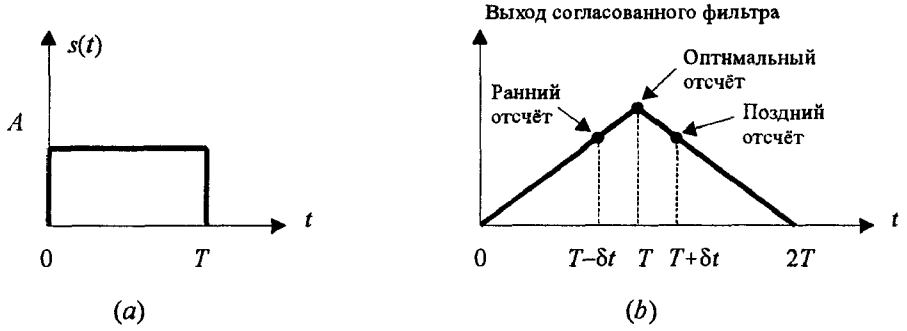


Рис. 6.3.4. Прямоугольный импульс сигнала (*a*) и выход согласованного с ним фильтра (*b*)

В присутствии шума идентификация пикового значения сигнала в общем случае затруднена. Допустим, что вместо стробирования сигнала в точке пика мы берём отсчёт раньше (в точке $t = T - \delta$) и позже (в точке $t = T + \delta$). Абсолютные значение ранних отсчётов $|y(m(T-\delta))|$ и поздних отсчетов $|y(m(T+\delta))|$ будут меньше (в среднем в присутствии шума), чем абсолютное значение в пике $|y(mT)|$. Поскольку автокорреляционная функция чётна относительно оптимального времени взятия отсчётов $t = T$, абсолютные значения корреляционной функции в точке $t = T - \delta$ и $t = T + \delta$ равны. С учетом этого условия хорошая точка отсчёта – средняя точка между $t = T - \delta$ и $t = T + \delta$. Это условие образует основу синхронизатора с окнами на задержку-опережение.

Рисунок 6.3.5 иллюстрирует блок-схему синхронизатора с окнами на задержку-опережение. На этом рисунке корреляторы используются вместо эквивалентных согласованных фильтров. Два коррелятора интегрируют по символьному интервалу T, но один коррелятор начинает интегрирование на δ секунд раньше относительно оцениваемого оптимального времени отсчёта, а второй интегратор начинает интегрирование на δ секунд позже относительно оцениваемого оптимального времени отсчета. Сигнал ошибки формируется путем взятия разности между абсолютными значениями выходов двух корреляторов. Чтобы сгладить влияние шума на отсчёты сигналов, сигнал ошибки пропускается через фильтр нижних частот. Если время отсчёта отличается от оптимального времени отсчёта, усредненный сигнал ошибки на выходе фильтра нижних частот не равен нулю, и таймерная последовательность смещается в сторону отставания или опережения, в зависимости от знака ошибки. Таким образом, сглаженный сигнал ошибки используется для управления ТУН, чей выход является желательным таймерным сигналом, который используется для стробирования. Выход ТУН также используется как таймерный сигнал для генератора символьного сигнала, который выдает ту же базовую форму импульса, что на выходе фильтра передатчика. Эта форма импульса смещается во времени на δ в сторону опережения и отставания, и полученные

образцы ожидаемого сигнала поступают на два коррелятора, как показано на рис. 6.3.5. Заметим, что, если сигнальные импульсы прямоугольные, нет надобности в генераторе сигнального импульса внутри отслеживающей петли.

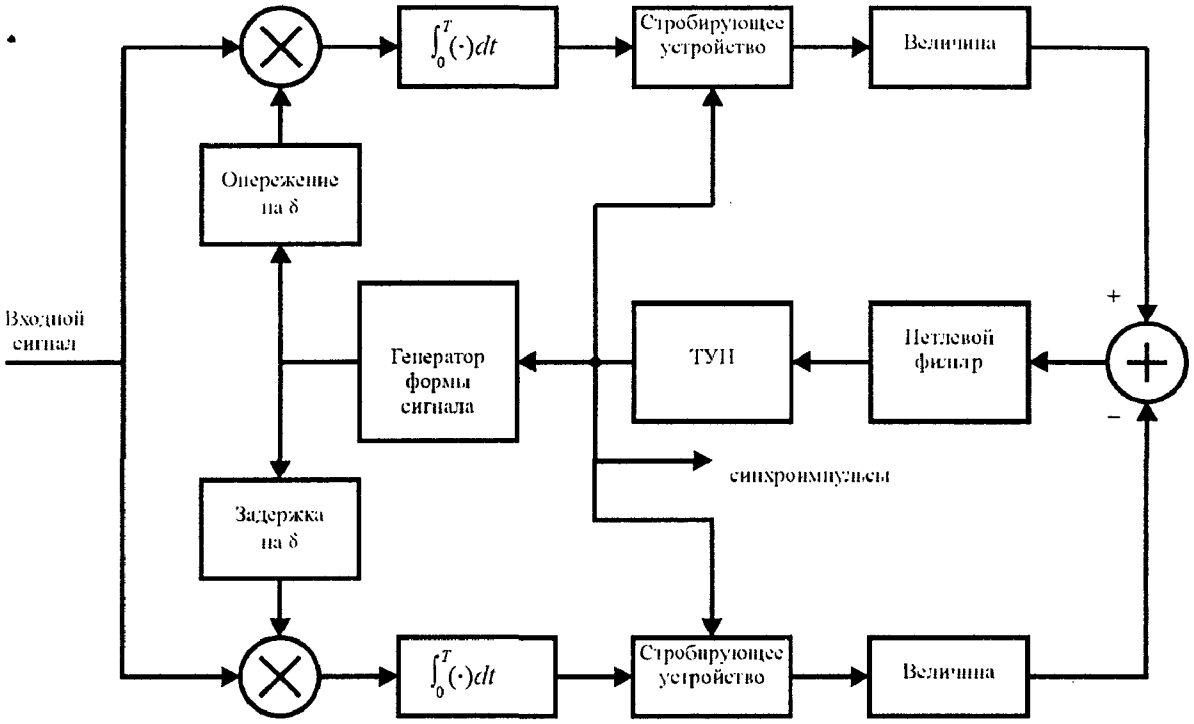


Рис. 6.3.5. Блок-схема синхронизатора с окнами на задержку–опережение

Мы видели, что синхронизатор с окнами на задержку-опережение имеет в своей основе систему замкнутого петлевого управления, чья полоса относительно узка по сравнению со скоростью передачи символов $1/T$. Полоса петли определяет качество оценки времени задержки. Узкополосная петля обеспечивает большее усреднение по аддитивному шуму и, таким образом, улучшает качество оцениваемых отсчётных величин в предположении, что время распространения в канале неизменно и таймерный генератор на передаче не дрейфует со временем (или дрейфует очень медленно во времени). С другой стороны, если время распространения в канале меняется со временем и (или) таймер передатчика также дрейфует со временем, тогда полосу петли следует увеличить, чтобы обеспечить отслеживание быстрых изменений во времени параметров синхронизации.

В устройствах отслеживания два коррелятора эффективно взаимодействуют при соседних символах. Однако, если последовательность информационных символов имеет нулевое среднее, как в случае с АМ и при других видах модуляции, вклад в выходы корреляторов от соседних импульсов усредняется до нуля в фильтре нижних частот.

Эквивалентная реализация для синхронизатора с окнами на задержку-опережение, которая несколько проще в реализации, дана на рис. 6.3.6. В этом случае таймерный сигнал от ТУН опережает и запаздывает на δ, и эти таймерные сигналы используются для стробирования выходов двух корреляторов.

Синхронизатор с окнами на задержку-опережение, описанный выше, является оценивателем задержки сигнала, не управляемым решениями, который аппроксимирует максимально правдоподобный оцениватель. Это утверждение можно продемонстрировать путём аппроксимации производной от логарифма функции правдоподобия конечной разностью, т.е.

$$\frac{d\overline{\Lambda}_L(\tau)}{d\tau} = \frac{\overline{\Lambda}_L(\tau+\delta) - \overline{\Lambda}_L(\tau-\delta)}{2\delta}. \qquad (6.3.10)$$

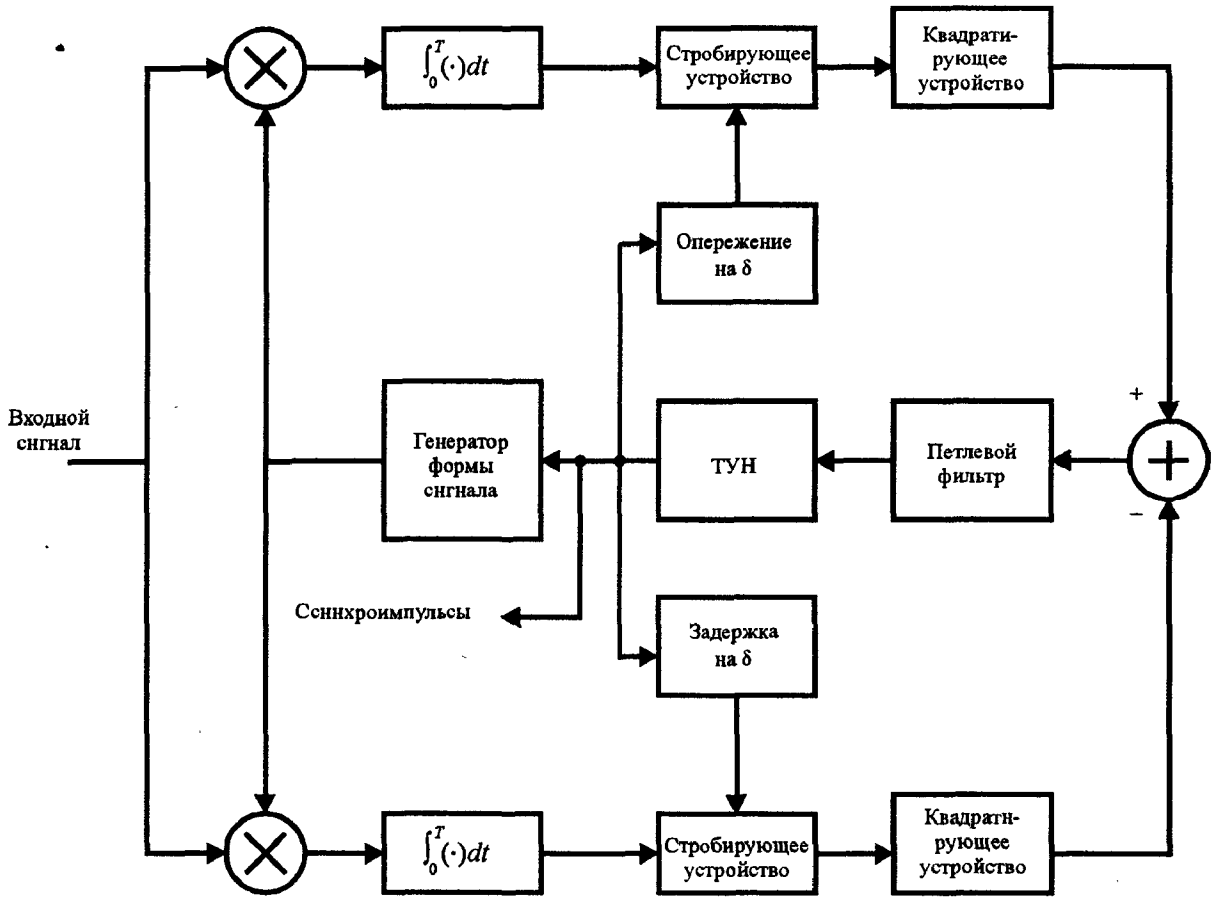


Рис. 6.3.6. Блок-схема синхроннзатора с окнами на задержку–опережение – альтернативный вариант

Если подставим выражение до $\overline{\Lambda}_L(\tau)$ из (6.3.8) в (6.3.10), получим следующую аппроксимацию для производной:

$$\frac{d\overline{\Lambda}_L(\tau)}{d\tau} = \frac{C^2}{4\delta}\sum_n\left[y_n^2(\tau+\delta) - y_n^2(\tau-\delta)\right] \approx \frac{C^2}{4\delta}\sum_n\left\{\left[\int_{T_0} r(t)g(t-nT-\tau-\delta)\,dt\right]^2 - \right.$$
$$\left. -\left[\int_{T_0} r(t)g(t-nT-\tau+\delta)\,dt\right]^2\right\}. \qquad (6.3.11)$$

Но математические выражения (6.3.11) принципиально описывают преобразования, выполняемые синхронизатором с окнами на задержку-опережение, иллюстрируемые на рис. 6.3.5 и 6.3.6.

6.4. СОВМЕСТНОЕ ОЦЕНИВАНИЕ ФАЗЫ НЕСУЩЕЙ И ФАЗЫ СИНХРОНИЗАЦИИ

Оценку фазы несущей и фазы синхронизации можно получать отдельно, как описано выше, или совместно. Совместные МП оценки двух или более сигнальных параметров бывают не хуже, а обычно лучше, чем оценки, полученные раздельной оптимизацией функций правдоподобия. Другими словами, дисперсии сигнальных параметров,

полученных при совместной оптимизации, меньше или равны дисперсиям оценок параметров, полученных при раздельной оптимизации функций правдоподобия.

Рассмотрим совместное оценивание фазы несущей и параметра задержки. Логарифм функции правдоподобия для этих двух параметров можно выразить через эквивалентные низкочастотные сигналы так:

$$\overline{\Lambda}_L(\phi,\tau) = \mathrm{Re}\left[\frac{1}{N_0}\int_{T_0} r(t)\, s_l^*(t;\phi,\tau)\,dt\right], \tag{6.4.1}$$

где $s_l(t;\phi,\tau)$ – эквивалентный низкочастотный сигнал, который имеет общую форму:

$$s_l^*(t;\phi,\tau) = \mathrm{e}^{-j\phi}\left[\sum_n I_n g(t-nT-\tau) + j\sum_n I_n g(t-nT-\tau)\right], \tag{6.4.2}$$

где $\{I_n\}$ и $\{J_n\}$ – две информационные последовательности.

Заметим, что для АМ можем считать $J_n = 0$ для всех n, а последовательность $\{I_n\}$ вещественная.

Для КАМ и ФМ положим $J_n = 0$ для всех n, а последовательность $\{I_n\}$ комплексная. Для офсетной КФМ обе последовательности $\{I_n\}$ и $\{J_n\}$ – ненулевые, и $w(t) = g(t-\frac{1}{2}T)$.

Для МП оценивания ϕ и τ, управляемого решениями, логарифм отношения правдоподобия равен

$$\overline{\Lambda}_L(\phi,\tau) = \mathrm{Re}\left\{\frac{\mathrm{e}^{j\phi}}{N_0}\sum_n\left[I_n^* y_n(\tau) + jJ_n^* x_n(\tau)\right]\right\}, \tag{6.4.3}$$

$$\begin{aligned} y_n(\tau) &= \frac{C^2}{4\delta}\int_{T_0} r(t)\, g^*(t-nT-\tau)\,dt, \\ x_n(\tau) &= \frac{C^2}{4\delta}\int_{T_0} r(t)\, w^*(t-nT-\tau)\,dt. \end{aligned} \tag{6.4.4}$$

Необходимые условия для того, чтобы оценки ϕ и τ были МП оценками, таковы:

$$\frac{\partial\Lambda_L(\phi,\tau)}{\partial\phi} = 0, \qquad \frac{\partial\Lambda_L(\phi,\tau)}{\partial\tau} = 0. \tag{6.4.5}$$

Удобно ввести определение

$$A(\tau) + jB(\tau) = \frac{1}{N_0}\sum_n\left[I_n^* y_n(\tau) + jI_n^* x_n(\tau)\right]. \tag{6.4.6}$$

С этим определением (6.4.3) можно выразить в простой форме

$$\Lambda_L(\phi,\tau) = A(\tau)\cos\phi - B(\tau)\sin\phi. \tag{6.4.7}$$

Теперь условия (6.4.5) для совместных МП оценок можно записать в виде

$$\frac{\partial\Lambda_L(\phi,\tau)}{\partial\phi} = -A(\tau)\sin\phi - B(\tau)\cos\phi = 0, \tag{6.4.8}$$

$$\frac{\partial\Lambda_L(\phi,\tau)}{\partial\tau} = \frac{\partial A(\tau)}{\partial\tau}\cos\phi - \frac{\partial B(\tau)}{\partial\tau}\sin\phi = 0. \tag{6.4.9}$$

Из (6.4.8) получаем

$$\hat{\phi}_{\mathrm{МП}} = -\mathrm{arctg}\left[\frac{B(\hat{\tau}_{\mathrm{МП}})}{A(\hat{\tau}_{\mathrm{МП}})}\right]. \tag{6.4.10}$$

Подставляя (6.4.10) в (6.4.9), находим решение для $\hat{\tau}_{\mathrm{МП}}$ в виде

$$\left[A(\tau)\frac{\partial A(\tau)}{\partial \tau}+B(\tau)\frac{\partial B(\tau)}{\partial \tau}\right]_{\tau=\tau_{\text{МП}}}=0. \qquad (6.4.11)$$

Отслеживающая петля с прямым решением для КАМ (или ФМ), полученная из этого уравнения, иллюстрируется на рис. 6.4.1

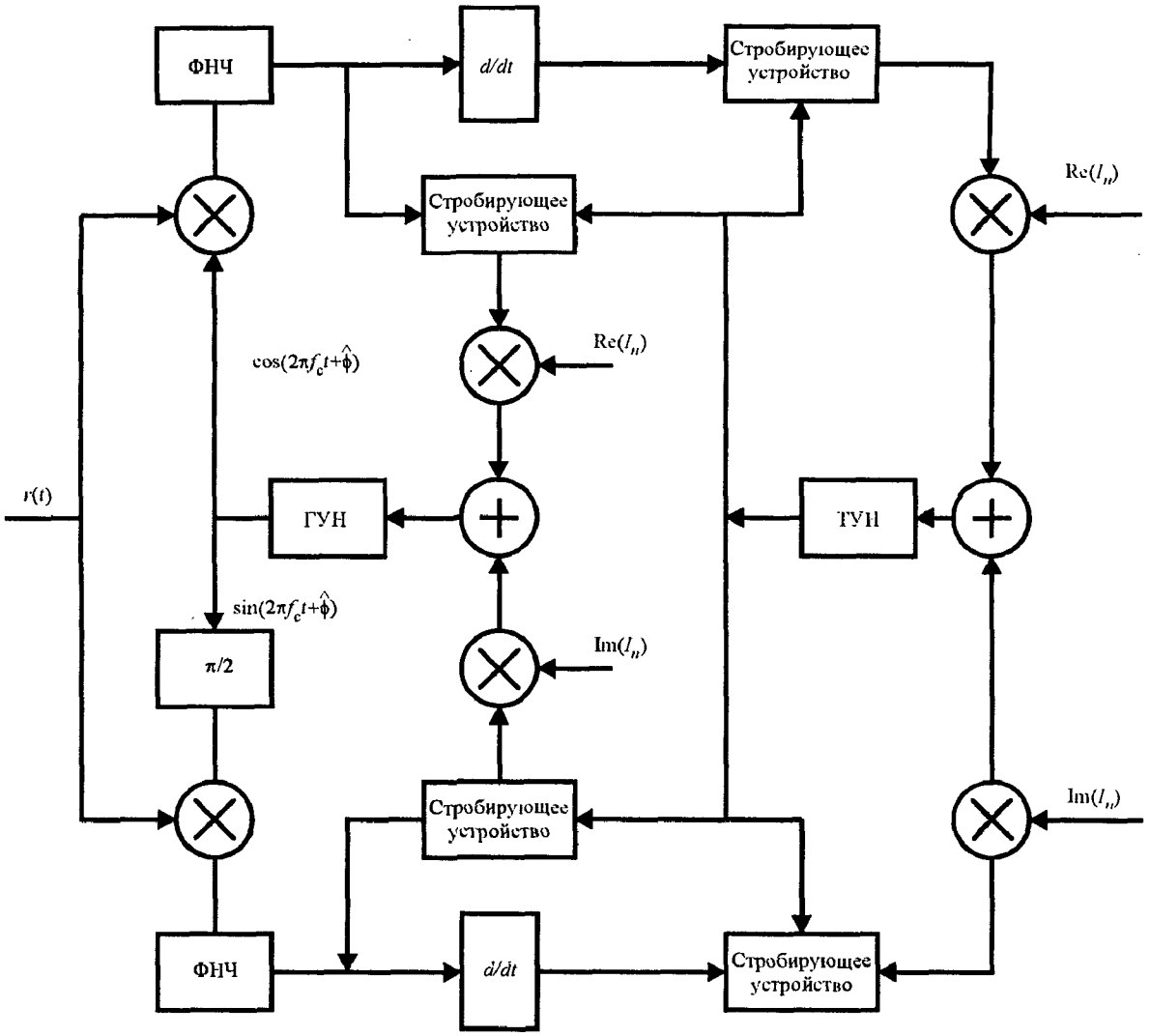


Рис. 6.4.1. Совместная отслеживающая петля, управляемая решением, для оценивания фазы несущей и параметра задержки для КАМ и ФМ

Офсетная КФМ требует немного более сложную структуру для совместного оценивания ϕ и τ. Структуру легко получить из (6.4.6) и (6.4.11).

Кроме совместного оценивания, управляемого решениями, также возможно получить совместные оценки для фазы несущей и параметра задержки без управления решениями, хотя мы не будем обсуждать этот подход.

Мы хотим также напомнить, что возможно совместить проблему оценки параметров с демодуляцией информационной последовательности $\{I_n\}$. Это значит, что можно рассматривать совместно максимально правдоподобное оценивание $\{I_n\}$, фазы несущей ϕ и параметра задержки τ.

Результаты по вопросам совместного оценивания появились в технической литературе, например у Кобаяси (1971), Фальконера (1976) и Фальконера и Сальца (1977).

6.5. ХАРАКТЕРИСТИКИ КАЧЕСТВА МП ОЦЕНИВАТЕЛЕЙ

Качество оценки параметра сигнала обычно измеряется её смещением и дисперсией. Чтобы определить эти характеристики, предположим, что мы имеем последовательность наблюдений $[x_1\ x_2\ x_3\ \ldots x_n\ ]=\mathbf{x}$ с ФПВ $p(\mathbf{x}|\phi)$, из которых извлекаем оценку параметра ϕ. Смещение оценки $\hat{\phi}(\mathbf{x})$ определяется так:

$$\text{смещение} = E[\hat{\phi}(\mathrm{x})]-\phi, \tag{6.5.1}$$

где ϕ – истинное значение параметра. Если $E[\hat{\phi}(\mathrm{x})]=\phi$, мы говорим, что оценка *несмещённая*. Дисперсия оценки $\phi(\mathbf{x})$ определяется так:

$$\sigma_\phi^2 = E\left\{\left[\hat{\phi}(\mathbf{x})\right]^2\right\}-\left\{E\left[\hat{\phi}(\mathbf{x})\right]\right\}^2. \tag{6.5.2}$$

В общем σ_ϕ^2 трудно вычислить. Однако хорошо известным результатом в теории оценивания параметра (см. Хелстром, 1968) является нижняя граница Крамера-Рао для среднеквадратической ошибки:

$$E\left\{\left[\hat{\phi}(\mathbf{x})-\phi\right]^2\right\}\geq\left[\frac{\partial}{\partial\phi}E\left[\hat{\phi}(\mathbf{x})\right]\right]^2\Big/E\left\{\left[\frac{\partial}{\partial\phi}\ln p(\mathbf{x}|\phi)\right]^2\right\}. \tag{6.5.3}$$

Заметим, что если оценка несмещённая, то числитель в (6.5.3) равен единице, и граница (6.5.3) приводит к нижней границе для дисперсии σ_ϕ^2 оценки $\hat{\phi}(\mathbf{x})$, т.е.

$$\sigma_\phi^2\geq 1/E\left\{\left[\frac{\partial}{\partial\phi}\ln p(\mathbf{x}|\phi)\right]^2\right\}. \tag{6.5.4}$$

Поскольку $\ln p(\mathbf{x}|\phi)$ отличается от логарифма отношения правдоподобия постоянным множителем, не зависящим от ϕ, то получим

$$E\left\{\left[\frac{\partial}{\partial\phi}\ln p(\mathrm{x}|\phi)\right]^2\right\}=E\left\{\left[\frac{\partial}{\partial\phi}\ln\Lambda(\phi)\right]^2\right\}=-E\left\{\frac{\partial^2}{\partial\phi^2}\ln\Lambda(\phi)\right\}. \tag{6.5.5}$$

Следовательно, нижняя граница для дисперсии равна

$$\sigma_\phi^2\geq 1/E\left\{\left[\frac{\partial}{\partial\phi}\ln\Lambda(\phi)\right]^2\right\}=-1\Big/E\left\{\frac{\partial^2}{\partial\phi^2}\ln\Lambda(\phi)\right\}. \tag{6.5.6}$$

Эта нижняя граница – очень полезный результат. Она даёт оценку близости при сравнении дисперсии практической оценки относительно нижней границы. Несмещённая оценка, дисперсия которой достигает нижней границы, называется *эффективной*.

В общем эффективные оценки являются редкими. Если они существуют, то являются оценками максимального правдоподобия. Хорошо известный результат из теории оценивания параметра – это то, что МП оценка параметра асимптотически (при произвольно большом числе наблюдений) не смещена и эффективна. В значительной степени эти желательные свойства определяют важность МП оценки параметра. Также известно, что МП оценка имеет асимптотически гауссовское распределение [со средним ϕ и дисперсией, равной нижней границе, определяемой (6.5.6)].

В случае МП оценок, описанных в этой главе для двух сигнальных параметров, дисперсии в общем обратно пропорциональны отношению сигнал-шум или, что эквивалентно, обратно пропорциональны мощности сигнала, умноженной на интервал наблюдения T_0. Далее, дисперсии оценок, управляемых решениями, при малых вероятностях ошибки в целом ниже, чем дисперсии оценок, не управляемых решениями.

Фактически качество МП оценок, управляемых решениями, для ϕ и τ достигает нижней границы.

Следующие примеры относятся к расчёту нижней границы Крамера-Рао для МП оценки фазы несущей.

Пример 6.5.1. МП оценка фазы немодулированной несущей, как было показано в (6.2.11), удовлетворяет условию

$$\int_{T_0} r(t)\sin\left(2\pi f_c t + \hat{\phi}_{\text{МП}}\right)dt = 0, \tag{6.5.7}$$

где

$$r(t) = s(t;\phi) + n(t) = A\cos\left(2\pi f_c t + \phi\right) + n(t). \tag{6.5.8}$$

Условие (6.5.7) получено при взятии производной логарифма функции правдоподобия

$$\Lambda_L(\phi) = \frac{2}{N_0}\int_{T_0} r(t)\,s(t;\phi)\,dt. \tag{6.5.9}$$

Дисперсия $\hat{\phi}_{\text{МП}}$ имеет нижнюю границу

$$\begin{aligned}\sigma^2_{\hat{\phi}_{\text{МП}}} &\geq \left\{\frac{2A}{N_0}\int_{T_0} E[r(t)]\cos\left(2\pi f_c t + \phi\right)dt\right\}^{-1} = \left\{\frac{A^2}{N_0}\int_{T_0} dt\right\}^{-1} = \\ &= \frac{N_0}{A^2 T_0} = \frac{N_0/2T_0}{\frac{1}{2}A^2} = \frac{N_0 B_{\text{экв}}}{\frac{1}{2}A^2}.\end{aligned} \tag{6.5.10}$$

Множитель $1/2T_0$ – это эквивалентная (односторонняя) шумовая полоса идеального интегратора.

Из этого примера мы видим, что дисперсия МП оценки фазы имеет нижнюю границу

$$\sigma^2_{\hat{\phi}_{\text{МП}}} \geq 1/\gamma_L, \tag{6.5.11}$$

где $\gamma_L = A^2/2N_0B_{\text{экв}}$ – петлевое ОСШ. Это также дисперсия, получаемая при оценке фазы несущей посредством ФЗП с оценкой, управляемой решениями. Как мы уже видели, оценки, не управляемые решениями, нельзя выполнить так хорошо из-за потерь в нелинейностях, требуемых для снятия модуляции, например потерь из-за квадратирования или возведения в M-ю степень.

Похожие результаты можно получить для качества оценок параметра задержки, рассмотренных выше. В дополнение к их зависимости от ОСШ качество оценки параметра задержки является функцией от огибающей сигнального импульса. Например, на практике обычно используется импульс, имеющий спектр в виде приподнятого косинуса (см. разд. 9.2).

Для такого импульса среднеквадратическая ошибка (σ_τ) оценивания параметра задержки как функция от ОСШ показана на рис. 6.5.1 для оценок, управляемых и не управляемых решениями.

Отметим значительное улучшение качества оценки, управляемой решениями, по сравнению с оценкой, не управляемой решениями. Теперь, если меняется полоса частот импульса τ, меняется огибающая импульса и, следовательно, среднеквадратическая ошибка оценки параметра задержки также меняется. Например, если меняется полоса частот импульса, который имеет спектр в форме приподнятого косинуса, среднеквадратическая ошибка меняется так, как показано на рис. 6.5.2. Заметим, что ошибка уменьшается по мере увеличения полосы частот импульса.

В заключение мы представили метод МП оценки сигнальных параметров и применили его для оценки фазы несущей и параметра задержки символов. Мы также описали их характеристики качества.

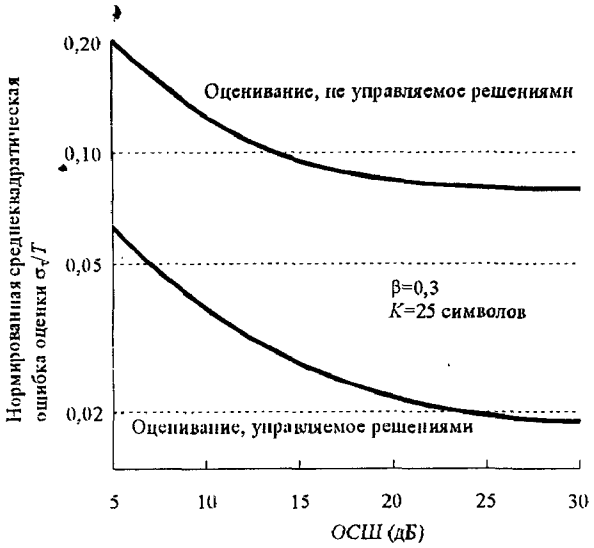


Рис. 6.5.1. Качество оценки параметра задержки для фиксированного сигнала и фиксированной петлевой полосы [L.Franks, Synchronization Subsystem: Analysis and Design, 1983]

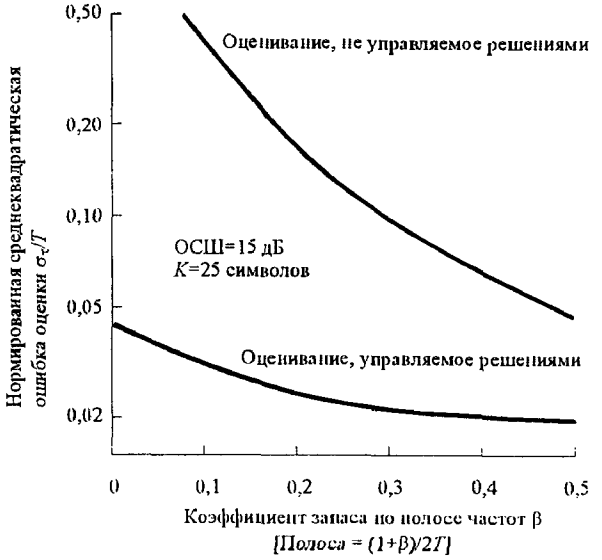


Рис. 6.5.2. Качество оценки параметра задержки для фиксированного ОСШ и фиксированной петлевой полосы [L.Franks, Synchronization Subsystem: Analysis and Design, 1983]

6.6. БИБЛИОГРАФИЧЕСКИЕ ЗАМЕЧАНИЯ И ССЫЛКИ

Восстановление несущей и фазовая синхронизация являются двумя темами, которые постоянно прорабатывались за последние три десятилетия. Петля Костаса была изобретена в 1956 г., а методы оценивания фазы, управляемого решениями, были описаны Прокисом и др. (1964) и Натали и Валбессером (1969). Работа по оценкам, управляемым решениями (ОСР), была мотивирована ранней работой Прайса (1962a,b). Всесторонний анализ фазозамкнутой петли сначала появился в книгах Витерби (1966) и Гарднера (1979). Книги, которые охватывают вопросы восстановления фазы несущей и техники синхронизации, были написаны Стиффлером (1971), Линдсеем (1972), Линдсеем и Саймоном (1973), Мейром и Ашайдом (1990)[1] .

Определённое число учебных статей появилось в журналах IEEE по ФЗП и синхронизации символов. Мы цитировали для примера статью Гупта (1975), который рассматривает как аналоговую, так и цифровую реализацию ФЗП, и статью Лидсея и Чая (1981), которая посвящена анализу цифровых ФЗП. В дополнение учебная статья Френкса (1980) описывала методы оценки фазы несущей и синхронизации символов, включая методы, основанные на оценках по правилу максимального правдоподобия. Статья Френкса содержится в специальном исследовании в IEEE *Transactions on Communications* (август 1980), посвященном синхронизации. Статья Мюллера и Маллера (1976) описывала алгоритм цифровой обработки принимаемого сигнала для выделения синхроимпульсов.

Применение правила максимального правдоподобия для оценки параметра было впервые описано в контексте оценки радиолокационного параметра (дальность и скорость перемещения объекта). Впоследствии это оптимальное правило было применено как к оцениванию фазы несущей и параметра синхронизации, так и к совместному оцениванию этих параметров и самих данных. Статьи на эти темы были опубликованы различными исследователями, включая Фальконера (1976), Менгали (1977), Фальконера и Сальца (1977), Мейерса и Френкса (1980).

[1] В этой области следует отметить работы В.В. Шахгильдяна, А.А. Ляховкина, М.С. Лохвицкого [50, 51], а также В.В. Гинзбурга и А.А. Каяцкаса [49] (прп).

ЗАДАЧИ

6.1. Докажите отношение (6.1.7).

6.2. Нарисуйте эквивалентную реализацию приемника для двоичных сигналов ФМ (рис. 6.1.1), которая содержит согласованный фильтр вместо коррелятора.

6.3. Допустим, что петлевой фильтр (см. (6.2.14)) для ФЗП имеет передаточную функцию

$$G(s)=\frac{1}{s+\sqrt{2}}.$$

a) Определите передаточную функцию $H(s)$ для замкнутой петли и укажите, является ли петля устойчивой.

b) Определите коэффициент затухания и резонансную частоту петли.

6.4. Рассмотрите ФЗП для оценки фазы несущей сигиала, для которого петлевой фильтр задан так:

$$G(s)=\frac{K}{1+\tau_1 s}.$$

a) Определите передаточную функцию $H(s)$ для замкнутой петли и ее усилеиие при $f=0$.

b) Для какого диапазона значеиий τ_1 и K петля устойчива?

6.5. Петлевой фильтр $G(s)$ в ФЗП реализован цепью, показанной на рис. Р6.5. Определите системную функцию $G(s)$ и выразите постояиные времени τ_1 и τ_2 через параметры цепи.

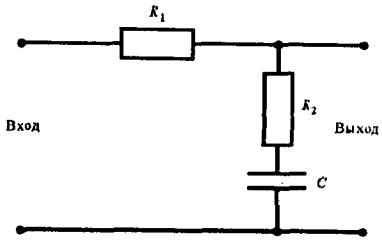


Рис. Р6.5

6.6. Петлевой фильтр $G(s)$ в ФЗП реализован активным фильтром, показанным на рис. Р.6.6. Определите системную функцию $G(s)$ и выразите постояиные временн через параметры цепи.

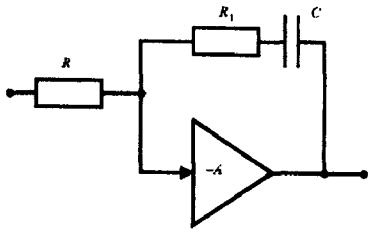


Рис. Р.6.6

6.7. Покажите, что синхронизатор с окнами на задержку-опережение, показанный на рис. 6.3.5, является близкой аппроксимацией системы синхронизации, иллюстрированной на рис. Р6.7.

6.8. Основываясь иа правиле МП, определите метод оценки фазы иесущей для двоичной модуляции с пассивной паузой.

6.9. При передаче и приеме сигналов между двигающимися объектами передаваемый сигнал получает сдвиг частоты, прямо пропорциональный скорости перемещения. Так называемый доплеровский сдвиг частоты, возникающий в этом случае, определяется формулой $f_D=\pm\frac{\upsilon}{\lambda}$,

где υ – скорость взаимного перемещения передатчика и приёмника, λ – длина волны, а знак зависит от иаправления (движение вперёд или назад) взаимного перемещения приёмннка и передатчика. Допустим, что в мобильной сотовой системе связи индивидуальный получатель перемещается со скоростью 100 км/час относительно базовой станции. Узкополосный сигнал передаётся на несущей 1 ГГц.

a) Определите доплеровское смещение частоты.

b) Какова полоса отслеживающей петли при доплеровских сдвигах, если петля рассчнтана иа отслеживание доплеровских сдвигов пользователей, двигающихся со скоростями до 100 км/час?

c) Положим, что полоса переданного сигнала 2 МГц сосредоточеиа около 1 ГГц. Определите величину доплеровского рассеяния между верхней и нижней частями сигиала.

6.10. Покажите, что среднее значение МП оцеики в (6.2.38) равно ϕ, т.е. что оцеика несмещённая.

6.11. Определите ФПВ МП оценки фазы в (6.2.38).

6.12. Определите МП оценку фазы для офсетной КФМ.

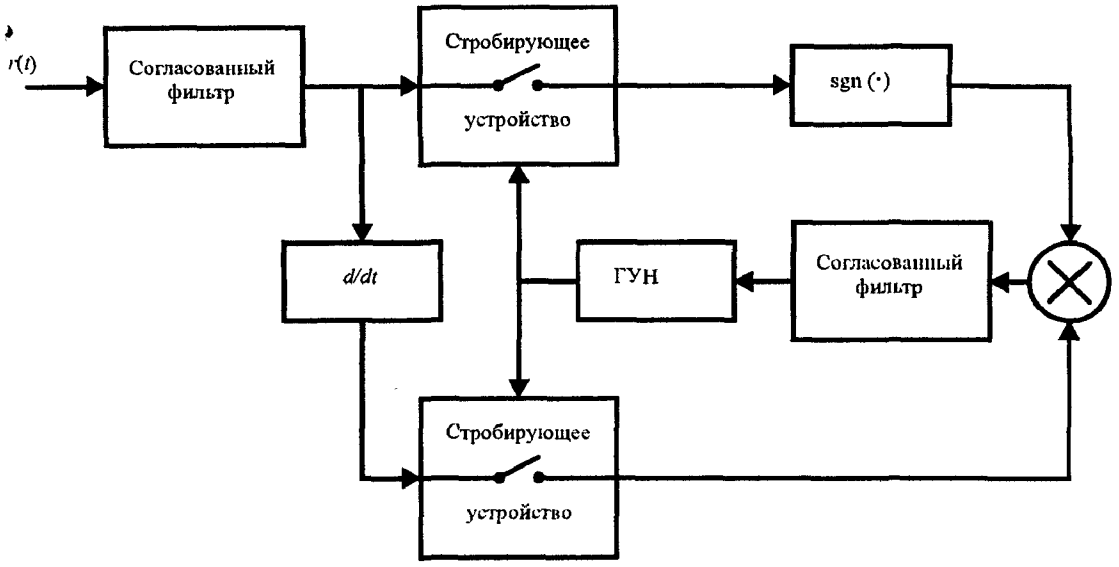


Рис. Р.6.7

6.13. Однополосный сигнал АМ можно представить так:

$$u_m(t) = A_m\left[g_T(t)\cos 2\pi f_c t - \hat{g}_T(t)\sin 2\pi f_c t\right],$$

где $\hat{g}_T(t)$ – преобразование Гильберта от $g_T(t)$, а A_m – уровень амплитуды, который содержит информацию. Покажите математически, что петлю Костаса можно использовать для демодуляции ОБП сигнала АМ.

6.14. Информация передаётся квадратурной компонентой в системе связи посредством двоичной ФМ. Таким образом, принимаемый сигнал имеет вид

$$r(t) = \pm\sqrt{2P_s}\cos(2\pi f_c + \phi) + \sqrt{2P_c}\sin(2\pi f_c + \phi) + n(t),$$

где ϕ – фаза несущей, а $n(t)$ – АБГШ. Немодулированная компонента несущей используется как некий сигнал на прнёме для оценивания фазы несущей.

a) Нарисуйте блок-схему приёмнпка, включая оцениватель фазы иесущей;

b) Покажите математически операции, определяющне оценивание фазы несущей ϕ:

c) Выразите вероятность ошибки детектнрования двоичного сигнала ФМ как функцию от суммарной переданной мощности $P_T = P_s + P_c$. Какова потеря качества, обусловленная концентрацией частоты переданной мощности в петле сигнала? Рассчитайте потерю при $P_c / P_T = 0{,}1$.

6.15. Определите сигнальную и шумовую компоненты на входе ФЗП (с возведеиием в четвёртую степень $(M = 4)$), используемой для генерирования фазы несущей при демодуляцни КФМ. Игнорируя шумовыс компоненты, кроме тех, которые линейно связаны с шумом $n(t)$, определите дисперсию оценки фазы выхода ФЗП.

6.16. Вероятность ошибки демодуляции и детектирования двоичной ФМ, если имеется ошибка в определенин фазы несущей на ϕ_c, равна

$$P_2(\phi_c) = Q\left(\sqrt{\frac{2\mathscr{E}_b}{N_0}\cos^2\phi_c}\right).$$

Предположите, что фазовая ошибка от ФЗП моделнруется гауссовской случайной величиной с нулевым средним и дисперсией $\sigma_\phi^2 << \pi^2$. Определите выражение для средней вероятности ошибки (в интегральном виде).

7

ПРОПУСКНАЯ СПОСОБНОСТЬ КАНАЛА И КОДИРОВАНИЕ

В главе 5 мы обсуждали проблему цифровой модуляции $M = 2^k$ сигналов, когда каждый сигнал содержит k бит информации. Мы видели, что некоторые методы модуляции обеспечивают лучшее качество, чем другие. В частности, мы показали, что ортогональный ансамбль сигналов позволяет нам сделать вероятность ошибки произвольно малой, взяв число сигналов $M \to \infty$ при обеспечении ОСШ на бит $\gamma_b \geq -1.6$ дБ. Таким образом, мы можем оперировать с пропускной способностью канала с аддитивным белым гауссовским шумом в пределе, когда показатель расширения канала $B_c = W / R \to \infty$. Приходится платить высокую цену, поскольку B_c растет экспоненциально с длиной блока k. Такое неэкономное использование полосы канала крайне нежелательно.

В этой и последующих главах мы рассмотрим сигналы, создаваемые посредством двоичных или недвоичных последовательностей. Результирующие сигналы, обычно, характеризуются показателем расширения полосы, который растет только линейно с k. Следовательно, кодированные сигналы предлагают в потенциале бóльшую частотную эффективность, чем ортогональные M-позиционные сигналы. Мы увидим, что, в общем, кодированные сигналы предлагают преимущество качества не только в системах с ограничением мощности, когда $R/W < 1$, но также в системах с ограничением полосы, когда $R/W > 1$.

Мы начнём с установления нескольких моделей каналов, которые будут использованы для расчета выгоды канального кодирования, и мы хотим ввести концепцию пропускной способности канала для различных моделей канала. Затем мы будем обсуждать вопросы выбора (синтеза) кодов для эффективной связи.

7.1. МОДЕЛИ КАНАЛОВ И ПРОПУСКНАЯ СПОСОБНОСТЬ КАНАЛОВ

Напомним, что в модели цифровой системы связи, описанной в разделе 1.1, передатчик содержит кодер канала, имеющий дискретный вход и дискретный выход, за которым следует модулятор. Функция дискретного кодера канала состоит во введении в двоичную информационную последовательность некоторой контролируемой избыточности, которую можно использовать в приёмнике для преодоления влияния шума и интерференции, возникающих при передаче сигнала по каналу. В общем процесс кодирования включает в себя взятие k информационных бит на определенном временном интервале и отображение каждой k-битовой информационной последовательности в уникальную (взаимно однозначную) n-битовую последовательность, называемую *кодовым словом*. Величина избыточности, вводимая кодированием данных таким путем, измеряется отношением n/k. Величину, обратную этому отношению, а именно k/n, называют *скоростью кода*.

Двоичная последовательность выхода кодера канала питает модулятор, который служит интерфейсом к каналу связи. Как мы уже обсудили, модулятор может просто отобразить каждый двоичный символ в один из двух возможных сигналов, т.е. «0»

отображается сигналом $s_1(t)$, а «1» отображается сигналом $s_2(t)$. Альтернативно, модулятор может передавать q-битовые блоки за определенное время, используя $M = 2^q$ возможных сигналов.

На приёмной стороне цифровой системы связи демодулятор обрабатывает сигналы, искажённые каналом, и преобразует каждый принятый сигнал в скаляр или вектор, который представляет оценку переданных символов данных (двоичных или M-позиционных). Детектор, который следует за демодулятором, должен решить, передан «0» или «1». В этом случае детектор выносит *жёсткое решение*. Если мы посмотрим на процесс вынесения детектором решения как на форму квантования, мы заметим, что жёсткое решение соответствует двоичному квантованию выхода демодулятора. В более общем виде мы можем рассмотреть детектор, который квантует выход на $Q > 2$ уровней, т.е. Q-ичный детектор. Если используются M-позиционные сигналы, тогда $Q \geq M$. В экстремальном случае, когда вообще не производится квантования выхода демодулятора, $Q = \infty$. В случае, когда $Q > M$, мы говорим, что детектор выносит *мягкое* решение.[1]

Квантованный выход детектора затем подаётся на канальный декодер, который использует имеющуюся в его распоряжении избыточность для коррекции искажений в канале.

В следующих разделах мы опишем три модели канала, которые будут использоваться для установления максимально возможной битовой скорости для канала.

7.1.1. Модели канала

В этом разделе мы опишем модели канала, которые будут полезны при синтезе кодов. Наиболее простая – это модель *двоичного симметричного канала* (ДСК), которая соответствует случаю, когда $M = 2$, и жёсткому решению детектора.

Двоичный симметричный канал. Рассмотрим канал с аддитивным шумом, и пусть модулятор и демодулятор/детектор включены, как части канала.

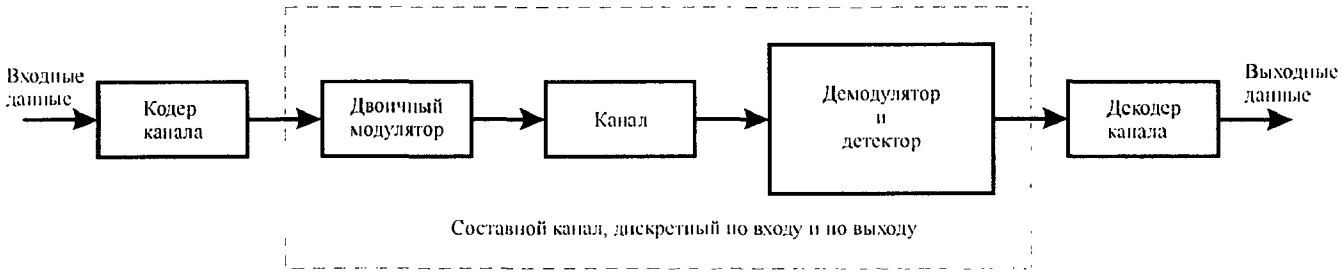


Рис. 7.1.1. Составной канал, дискретный по входу и по выходу, образованный путём включения в него модулятора и демодулятора/детектора как частей канала

Если модулятор применяет двоичные сигналы, и детектор делает жёсткие решения, то составной канал, показанный на рис. 7.1.1, имеет на входе и выходе двоичную последовательность с дискретным временем. Такой составной канал характеризуется набором $X = \{0, 1\}$ возможных входов, набором $Y = \{0, 1\}$ возможных выходов и набором условных вероятностей возможных выходов при условии возможных входов. Если канальный шум и другие нарушения вызывают статистически независимые ошибки при передаче двоичной последовательности со средней вероятностью p, тогда

[1] Чаще всего, говоря о декодировании декодером мягких решений детектора, имеют в виду, что $Q=\infty$

$$P(Y=0|X=1)=P(Y=1|X=0)=p,$$
$$P(Y=1|X=1)=P(Y=0|X=0)=1-p. \qquad (7.1.1)$$

Таким образом, мы свели каскадное соединение двоичного модулятора, канала и двоичного демодулятора и детектора в эквивалентный канал с дискретным временем, который представлен графом на рис. 7.1.2. Этот симметричный канал с двоичным входом и двоичным выходом обычно называют двоичным симметричным каналом (ДСК). Поскольку каждый выходной двоичный символ канала зависит только от соответствующего входного двоичного символа, мы говорим, что этот канал без памяти.

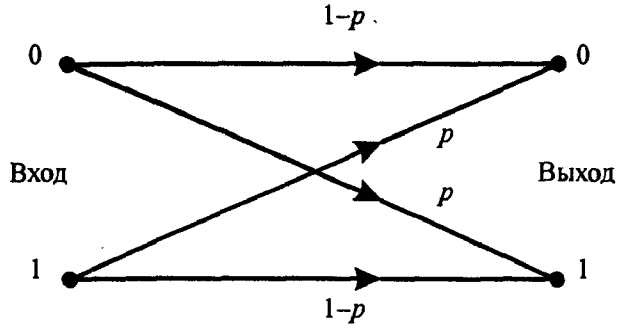


Рис. 7.1.2. Двоичный симметричный канал

Дискретные каналы без памяти. ДСК является частным случаем более общего канала с дискретным входом и дискретным выходом. Предположим, что входом кодера канала являются q-ичные символы, т.е. $X=\{x_0,x_1,\ldots,x_{q-1}\}$, а выходом детектора являются Q-ичные символы, где $Q\geq M=2^q$. Если канал и модуляция без памяти, тогда характеристика вход–выход составного канала, показанного на рис. 7.1.1, описывается рядом из qQ условных вероятностей

$$P(Y=y_i|X=x_j)\equiv P(y_i|x_j), \qquad (7.1.2)$$

где $i=0,1,\ldots,Q-1$ и $j=0,1,\ldots,q-1$. Такой канал называется дискретным каналом без памяти (ДКБП) и его графическое представление показано на рис. 7.1.3. Таким образом, если входом ДКБП является последовательность из n символов $u_1,u_2,\ldots,u_n$, выбираемых из алфавита X, и соответствующим выходом является последовательность $\upsilon_1, \upsilon_2, \ldots, \upsilon_n$ символов из алфавита Y, то совместные условные вероятности определяются так:

$$P(Y_1=\upsilon_1,Y_2=\upsilon_2,\ldots,Y_n=\upsilon_n|X_1=u_1,X_2=u_2,\ldots,X_n=u_n)=\prod_{k=1}^{n}P(Y=\upsilon_k|X=u_k) \qquad (7.1.3)$$

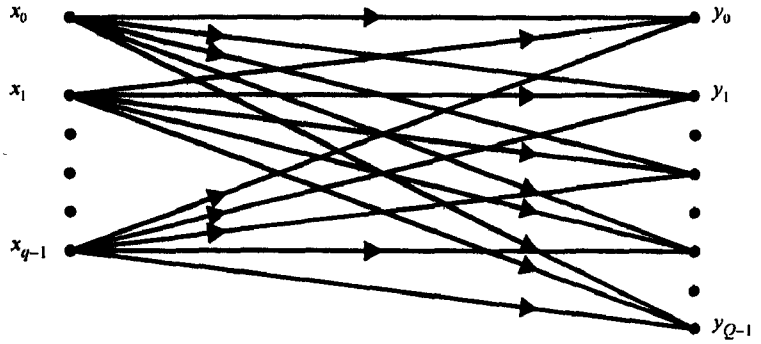


Рис. 7.1.3. Дискретный канал, q-ичный по входу и Q-ичный по выходу

Это выражение – просто математическая констатация условия отсутствия памяти.

В общем, условные вероятности $\{P(y_i|x_j)\}$, которые характеризуют ДКБП, могут быть упорядочены в форме матрицы $\mathbf{P} = [p_{ji}]$, где, по определению, $p_{ji} \equiv P(y_i|x_j)$. $\mathbf{P}$ называется матрицей переходных вероятностей канала.

Канал с дискретным входом и непрерывным выходом. Теперь предположим, что на вход модулятора подаются символы, выбираемые из конечного и дискретного входного алфавита $X = \{x_0, x_1, \ldots, x_{q-1}\}$, а выход детектора не квантован $(Q = \infty)$. Тогда входом декодера канала можно считать любую величину на вещественной оси, т.е. $Y = \{-\infty, \infty\}$. Это ведет нас к определению составного канала без памяти с дискретным временем, который характеризуется дискретным входом X, непрерывным выходом Y и рядом условных ФПВ $p(y|X = x_k), \quad k = 0, 1, \ldots, q-1$.

Наиболее важный канал этого типа – это канал с аддитивным белым гауссовским шумом (АБГШ), для которого

$$Y = X + G, \tag{7.1.4}$$

где G – гауссовские случайные величины с нулевым средним и дисперсией σ^2, а $X = x_k, \quad k = 0, 1, \ldots, q-1$. Для данного $X = x_k$ следует, что Y является гауссовской случайной величиной со средним x_k и дисперсией σ^2. Это значит

$$p(y|X = x_k) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(y-x_k)^2}{2\sigma^2}}. \tag{7.1.5}$$

Для любой входной последовательности X_i, $i = 0,1,\ldots,n$ имеется соответствующая выходная последовательность

$$Y_i = X_i + G_i, \quad i = 1, 2, \ldots, n. \tag{7.1.6}$$

Условие, что канал без памяти, можно выразить так:

$$p(y_1, y_2, \ldots, y_n | X_1 = u_1, X_2 = u_2, \ldots, X_n = u_n) = \prod_{i=1}^{n} p(y_i | X_i = u_i). \tag{7.1.7}$$

Сигнальные каналы. Мы можем отделить модулятор и демодулятор от физического канала и рассмотреть модель канала, в котором входы и выходы являются сигналами. Предположим, что такой канал имеет заданную полосу частот W с идеальной частотной характеристикой $C(f) = 1$ внутри полосы W, а сигнал на его выходе искажен аддитивным белым гауссовским шумом. Предположим, что $x(t)$ является частотно-ограниченным входом для этого канала, а $y(t)$ – соответствующий выход. Тогда

$$y(t) = x(t) + n(t), \tag{7.1.8}$$

где $n(t)$ представляет реализацию аддитивного шумового случайного процесса. Подходящий метод для определения ряда вероятностей, которые характеризуют канал, – это разложить $x(t)$, $y(t)$ и $n(t)$ в полный ряд ортонормированных функций. Это значит, мы выражаем $x(t)$, $y(t)$ и $n(t)$ в форме

$$y(t) = \sum_i y_i f_i(t); \qquad x(t) = \sum_i x_i f_i(t); \qquad n(t) = \sum_i n_i f_i(t), \tag{7.1.9}$$

где $\{x_i\}$, $\{y_i\}$ и $\{n_i\}$ – ряд коэффициентов в соответствующих выражениях, например

$$y_i = \int_0^T y(t) f_i^*(t)dt = \int_0^T [x(t) + n(t)] f_i^*(t)dt = x_i + n_i. \qquad (7.1.10)$$

Функции $\{f_i(t)\}$ образуют полный ортонормированный ансамбль на интервале $(0,T)$, т.е.

$$\int_0^T f_i(t) f_i^*(t)dt = \delta_{ij} = \begin{cases} 1 & (i=j), \\ 0 & (i \neq j), \end{cases} \qquad (7.1.11)$$

где δ_{ij} – дельта-функция Кронекера. Поскольку гауссовский шум белый, то в выражениях (7.1.9) можно использовать любой полный ансамбль ортонормированных функций.

Теперь используем коэффициенты в указанных выражениях для характеристики канала. Поскольку $y_i = x_i + n_i$, где n_i – гауссовские случайные величины, то следует

$$p(y_i|x_i) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{\frac{-(y-x_k)^2}{2\sigma_i^2}}, \quad i = 1, 2, \ldots \qquad (7.1.12)$$

Поскольку функции $\{f_i(t)\}$ в разложениях являются ортонормированными, то следует, что $\{n_i\}$ некоррелированы. Поскольку они гауссовские, они также статистически независимы. Следовательно,

$$p(y_1, y_2, \ldots, y_N | x_1, x_2, \ldots, x_N) = \prod_{i=1}^{N} p(y_i|x_i) \qquad (7.1.13)$$

для любого N. Таким образом, описание сигналов канала сведено к эквивалентному каналу с дискретным временем, характеризуемым совместной ФПВ, даваемой (7.1.12).

Когда аддитивный шум белый и гауссовский со спектральной плотностью $\frac{1}{2}N_0$, дисперсия $\sigma_i^2 = \frac{1}{2}N_0$ для всех i в (7.1.12). В этом случае отсчёты $x(t)$ и $y(t)$ можно брать со скоростью Найквиста $2W$ отсчетов/с так, что $x_i = x(i/2W)$ и $y_i = y(i/2W)$. Поскольку шум белый, отсчёты шума статистически независимы. Таким образом, (7.1.12) и (7.1.13) описывают статистику отсчётов сигналов. Заметим, что на временном интервале длительностью T имеются $N = 2WT$ отсчетов. Этот параметр используется ниже для получения пропускной способности частотно-ограниченного канала с АБГШ.

Выбор модели канала для его использования на определенном временном интервале зависит от объекта исследования. Если мы интересуемся синтезом и анализом качества кодера и декодера дискретного канала, приемлемо рассмотреть модели канала, в которых модулятор и демодулятор являются частью составного канала. С другой стороны, если наша цель – синтез и анализ качества цифрового модулятора и цифрового демодулятора, мы используем модель сигнального канала.

7.1.2. Пропускная способность канала

Теперь рассмотрим ДКБП с входным алфавитом $X = \{x_0, x_1, \ldots, x_{q-1}\}$ и выходным алфавитом $Y = \{y_0, y_1, \ldots, y_{Q-1}\}$ и рядом переходных вероятностей $P(y_i|x_j)$, определённых в (7.1.2). Предположим, что передан символ x_j, а принят символ y_i. Взаимная информация о событии $X = x_j$, когда имеет место событие $Y = y_i$, равно $\log[P(y_i|x_j)/P(y_i)]$, где

$$P(y_i) \equiv P(Y = y_i) = \sum_{k=0}^{q-1} P(x_k) P(y_i|x_k). \qquad (7.1.14)$$

Следовательно, средняя взаимная информация, получаемая по выходу Y о входе X, равна

$$I(X;Y)=\sum_{j=0}^{q-1}\sum_{i=0}^{Q-1}P(x_j)P(y_i|x_j)\log\left[P(y_i|x_j)/P(y_i)\right]. \qquad (7.1.15)$$

Характеристики канала определяют переходные вероятности $P(y_i|x_j)$, но вероятности входных символов определяются дискретным кодером канала. Величина $I(X;Y)$, максимизируемая по набору вероятностей входных символов $P(x_j)$ является величиной, которая зависит только от характеристик ДКБП через условные вероятности $P(y_i|x_j)$. Эта величина названа *пропускной способностью канала* и обозначается C. Таким образом, пропускная способность ДКБП определяется так

$$C=\max_{P(x_j)} I(X;Y)=\max_{P(x_j)}\sum_{j=0}^{q-1}\sum_{i=0}^{Q-1}P(x_j)P(y_i|x_j)\log\left[P(y_i|x_j)/P(y_i)\right]. \qquad (7.1.16)$$

Максимизация $I(X;Y)$ выполняется при условиях

$$P(x_j)\ge 0;\quad \sum_{j=0}^{q-1}P(x_j)=1.$$

Размерность C – бит/символ, если берется логарифм с основанием 2, и нат/символ, если берётся логарифм с основанием **e**. Если символы поступают в канал каждые τ_s секунд, то пропускная способность канала в единицу времени в бит/с и нат/с равна $C/\tau_s = C'$.

Пример 7.1.1

Для ДСК с переходными вероятностями

$$P(0/1)=P(1/0)=p$$

средняя взаимная информация максимизируется, если входные вероятности $P(0)=P(1)=\frac{1}{2}$. Следовательно, пропускная способность ДСК равна

$$C=p\log 2p+(1-p)\log 2(1-p)=1-H(p), \qquad (7.1.17)$$

где $H(p)$ – двоичная энтропийная функции. Кривая для C в зависимости от p иллюстрируется на рис. 7.1.4. Заметим, что при $p=0$ пропускная способность равна 1 бит/символ. С другой стороны, при $p=\frac{1}{2}$ взаимная информация между выходом и входом равна 0. Следовательно, пропускная способность равна 0. При $\frac{1}{2}<p\le 1$ мы можем поменять местами на входе ДСК 0 и 1, так что C оказывается симметричной функцией относительно точки $p=\frac{1}{2}$.

В нашей трактовке двоичной модуляции и демодуляции, данной в главе 5, мы показали, что p является монотонной функцией от отношения сигнал–шум (ОСШ), как показано на рис. 7.1.5(a). Следовательно, когда C строится как функция ОСШ, она возрастает монотонно по мере увеличения ОСШ. Зависимость C от ОСШ иллюстрируется на рис. 7.1.5(b).

Далее рассмотрим канал без памяти с АБГШ и дискретным временем, описываемый переходными ФПВ, определяемыми (7.1.5). Средняя максимальная взаимная информация между дискретным входом $X=\{x_0,x_1,\ldots,x_{q-1}\}$ и выходом $Y=\{-\infty,\infty\}$ определяется пропускной способностью канала в бит/символ и равна

$$C = \max_{P(x_i)} \sum_{i=0}^{q-1} \int_{-\infty}^{\infty} p(y|x_i) P(x_i) \log_2 \frac{p(y|x_i)}{p(y)} dy, \qquad (7.1.18)$$

где

$$p(y) = \sum_{k=0}^{q-1} p(y|x_k) P(x_k). \qquad (7.1.19)$$

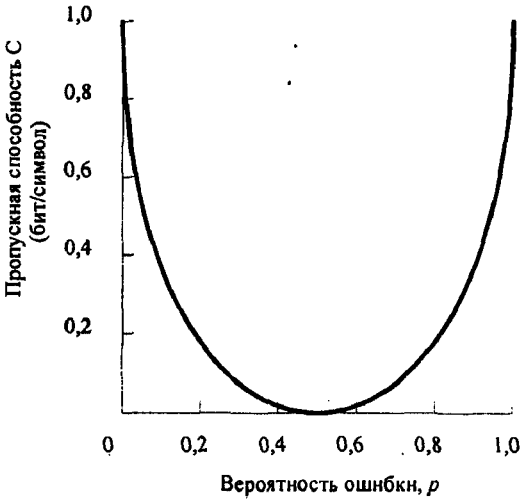


Рис. 7.1.4. Пропускиая способность ДСК как функция вероятности ошибки p

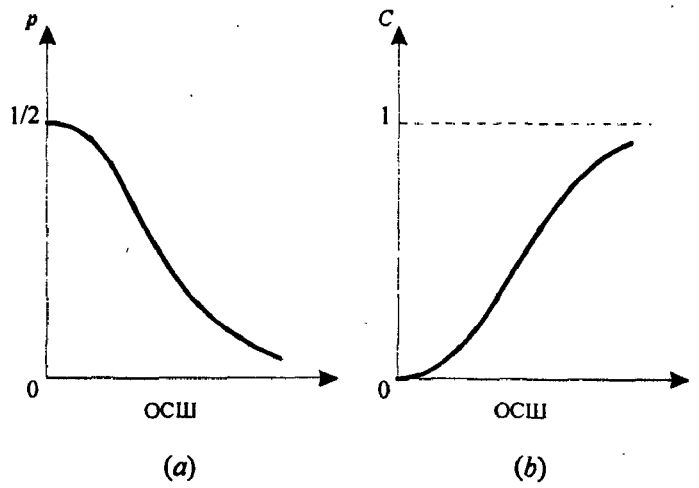


Рис. 7.1.5. Общее поведение вероятиости ошибки и пропускной способности канала, как функции от отношения сигнал/шум (ОСШ)

Пример 7.1.2

Рассмотрим канал без памяти с АБГШ и с возможными входами $X = A$ и $X = -A$. Средняя взаимная информация $I(X;Y)$ максимизируется, если вероятности входов $P(X = A) = P(X = -A) = \frac{1}{2}$. Следовательно, пропускная способность такого канала в бит/символ равна

$$C = \tfrac{1}{2} \int_{-\infty}^{\infty} p(y|A) \log_2 \frac{p(y|A)}{p(y)} dy + \tfrac{1}{2} \int_{-\infty}^{\infty} p(y|-A) \log_2 \frac{p(y|-A)}{p(y)} dy. \qquad (7.1.20)$$

Рис. 7.1.6 иллюстрирует C как функцию отношения $A^2/2\sigma^2$.

Интересно отметить, что в двух моделях канала, описанных выше, выбор одинаковой вероятности для входных символов максимизирует среднюю взаимную информацию. Таким образом, пропускная способность канала получается, когда входные символы равновероятны. Однако, такое решение для пропускной способности канала, даваемое формулами (7.1.16) и (7.1.17), не всегда имеет место.

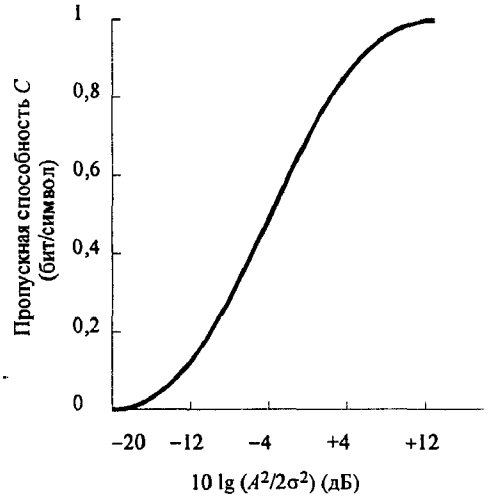


Рис. 7.1.6. Пропускная способность как функция ОСШ ($A^2/2\sigma^2$) для канала без памяти с АБГШ и двоичным входом

Ничего нельзя сказать в общем относительно задания вероятностей входа, которые максимизируют среднюю взаимную информацию. Однако, в двух моделях канала, рассмотренных выше, переходные вероятности канала проявляют форму симметрии, которая влияет на максимум $I(X;Y)$, который получается, когда символы входа равновероятны. Условия симметрии можно выразить через элементы матрицы переходных вероятностей P канала. Когда каждая строка этой матрицы является перестановкой других строк и каждый столбец является перестановкой других столбцов, матрица переходных вероятностей симметрична, а входные символы с равной вероятностью максимизируют $I(X;Y)$.

В общем необходимые и достаточные условия для совокупности вероятностей входных символов $\{P(x_j)\}$, при которых максимизируется $I(X;Y)$ и, следовательно, достигается пропускная способность ДСК, таковы (задача 7.1):

$$\begin{aligned} I(x_j;Y) &\le C \qquad \text{для всех } j \text{ с } P(x_j) > 0, \\ I(x_j;Y) &= C \qquad \text{для всех } j \text{ с } P(x_j) = 0, \end{aligned} \tag{7.1.21}$$

где C – пропускная способность канала, и

$$I(x_j;Y) = \sum_{i=0}^{Q-1} P(y_i|x_j) \log \frac{P(y_i|x_j)}{P(y_i)}. \tag{7.1.22}$$

Обычно относительно просто проверить, удовлетворяет ли совокупность входных символов с равными вероятностями условиям (7.1.21). Если они не удовлетворяются, то ряд входных символов с неравными вероятностями $\{P(x_j)\}$ могут удовлетворять (7.1.21).

Теперь рассмотрим ограниченный по полосе частот канал с аддитивным белым гауссовским шумом. Формально, пропускная способность такого канала в единицу времени определена Шенноном (1948) так:

$$C = \lim_{T\to\infty} \max_{p(x)} \frac{1}{T} I(X;Y), \tag{7.1.23}$$

где усреднённая взаимная информация определена (3.2.17). Альтернативно, мы можем использовать отсчёты или коэффициенты $\{y_i\}$, $\{x_i\}$ и $\{n_i\}$ в ряде разложений $x(t)$, $y(t)$ и $n(t)$ соответственно для определения средней взаимной информации между $\mathbf{x}_N = [x_1 x_2 \ldots x_N]$ и $\mathbf{y}_N = [y_1 y_2 \ldots y_N]$, где $N = 2WT$, $y_i = x_i + y_i$ и $p(y_i|x_j)$ определены (7.1.12). Средняя взаимная информация между x_N и y_N для канала с АБГШ равна

$$\begin{aligned} I(\mathrm{X_N};\mathrm{Y_N}) &= \int_{x_N}\ldots\iint_{y_N}\ldots\int p(\mathrm{y}_N|\mathrm{x}_N)\log\frac{p(\mathrm{y}_N|\mathrm{x}_N)}{p(\mathrm{y}_N)}d\mathrm{x}_N d\mathrm{y}_N = \\ &= \sum_{i=1}^{N}\int_{-\infty-\infty}^{\infty\ \infty} p(y_i|x_i)p(x_i)\log\frac{p(y_i|x_i)}{p(y_i)}dy_i dx_i, \end{aligned} \tag{7.1.24}$$

где

$$p(y_i|x_i) = \frac{1}{\sqrt{\pi N_0}} e^{-(y_i - x_i)^2/N_0}. \tag{7.1.25}$$

Максимум $I(X;Y)$ по входной ФПВ получается, когда входы $\{x_i\}$ статистически независимые гауссовские случайные величины с нулевыми средними, т.е.

$$p(x_i) = \frac{1}{\sqrt{2\pi\sigma_x}} e^{-x_i^2/\sigma_x^2}, \tag{7.1.26}$$

где σ_x^2 – дисперсия каждого x_i. Затем из (7.1.24) следует, что

$$\max_{p(x)} I(\mathbf{X}_N;\mathbf{Y}_N) = \sum_{i=0}^{N}\tfrac{1}{2}\log\left(1+\frac{2\sigma_x^2}{N_0}\right) = \tfrac{1}{2}N\log\left(1+\frac{2\sigma_x^2}{N_0}\right) = WT\log\left(1+\frac{2\sigma_x^2}{N_0}\right). \tag{7.1.27}$$

Предположим, что мы накладываем ограничение на среднюю мощность входных сигналов $x(t)$, т.е.

$$P_{\text{av}} = \frac{1}{T}\int_0^T E|x^2(t)|dt = \frac{1}{T}\sum_{i=1}^{N} E(x_i^2) = \frac{N\sigma_x^2}{T}. \tag{7.1.28}$$

Следовательно,

$$\sigma_x^2 = \frac{TP_{\text{ср}}}{N} = \frac{P_{\text{ср}}}{2W}. \tag{7.1.29}$$

Подставив этот результат в (7.1.27) для σ_x^2, получаем

$$\max_{p(x)} I(\mathbf{X}_N;\mathbf{Y}_N) = WT\log\left(1+\frac{P_{\text{ср}}}{WN_0}\right). \tag{7.1.30}$$

В заключение можно получить пропускную способность канала в единицу времени путем деления результата (7.1.30) на T. Таким образом,

$$C' = W\log\left(1+\frac{P_{\text{ср}}}{WN_0}\right). \tag{7.1.31}$$

Это базовая формула для пропускной способности частотно-ограниченного канала с АБГШ с частотно-ограниченным и ограниченным по средней мощности входом. Она была впервые получена Шенноном (1946).

График пропускной способности (бит/с), нормированной к полосе W, как функция от отношения средних мощностей сигнала $P_{\text{ср}}$ и шума WN_0, дан на рис. 7.1.7.

Заметим, что пропускная способность увеличивается монотонно с увеличением ОСШ. Таким образом, при фиксированной полосе пропускная способность канала увеличивается

с увеличением переданной мощности сигнала. С другой стороны, если $P_{ср}$ фиксирована, пропускную способность можно увеличить за счёт увеличения полосы W.

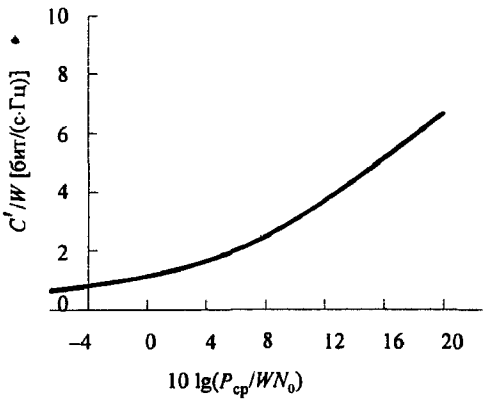


Рис. 7.1.7. Нормированная пропускная способность канала как функция ОСШ для ограниченного по полосе частот канала с АБГШ

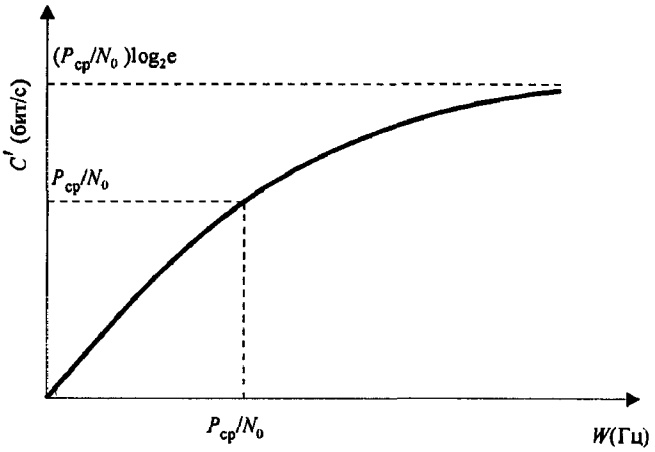


Рис. 7.1.8. Пропускная способность канала как функция полосы частот при фиксированной средней мощности сигнала

Рис. 7.1.8 дает зависимость C' от W. Заметим, что если W становится неограниченной, пропускная способность канала приближается к предельной величине

$$C = \frac{P_{ср}}{N_0}\log_2 e = \frac{P_{ср}}{N_0 \ln 2}. \qquad (7.1.32)$$

Поучительно выразить нормированную пропускную способность канала C/W как функцию от ОСШ на бит. Поскольку $P_{ср}$ представляет среднюю мощность (в Ваттах), а C определено в бит/символ, то следует

$$P_{ср} = C\mathscr{E}_b, \qquad (7.1.33)$$

где $\mathscr{E}_b$ – энергия сигнала на бит. Следовательно, (7.1.31) можно выразить так:

$$\frac{C'}{W} = \log_2\left(1 + \frac{C'}{W}\frac{\mathscr{E}_b}{N_0}\right). \qquad (7.1.34)$$

Следовательно,

$$\frac{\mathscr{E}_b}{N_0} = \frac{2^{C/W} - 1}{C/W}. \qquad (7.1.35)$$

Когда $C'/W = 1$, $\mathscr{E}_b / N_0 = 1$ (0 дБ). При $C'/W \to \infty$

$$\frac{\mathscr{E}_b}{N_0} \approx \frac{2^{C'/W}}{C'/W} \approx \exp\left(\frac{C'}{W}\ln 2 - \ln\frac{C'}{W}\right). \qquad (7.1.36)$$

Таким образом, $\mathscr{E}_b / N_0$ растёт экспоненциально, когда $C'/W \to \infty$. С другой стороны, когда $C'/W \to 0$,

$$\frac{\mathscr{E}_b}{N_0} = \lim_{C'/W \to 0}\frac{2^{C'/W} - 1}{C'/W} = \ln 2, \qquad (7.1.37)$$

которое равно –1,6 дБ. Зависимость C'/W от $\mathscr{E}_b / N_0$ показана на рис. 5.2.17.

Итак, мы получили выражение для пропускной способности для трех важных моделей канала, которые рассматриваются в этой книге. Первая – это модель канала с дискретными входом и выходом, для которой ДСК частный случай. Вторая, с дискретным входом и непрерывным выходом, – это модель канала без памяти с АБГШ. При помощи этих двух

моделей канала мы можем судить о качестве кода при получении жёстких и мягких решений (детектора) в цифровых системах связи.

Третья модель канала сфокусирована на нахождении пропускной способности в бит/с непрерывного (по входу и выходу) канала. В этом случае мы предположили ограничение полосы частот канала, что сигнал искажается в канале аддитивным белым гауссовским шумом и что средняя мощность передатчика ограничена. При этих условиях мы получили результат, даваемый (7.1.31).

Главное значение формул для пропускной способности канала, данных выше, это то, что они служат верхней границей скорости передачи для реализуемой связи по каналу с шумом. Фундаментальная роль, которую играет пропускная способность канала, определена теоремами кодирования в канале с шумами, данными Шенноном (1948 а).

Теоремы кодирования в канале с шумами

Существуют кодеры канала (и декодеры),которые делают возможным достичь надежную связь со столь малой, насколько желательно, вероятностью ошибки, если скорость передачи $R < C$, где C – пропускная способность канала. Если $R > C$, то нет возможности обеспечить стремление вероятности ошибки к нулю каким угодно кодом.

В следующем разделе мы исследуем выгоду кодирования для моделей каналов с аддитивным шумом, описанных выше, и используем пропускную способность канала, чтобы судить о доступном качестве реального кода.

7.1.3. Пропускная способность канала, достигаемая при помощи ортогональных сигналов

В разделе 5.2 мы использовали простую объединённую границу, чтобы показать, что для ортогональных сигналов вероятность ошибки можно сделать сколь угодно малой путем увеличения числа сигналов M при условии, что $\mathscr{E}_b / N_0 > 2\ln 2$. Мы указывали, что простая объединенная граница не дает наилучшую нижнюю границу для ОСШ на бит. Проблема в том, что верхняя граница, использованная для аппроксимации $Q(x)$ не очень плотная при малых x.

Альтернативный подход сводится к использованию двух различных верхних границ для $Q(x)$ в зависимости от величины x. Начиная с (5.2.21), мы видели, что

$$1-\left[1-Q(y)\right]^{M-1} \le (M-1)\,Q(y) < Me^{-y^2/2}. \qquad (7.1.38)$$

Это как раз объединенная граница, которая плотная, когда y велико, т.е. $y > y_0$, где y_0 зависит от M. Если y мало, объединенная граница превышает единицу для больших M. Поскольку

$$1-\left[1-Q(y)\right]^{M-1} \le 1 \qquad (7.1.39)$$

для всех y, мы можем использовать эту границу для $y < y_0$, т.к. она плотнее, чем единичная граница. Тогда (5.2.21) можно оценить верхней границей так:

$$P_M < \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y_0} e^{-(y-\sqrt{2\mathscr{E}})^2/2}dy + \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y_0} e^{-y^2/2}e^{-(y-\sqrt{2\mathscr{E}})^2/2}dy. \qquad (7.1.40)$$

Величину y_0, которая минимизирует эту верхнюю границу, можно найти дифференцированием правой части (7.1.40) и приравниванием производной нулю. Это выполняется легко и решение таково:

$$e^{-y_0^2/2} = M \qquad (7.1.41)$$

или, что эквивалентно,

$$y_0 = \sqrt{2\ln M} = \sqrt{2\ln 2 \log_2 M} = \sqrt{2k\ln 2}. \qquad (7.1.42)$$

Определив y_0, вычислим простые экспоненциальные верхние границы для интегралов в (7.1.40). Для первого интеграла имеем

$$\begin{aligned} \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y_0} e^{-(y-\sqrt{2\gamma})^2/2}dy &= \frac{1}{\sqrt{\pi}}\int_{-\infty}^{-(\sqrt{2\gamma}-y_0)/\sqrt{2}} e^{-x^2}dx \\ &= Q\left(\sqrt{2\gamma}-y_0\right), \quad y_0 \le \sqrt{2\gamma}, \\ &< e^{-(\sqrt{2\gamma}-y_0)^2/2}, \quad y_0 \le \sqrt{2\gamma}. \end{aligned} \qquad (7.1.43)$$

Второй интеграл ограничен сверху так:

$$\begin{aligned} \frac{M}{\sqrt{2\pi}}\int_{y_0}^{\infty} e^{-y^2/2}e^{-(y-\sqrt{2\gamma})^2/2}dy &= \frac{M}{\sqrt{2\pi}}e^{-\gamma/2}\int_{y_0-\sqrt{\gamma/2}}^{x} e^{-x^2}dx < \\ &< \begin{cases} M e^{-\gamma/2} & (y_0 \le \sqrt{\gamma/2}) \\ M e^{-\gamma/2}e^{(y_0-\sqrt{\gamma/2})^2} & (y_0 \ge \sqrt{\gamma/2}). \end{cases} \end{aligned} \qquad (7.1.44)$$

Объединяя границы для двух интегралов и подставив $e^{-y_0^2/2}$ для M, мы получим

$$P_M < \begin{cases} e^{-(\sqrt{2\gamma}-y_0)^2/2} + e^{(y_0^2-\gamma)/2} & \left(0 \le y_0 \le \sqrt{\frac{1}{2}\gamma}\right) \\ e^{-(\sqrt{2\gamma}-y_0)^2/2} + e^{(y_0^2-\gamma)/2}e^{-(y_0-\sqrt{\gamma/2})^2} & \left(\sqrt{\frac{1}{2}\gamma} \le y_0 \le \sqrt{2\gamma}\right). \end{cases} \qquad (7.1.45)$$

В области $0 \le y_0 \le \sqrt{\frac{1}{2}\gamma}$

$$P_M < e^{(y_0^2-\gamma)/2}\left(1+e^{-(y_0-\sqrt{\gamma/2})^2}\right) < 2e^{(y_0^2-\gamma)/2}, \quad \left(0 \le y_0 \le \sqrt{\tfrac{1}{2}\gamma}\right). \qquad (7.1.46)$$

В области $\sqrt{\frac{1}{2}\gamma} \le y_0 \le \sqrt{2\gamma}$ два слагаемых в (7.1.45) идентичны. Следовательно,

$$P_M < 2e^{-(\sqrt{2\gamma}-y_0)^2/2}, \quad \left(\sqrt{\tfrac{1}{2}\gamma} \le y_0 \le \sqrt{2\gamma}\right). \qquad (7.1.47)$$

Поскольку $y_0 = \sqrt{2\ln M} = \sqrt{2k\ln 2}$ и $\gamma = k\gamma_b$, границы (7.1.46) и (7.1.47) можно выразить так:

$$P_M < \begin{cases} 2e^{-k(\gamma_b-2\ln 2)^2/2} & \left(\ln M \le \frac{1}{4}\gamma\right) \\ 2e^{-k\left(\sqrt{\gamma_b}-\sqrt{\ln 2}\right)^2} & \left(\frac{1}{4}\gamma \le \ln M \le \gamma\right). \end{cases} \qquad (7.1.48)$$

Первая верхняя граница совпадает с объединённой границей, представленной ранее, но она шире для больших значений M. Заметим, что $P_M \to 0$, когда $k \to \infty$ ($M \to \infty$) при условии, что $\gamma_b > 2\ln 2$. Но ln2 – это предельное значение ОСШ на бит, требуемое для надёжной передачи, когда скорость передачи равна пропускной способности канала с АБТШ при неограниченной полосе частот, как было показано в разделе (7.1.2). Действительно, если выражения

$$\begin{aligned} y_0 &= \sqrt{2k\ln 2} = \sqrt{2RT\ln 2}, \\ \gamma &= \frac{TP_{\text{ср}}}{N_0} = TC'_\infty \ln 2 = C_\infty \ln 2 \end{aligned} \qquad (7.1.49)$$

подставить в две верхние границы, даваемые (7.1.46) и (7.1.47), где $C = P_{\text{ср}}/(N_0 \ln 2)$ – пропускная способность канала с АБГШ при неограниченной полосе частот, получаем результат

$$P_M < \begin{cases} 2\cdot 2^{-T\left(\frac{1}{2}C_\infty - R\right)} & \left(0 \le R \le \frac{1}{4}C_\infty\right), \\ 2\cdot 2^{-T\left(\sqrt{C_\infty}-\sqrt{R}\right)^2} & \left(\frac{1}{4}C_\infty \le R \le C_\infty\right). \end{cases} \qquad (7.1.50)$$

Таким образом, мы выразили границы через C_∞ и битовую скорость по каналу R. Первая верхняя граница приемлема для скоростей ниже $\frac{1}{4}C_\infty$, в то время как вторая плотнее, чем первая, для скоростей между $\frac{1}{4}C_\infty$ и C_∞ Ясно, что вероятность ошибки можно сделать произвольно малой, взяв $T\to\infty$ ($M\to\infty$ для фиксированного R), в предположении, что $R < C_\infty = P_{\text{ср}}/(N_0 \ln 2)$. Более того, мы видели, что ансамбль ортогональных сигналов достигает границ пропускной способности канала, когда $M\to\infty$, когда скорость $R < C_\infty$.

7.1.4. Функции надёжности канала

Экспоненциальные границы для вероятности ошибки M-ичной ортогональной системы сигналов в канале с АБГШ и неограниченной полосой, даваемые (7.1.50), можно выразить так

$$P_M < 2\cdot 2^{-TE(R)} \qquad (7.1.51)$$

Показатель экспоненты

$$E(R) = \begin{cases} \frac{1}{2}C_\infty - R & \left(0 \le R \le \frac{1}{4}C_\infty\right) \\ \left(\sqrt{C_\infty} - \sqrt{R}\right) & \left(\frac{1}{4}C_\infty \le R \le C_\infty\right) \end{cases} \qquad (7.1.52)$$

в (7.1.51) назван *функцией надёжности канала* с АБГШ и неограниченной полосой. График $E(R)/C_\infty$ дан на рис. 7.1.9. Показан также показатель экспоненты для объединённой границы для P_M, даваемый (5.2.27), которую можно выразить так:

$$P_M \le \tfrac{1}{2}\cdot 2^{-T\left(\frac{1}{2}C_\infty - R\right)}, \qquad 0 \le R \le \tfrac{1}{2}C_\infty. \qquad (7.1.53)$$

Ясно, что показатель экспоненты в (7.1.53) не так плотен, как $E(R)$ из-за изменения аргумента в объединённой границе в широких пределах.

Как показал Галлагер (1965), границы, даваемые (7.1.51) и (7.1.52), являются экспоненциально плотными. Это подразумевает, что не существует другая функция надёжности, скажем $E_1(R)$, удовлетворяющая условию $E_1(R) > E(R)$ для произвольного R. Следовательно, вероятность ошибки ограничена сверху и снизу как

$$K_{\text{н}}\cdot 2^{-TE(R)} \le P_e \le K_{\text{в}}\cdot 2^{-TE(R)}, \qquad (7.1.54)$$

где константы имеют только слабую зависимость от T, т.е. они меняются медленно с изменением T.

Поскольку ортогональные сигналы обеспечивают по существу то же качество, что и оптимальные симплексные сигналы для больших M, нижняя граница в (7.1.54) приемлема для любого ансамбля сигналов. Следовательно, функция надёжности $E(R)$, определяемая (7.1.52), определяет экспоненциальные характеристики вероятности ошибки для цифровых сигналов в канале с АБГШ и с неограниченной полосой частот.



Рис. 7.1.9. Функция надёжности канала с АБГШ при неограниченной полосе частот

Хотя вероятность ошибки можно сделать как угодно малой увеличивая число ортогональных или биортогональных или симплексных сигналов при $R < C_\infty$, для относительно умеренного числа сигналов, имеется большое расхождение между реальным качеством и лучшим достижимым качеством, даваемой формулой для пропускной способности канала. Для примера, из рис. 5.2.17 мы видим, что ансамбль из M=16 ортогональных сигналов требует для достижения вероятности ошибки $P_e = 10^{-5}$ ОСШ на бит при когерентном детектировании примерно 7,5 дБ. В контрасте формула для пропускной способности канала указывает на то, что для $C/W = 0{,}5$ надёжная передача возможна с ОСШ порядка –0,8 дБ. Это представляет большую разницу в 8,3 дБ/бит и является стимулом для поиска более эффективных форм сигналов. В этой главе и главе 8 мы покажем, что кодированные сигналы могут значительно сократить расхождение.

Аналогичные расхождения в качестве существуют также в частотно – ограниченной области рис. 5.2.17, где $R/W > 1$. Однако, в этой области мы должны быть более искусны для того, чтобы использовать кодирование для улучшения качества, поскольку мы не можем расширить полосу частот, как в области с ограничением мощности сигнала. Польза от техники кодирования для эффективных по полосе частот систем связи также обсуждается в главе 8.

7.2. СЛУЧАЙНЫЙ ВЫБОР КОДОВ

Синтез систем кодированной модуляции для эффективной передачи информации можно разделить на два базовых подхода.

Один – это алгебраический подход, который первоначально касался синтеза техники кодирования и декодирования для специального класса кодов, такие как циклические блоковые коды и свёрточные коды. Второй подход является вероятностным и он касается анализа качества общего класса кодированных сигналов. Этот подход даёт границы для вероятности ошибки, которые можно достичь для связи по каналу с некоторыми специфическими характеристиками.

В этом разделе мы ознакомимся с вероятностным подходом к кодированной модуляции. Алгебраический подход, базирующийся на блоковых и сверхточных кодах, рассматривается в главе 8.

7.2.1. Случайное кодирование, основанное на использовании ансамбля из M двоичных кодовых слов

Рассмотрим ансамбль из M кодовых слов, образованных из n-мерных двоичных кодовых слов вида

$$\mathbf{C}_i = [c_{i1}\ c_{i2}\ \ldots\ c_{in}], \qquad i = 1,2,\ldots,M, \tag{7.2.1}$$

где c_{ij}=0 или 1. Каждый символ кодового слова отображается двоичным ФМ сигналом, так что сигналы, соответствующие кодовому слову $\mathbf{C}_i$ можно выразить так:

$$s_i(t) = \sum_{j=1}^{n} s_{ij} f_j(t), \qquad i = 1,2,\ldots,M, \tag{7.2.2}$$

где

$$s_{ij} = \begin{cases} \sqrt{\mathscr{E}_c}, & \text{если } c_{ij} = 1, \\ -\sqrt{\mathscr{E}_c}, & \text{если } c_{ij} = 0 \end{cases} \tag{7.2.3}$$

и $\mathscr{E}_c$ – энергия на кодовый символ. Таким образом, сигналы $s_i(t)$ эквивалентны n-мерным векторам

$$\mathbf{s}_i = [s_{i1}\ s_{i2}\ \ldots\ s_{in}], \qquad i = 1,2,\ldots,M, \tag{7.2.4}$$

которые соответствуют вершинам гиперкуба в n-мерном пространстве.

Теперь предположим, что информационная скорость на входе кодера равна R бит/с и мы кодируем блоки из k бит на определенном временном интервале T посредством одного из M сигналов. Следовательно, $k = RT$ и требуется $M = 2^k = 2^{RT}$ сигналов. Удобно определить параметр D следующим образом:

$$D = \frac{n}{T} \text{ измер./с} \tag{7.2.5}$$

Таким образом, $n = DT$ – это размерность пространства сигналов.

Гиперкуб имеет $2^n = 2^{DT}$ вершин, из которых $M = 2^{RT}$ могут быть использованы для передачи информации. Если мы навяжем условие $D > R$, то часть вершин, которые мы используем как сигнальные точки, равна

$$F = \frac{2^k}{2^n} = \frac{2^{RT}}{2^{DT}} = 2^{-(D-R)T}. \tag{7.2.6}$$

Ясно, если $D > R$, имеем $F \to 0$, когда $T \to \infty$.

Вопрос, который мы хотим ставить, следующий. Можно ли выбрать подмножество $M = 2^{RT}$ вершин из $2^n = 2^{DT}$ имеющихся в распоряжении вершин так, чтобы вероятность ошибки $P \to 0$ при $T \to \infty$ или, что эквивалентно, когда $n \to \infty$? Поскольку часть F используемых вершин приближается к нулю, когда $T \to \infty$, возможно выбрать M сигналов, имеющих минимальное расстояние, которое увеличивается, когда $T \to \infty$ и, следовательно, $P_e \to 0$.

Вместо того, чтобы пытаться найти простой ансамбль из M кодовых сигналов, для которых мы рассчитаем вероятность ошибки, рассмотрим ансамбль из $(2^n)^M$ различных путей, по которым мы можем выбрать M вершин из 2^n имеющихся в распоряжении вершин гиперкуба. С каждым из 2^{nM} выборов мы можем связать систему связи, содержащей модулятор, канал и демодулятор, которые оптимальны для выбранного

набора из M сигналов. Таким образом, имеется 2^{nM} систем связи, одна для каждого выбора M кодовых сигналов, как показано на рис. 7.2.1. Каждая система связи характеризуется своей вероятностью ошибки.

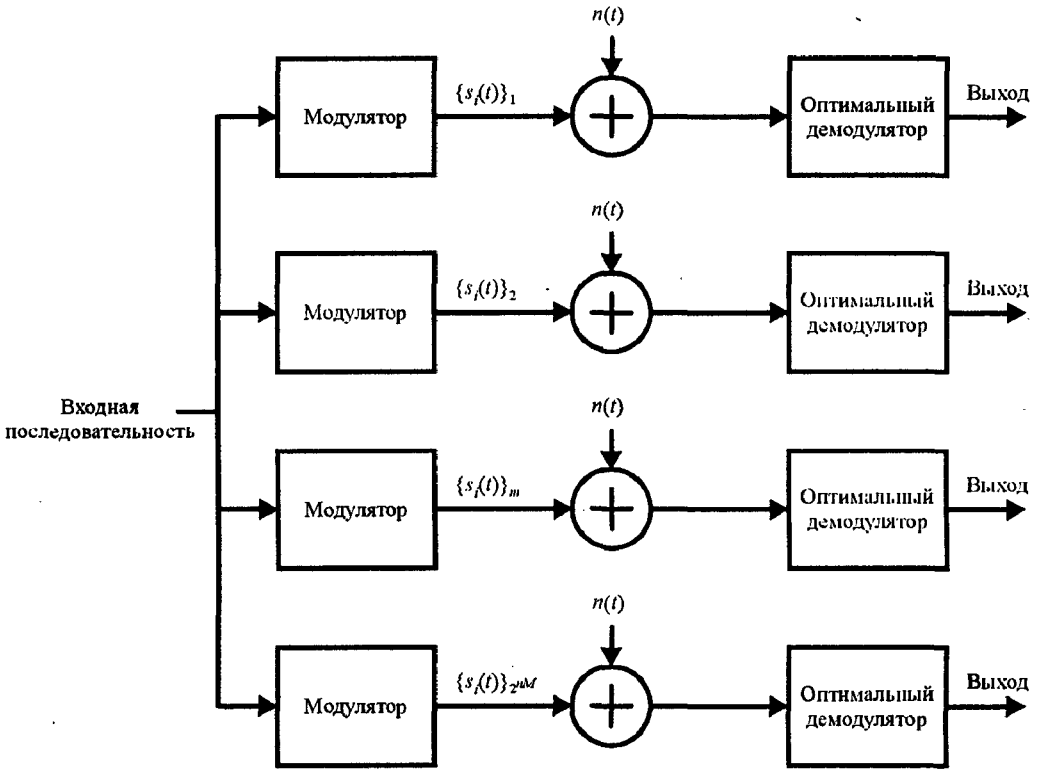


Рис. 7.2.1. Ансамбль 2^{nM} систем связи. Каждая система выбирает одну из 2^{nM} возможных последовательностей из M сигналов

Предположим, что наш выбор M кодовых сигналов основан на случайном выборе из множества 2^{nM} возможных ансамблей кодов. Таким образом, случайный выбор m-го кода, обозначенного $\{s_i\}_m$, происходит с вероятностью

$$P\left(\{s_i\}_m\right) = 2^{-nM}, \tag{7.2.7}$$

и соответствующую условную вероятность ошибки для этого выбора кодовых сигналов обозначим $P_e\left(\{s_i\}_m\right)$. Тогда средняя вероятность ошибки по всему набору кодов

$$\overline{P}_e = \sum_{m=1}^{2^{nM}} P_e\left(\{s_i\}_m\right) P\left(\{s_i\}_m\right) = 2^{-nM} \sum_{m=1}^{2^{nM}} P_e\left(\{s_i\}_m\right), \tag{7.2.8}$$

где верхняя черта над P_e означает усреднение по ансамблю кодов.

Ясно, что некоторые выборы кодов приведут к большим вероятностям ошибки. Например, код, который сопоставит все M k-битовых последовательностей одну и ту же вершину гиперкуба приведет к очень большой вероятности ошибки. В таком случае $P_e\left(\{s_i\}_m\right) > \overline{P}_e$. Однако, будут так же выборы кодов, для которых $P_e\left(\{s_i\}_m\right) < \overline{P}_e$. Следовательно, если мы получим верхнюю границу для $\overline{P}_e$, эта граница будет так же справедлива для тех кодов, для которых $P_e\left(\{s_i\}_m\right) < \overline{P}_e$. Далее, если $\overline{P}_e \to 0$ при $T \to \infty$,

тогда мы можем делать заключение, что для этих кодов $P_e\left(\{s_i\}_m\right) \to 0$, когда $T \to \infty$.

Для того, чтобы определить верхнюю границу для $\overline{P_e}$, рассмотрим передачу k-битового сообщения $\mathbf{X}_k \equiv [x_1 x_2 \ldots x_k]$, где $x_j = 0$ или 1 для $j = 1, 2, \ldots, k$. Условная вероятность ошибки, усредненная по ансамблям кодов, равна

$$\overline{P_e(X_k)} = \sum_{\text{все коды}} P_e\left(\mathbf{X}_k, \{s_i\}_m\right) P\left(\{s_i\}_m\right), \tag{7.2.9}$$

где $P_e\left(\mathbf{X}_k, \{s_i\}_m\right)$ – условная вероятность ошибки k-битового сообщения $\mathbf{X}_k$, которое передается посредством кода $\{s_i\}_m$. Для m-го кода вероятность ошибки $P_e\left(\mathbf{X}_k, \{s_i\}_m\right)$ имеет верхнюю границу

$$P_e\left(\mathbf{X}_k, \{s_i\}_m\right) \le \sum_{\substack{l=1 \\ l \ne k}}^{M} P_{2m}\left(\mathbf{s}_l, \mathbf{s}_k\right), \tag{7.2.10}$$

где $P_{2m}\left(\mathbf{s}_l, \mathbf{s}_k\right)$ – вероятность ошибки для двоичной системы связи, которая использует сигнальные векторы $\mathbf{s}_l$ и $\mathbf{s}_k$ для передачи одного из двух равновероятных k-битовых сообщений. Таким образом,

$$\overline{P_e\left(\mathbf{X}_k\right)} \le \sum_{\text{все коды}} P_e\left(\{s_i\}_m\right) \sum_{\substack{l=1 \\ l \ne k}}^{M} P_{2m}\left(\mathbf{s}_l, \mathbf{s}_k\right). \tag{7.2.11}$$

Меняя порядок суммирования в (7.2.11), получаем

$$\overline{P_e\left(\mathbf{X}_k\right)} \le \sum_{\substack{l=1 \\ l \ne k}}^{M} \left[\sum_{\text{все коды}} P_e\left(\{s_i\}_m\right) P_{2m}\left(\mathbf{s}_l, \mathbf{s}_k\right) \right] \le \sum_{\substack{l=1 \\ l \ne k}}^{M} \overline{P_2\left(\mathbf{s}_l, \mathbf{s}_k\right)}, \tag{7.2.12}$$

где $\overline{P_2\left(\mathbf{s}_l, \mathbf{s}_k\right)}$ представляет среднюю по ансамблю вероятность $P_{2m}\left(\mathbf{s}_l, \mathbf{s}_k\right)$, полученную усреднением по 2^{nM} кодам или по 2^{nM} системам связи.

Для канала с АБГШ вероятность ошибки двоичной системы $P_{2m}\left(\mathbf{s}_l, \mathbf{s}_k\right)$ равна

$$P_{2m}\left(\mathbf{s}_l, \mathbf{s}_k\right) = Q\left(\sqrt{\frac{d_{lk}^2}{2N_0}}\right), \tag{7.2.13}$$

где $d_{lk}^2 = \left|\mathbf{s}_l - \mathbf{s}_k\right|^2$. Если $\mathbf{s}_l$ и $\mathbf{s}_k$ отличаются по d координатам, то

$$d_{lk}^2 = \left\|\mathbf{s}_l - \mathbf{s}_k\right\|^2 = \sum_{j=1}^{n} \left(s_{lj} - s_{kj}\right)^2 = d\left(2\sqrt{\mathscr{E}_c}\right)^2 = 4d\mathscr{E}_c. \tag{7.2.14}$$

Следовательно,

$$P_{2m}\left(\mathbf{s}_l, \mathbf{s}_k\right) = Q\left(\sqrt{\frac{2d\mathscr{E}_c}{N_0}}\right). \tag{7.2.15}$$

Теперь мы можем усреднить $P_{2m}\left(\mathbf{s}_l, \mathbf{s}_k\right)$ по ансамблю кодов. Поскольку все коды равновероятны, сигнальный вектор $\mathbf{s}_l$ с равной вероятностью может соответствовать любой из 2^n возможных вершин гиперкуба, и это имеет место статистически независимо

от соответствия любой вершины сигнальному вектору $\mathbf{s}_k$. Следовательно, $P(s_{li} = s_{ki}) = \frac{1}{2}$ и $P(s_{li} \neq s_{ki}) = \frac{1}{2}$ независимо для всех $i = 1, 2, \ldots, n$. Как следствие, вероятность того, что $\mathbf{s}_l$ и $\mathbf{s}_k$ различаются в d позициях, равна

$$P(d) = \left(\frac{1}{2}\right)^n \binom{n}{d}. \tag{7.2.16}$$

Таким образом, математическое ожидание $P_{2m}(\mathbf{s}_l, \mathbf{s}_k)$ по ансамблю кодов можно выразить так:

$$\overline{P_2(\mathbf{s}_l, \mathbf{s}_k)} = \sum_{d=0}^{n} P(d) Q\left(\sqrt{\frac{2d\mathscr{E}_c}{N_0}}\right) = \frac{1}{2^n} \sum_{d=0}^{n} \binom{n}{d} Q\left(\sqrt{\frac{2d\mathscr{E}_c}{N_0}}\right). \tag{7.2.17}$$

Результат (7.2.17) можно упростить, если воспользоваться верхней границей для Q-функции

$$Q\left(\sqrt{\frac{2d\mathscr{E}_c}{N_0}}\right) < \mathrm{e}^{-d\mathscr{E}_c/N_0}.$$

Таким образом,

$$\overline{P_2(\mathbf{s}_l, \mathbf{s}_k)} \leq 2^{-n} \sum_{d=0}^{n} \binom{n}{d} \mathrm{e}^{-d\mathscr{E}_c/N_0} \leq 2^{-n} \left(1 + \mathrm{e}^{-\mathscr{E}_c/N_0}\right)^n \leq \left[\tfrac{1}{2}\left(1 + \mathrm{e}^{-\mathscr{E}_c/N_0}\right)\right]^n. \tag{7.2.18}$$

Мы видим, что правая часть (7.2.18) не зависит от индексов l и k. Следовательно, если мы подставим границу (7.2.18) в (7.2.12), то получим

$$\overline{P_e(\mathbf{X}_k)} \leq \sum_{\substack{l=1 \\ l \neq k}}^{M} \overline{P_2(\mathbf{s}_l, \mathbf{s}_k)} = (M-1)\left[\tfrac{1}{2}\left(1 + \mathrm{e}^{-\mathscr{E}_c/N_0}\right)\right]^n < M\left[\tfrac{1}{2}\left(1 + \mathrm{e}^{-\mathscr{E}_c/N_0}\right)\right]^n.$$

Наконец, безусловную среднюю вероятность ошибки $\overline{P}_e$ получим усреднением $\overline{P_e(\mathbf{X}_k)}$ по всем k-битовым информационным последовательностям. Таким образом,

$$\overline{P}_e = \sum_k \overline{P_e(\mathbf{X}_k)} P(\mathbf{X}_k) < M\left[\tfrac{1}{2}\left(1 + \mathrm{e}^{-\mathscr{E}_c/N_0}\right)\right]^n \sum_k P(\mathbf{X}_k) = M\left[\tfrac{1}{2}\left(1 + \mathrm{e}^{-\mathscr{E}_c/N_0}\right)\right]^n \tag{7.2.19}$$

Этот результат можно выразить в более удобной форме, прежде всего определив параметр R_0, который называется предельной скоростью и имеет размерность бит/измерение

$$R_0 = \log_2 \frac{2}{1 + \mathrm{e}^{-\mathscr{E}/N_0}} = 1 - \log_2\left(1 + \mathrm{e}^{-\mathscr{E}/N_0}\right). \tag{7.2.20}$$

Тогда (7.2.19) можно записать в виде

$$\overline{P}_e < M 2^{-nR_0} = 2^{RT} 2^{-nR_0}. \tag{7.2.21}$$

Поскольку $n = DT$, то (7.2.21) можно выразить так

$$\overline{P}_e < 2^{-T(DR_0 - R)} \tag{7.2.22}$$

Зависимость параметра R_0 как функции от $\mathscr{E}_c / N_0$ показана на рис. 7.2.2.

Видно, что $0 \leq R_0 \leq 1$. Следовательно, $\overline{P}_e \to 0$, $T \to \infty$, если информационная скорость $R < DR_0$.

Альтернативно (7.2.21) можно выразить так:

$$\overline{P_e} < 2^{-n(R_0 - R/D)}. \qquad (7.2.23)$$

Отношение R/D также имеет размерность бит/измерение и может быть определено так:

$$R_c = \frac{R}{D} = \frac{R}{n/T} = \frac{RT}{n} = \frac{k}{n}. \qquad (7.2.24)$$

Таким образом, R_c – это скорость кода, и

$$\overline{P_e} < 2^{-n(R_0 - R_c)} \qquad (7.2.25)$$

Мы замечаем, что, когда $R_c < R_0$, средняя вероятность ошибки $\overline{P_e} \to 0$, когда длина кодового блока $n \to \infty$. Поскольку среднюю величину вероятности ошибки можно сделать произвольно малой при $n \to \infty$, отсюда следует, что существуют коды в ансамбле 2^{nM} кодов, которые имеют вероятность ошибки не больше, чем $\overline{P_e}$.

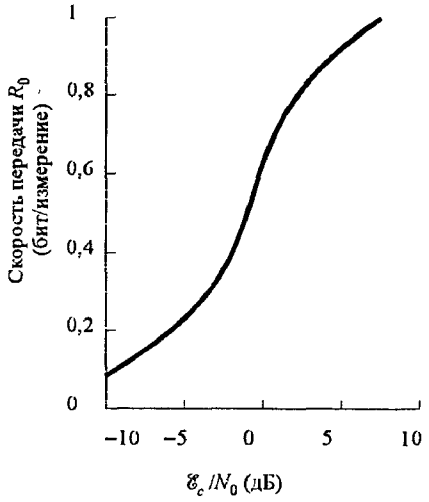


Рис. 7.2.2. Предельная скорость передачи R_0 как функция ОСШ на измерение в децибелах

Из определения средней вероятности ошибки, данного выше, мы заключаем, что хорошие коды существуют. Хотя мы нормально не можем выбрать коды случайно, интересно рассмотреть вопрос о том, может ли случайно выбранный код быть хорошим. Действительно, можно легко показать, что в ансамбле имеется много хороших кодов. Сначала заметим, что $\overline{P_e}$ – это средняя по ансамблю вероятность ошибки, полученная усреднением по всем кодам и что все отдельные вероятности. очевидно, положительные величины. Если код выбран случайно, вероятность того, что его вероятность ошибки $P_e > \alpha \overline{P_e}$, меньше, чем 1/a. Следовательно, не больше, чем 10 процентов кодов имеет вероятность ошибки, которая превосходит $10\overline{P_e}$ и не больше, чем 1 процент кодов имеют вероятность ошибки, которая превосходит $100\overline{P_e}$.

Мы хотим подчеркнуть, что коды с вероятностью ошибки, превосходящей $\overline{P_e}$, не является обязательно плохими кодами. Например, предположим, что среднюю вероятность ошибки $\overline{P_e} < 10^{-10}$ можно получить, используя коды с размерностью n_0, когда $R_0 > R_c$. Тогда, если мы выберем код с вероятностью ошибки $1000\overline{P_e} < 10^{-7}$, мы можем скомпенсировать это увеличение вероятности ошибки увеличением длины кода n от n_0 до $n = 10n_0/7$. Таким образом, скромным увеличением размерности мы имеем код с $\overline{P_e} < 10^{-10}$. Суммируя скажем, хороших кодов в изобилии и, следовательно, их можно легко найти даже случайным выбором.

Также интересно выразить среднюю вероятность ошибки в (7.2.25) через ОСШ на бит γ_b. Чтобы это сделать, выразим энергию кодового сигнала так:

$$\mathscr{E} = n\mathscr{E}_c = k\mathscr{E}_b. \qquad (7.2.26)$$

Следовательно, $n = k\mathscr{E}_b/\mathscr{E}_c$. Также заметим, что $R_c\mathscr{E}_b/\mathscr{E}_c = 1$. Таким образом, (7.2.25) можно выразить так:

$$\overline{P_e} < 2^{-k(\gamma_b/\gamma_0 - 1)}, \qquad (7.2.27)$$

где γ_0 – параметр, определяющей нормирование ОСШ, и равный

$$\gamma_0 = \frac{R_c}{R_0}\gamma_b = \frac{R_c\gamma_b}{1 - \log_2\left(1 + \mathscr{E}^{-R_c\gamma_b}\right)}. \qquad (7.2.28)$$

Теперь заметим, что $\overline{P_e} \to 0$, когда $k \to \infty$ при условии, что ОСШ на бит $\gamma_b > \gamma_0$. Зависимость γ_0 от $R_c\gamma_b$ показана на рис. 7.2.3.

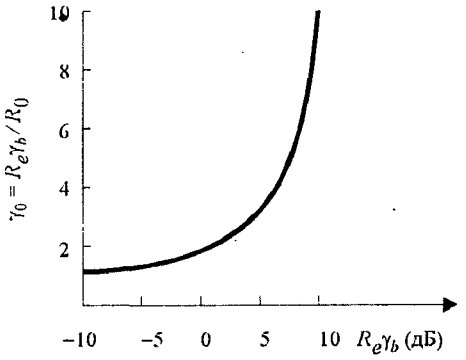


Рис. 7.2.3. Нижняя граница для ОСШ на бит для двоичных противоположных сигналов

Заметим, что когда $R_c\gamma_b \to 0$, $\gamma_0 \approx 2\ln 2$ Следовательно, вероятность ошибки для M двоичных кодовых сигналов эквивалентна вероятности ошибки, полученной из объединенной границы для M-ичных ортогональных сигналов, предполагая, что размерность кодового сигнала достаточно велика, так что $\gamma_0 \approx 2\ln 2$.

Параметр размерности D, который мы ввели в (7.2.5), пропорционален полосе канала, требуемой для передачи сигналов. Напомним, что согласно теореме отсчетов, сигнал с полосой W можно представить отсчётами, взятыми со скоростью $2W$ отсчётов/с. Таким образом, на временном интервале T имеется $n = 2WT$ отсчётов или, что эквивалентно, имеем n степеней свободы (размерность). Следовательно, D можно приравнять $2W$.

В заключение заметим, что двоичные кодовые сигналы, рассмотренные в этом разделе, используются, когда ОСШ на измерение мало, например, $\mathscr{E}_c/N_0 < 10$. Однако, когда $\mathscr{E}_c/N_0 > 10$, величина R_0 асимптотически приближается к 1 бит/измерение. Поскольку скорость кода ограничена так, что она меньше R_0, то двоичные кодовые сигналы получаются неэффективными при $\mathscr{E}_c/N_0 > 10$. В этом случае мы можем использовать недвоичные кодовые сигналы для того, чтобы достичь увеличение числа бит на измерение. Например, многоуровневый кодовый ансамбль можно синтезировать из недвоичных кодов путем отражения каждого кодового элемента в один из возможных уровней (как в АМ). Такие коды рассматриваются ниже.

7.2.2. Случайное кодирование, основанное на использовании M кодовых слов с многоуровневыми сигналами

Вместо того, чтобы синтезировать двоичные кодовые сигналы, предположим, что мы используем недвоичные коды с кодовыми словами вида (7.2.1), где кодовые элементы c_{ij} выбираются из ансамбля $\{0, 1, \ldots, q-1\}$. Каждый кодовый элемент отображается в одно из q возможных значений амплитуд. Таким образом, мы синтезируем сигналы, соответствующие n-мерным векторам $\{\mathbf{s}_i\}$, как в (7.2.4), где компоненты $\{s_{ij}\}$ выбираются из ряда амплитуд, принимающих q возможных значений. Мы имеем q^n возможных сигналов. Из них мы выбираем $M = 2^{RT}$ сигналов для передачи k-битовых информационных блоков. Величины q амплитуд, соответствующих кодовым элементам $\{0, 1, \ldots, q-1\}$, обозначим $\{a_1, a_2, \ldots a_q\}$ и будем предполагать, что они выбираются с вероятностями $\{p_i\}$. Предполагается, что уровни амплитуд равномерно распределены на интервале $[-\sqrt{\mathscr{E}_c}, \sqrt{\mathscr{E}_c}]$. Для примера, рис. 7.2.4 иллюстрирует значения амплитуд при $q = 4$.

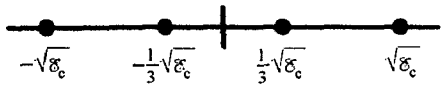


Рис. 7.2.4. Сигнальный алфавит, состоящий из 4 уровней амплитуды

В общем, соседние уровни амплитуд разделены интервалом $2\sqrt{\mathscr{E}_c}/(q-1)$. Такое задание гарантирует не только то, что каждая компонента s_{ij} ограничена пиковой энергией $\sqrt{\mathscr{E}_c}$, но также и то, что каждое кодовое слово ограничено по средней энергии, удовлетворяющей условию

$$\|s_i\|^2 < n\mathscr{E}_c. \tag{7.2.29}$$

Путем повторения подхода, данного выше, для случайного выбора кодов в канале с АБГШ, находим, что средняя вероятность ошибки ограничена сверху

$$\overline{P_e} < M2^{-nR_0} = 2^{RT}2^{-nT_0} = 2^{-n(R_0-R/D)}, \tag{7.2.30}$$

где R_0 определено как

$$R_0 = -\log_2\left(\sum_{l=1}^{q}\sum_{m=1}^{q} p_l p_m e^{-d_{lm}^2/4N_0}\right) \tag{7.2.31}$$

и

$$d_{lm} = |a_l - a_m|, \qquad l,m = 1,2,\ldots,q \tag{7.2.32}$$

В частном случае, когда все уровни амплитуд равновероятны,

$$p_l = p_m = 1/q,$$

и (7.2.31) сводится к

$$R_0 = -\log_2\left(\frac{1}{q^2}\sum_{l=1}^{q}\sum_{m=1}^{q} p_l p_m e^{-d_{lm}^2/4N_0}\right). \tag{7.2.33}$$

Для примера, когда $q=2$ и $a_1 = -\sqrt{\mathscr{E}_c}$, $a_2 = \sqrt{\mathscr{E}_c}$, имеем $d_{11} = d_{22} = 0$, $d_{12} = d_{21} = 2\sqrt{\mathscr{E}_c}$ и, следовательно,

$$R_0 = \log_2 \frac{2}{1+e^{-\mathscr{E}_c/N_0}}, \qquad q = 2,$$

что совпадает с нашим прежним результатом. Когда $q=4$, $a_1 = -\sqrt{\mathscr{E}_c}$, $a_2 = -\sqrt{\mathscr{E}_c}/3$, $a_3 = \sqrt{\mathscr{E}_c}/3$, $a_4 = \sqrt{\mathscr{E}_c}$, имеем $d_{mm} = 0$ для $m = 1, 2, 3, 4$, $d_{12}=d_{23}=d_{34}=d_{21}=d_{32}=d_{43}= 2\sqrt{\mathscr{E}_c}/3$, $d_{13}=d_{31}=d_{24}=d_{42}=4\sqrt{\mathscr{E}_c}/3$ и $d_{14} = d_{41} = 2\sqrt{\mathscr{E}_c}$. Следовательно,

$$R_0 = \log_2 \frac{2}{2+3e^{-\mathscr{E}_c/9N_0}+2e^{-4\mathscr{E}_c/9N_0}+e^{-\mathscr{E}_c/9N_0}}, \qquad q = 4. \tag{7.2.34}$$

Ясно, R_0 теперь насыщается до 2 бит/измерение по мере увеличения $\mathscr{E}_c/N_0$.

График R_0 как функции $\mathscr{E}_c/N_0$ при равноотстоящих и равномерно распределённых уровнях амплитуд показан на рис. 7.2.5 для $q = 2, 3, 4, 8, 16, 32$ и 64. Заметим, что таким образом, для больших ОСШ $P_e \to 0$, когда $n \to \infty$ при условии, что $R < DR_0 = 2WR_0$, бит/с.

Если мы изменим пиковую энергию элемента, но сохраним заданную среднюю энергию на кодовое слово, как дано в (7.2.29), возможно получить большую верхнюю границу на число бит/измерение. Для этого случая, результат, полученный Шенноном (1959b), равен

$$R_0^* = \frac{1}{2}\left[1+\frac{\mathscr{E}_c}{N_0}-\sqrt{1+\left(\frac{\mathscr{E}_c}{N_0}\right)^2}\right]\log_2 e+\frac{1}{2}\log_2\left[\frac{1}{2}\left(1+\sqrt{1+\left(\frac{\mathscr{E}_c}{N_0}\right)^2}\right)\right]. \quad (7.2.35)$$

График R_0^* в функции от ОСШ на измерение ($\mathscr{E}_c/N_0$) также дан на рис. 7.2.5. Ясно, что наш выбор равномерно отстоящих, равновероятных уровней амплитуд, который приводит к R_0, является субоптимальным. Однако, эти кодовые сигналы легче синтезировать и реализовать на практике. Это важное преимущество, которое оправдывает их использование.

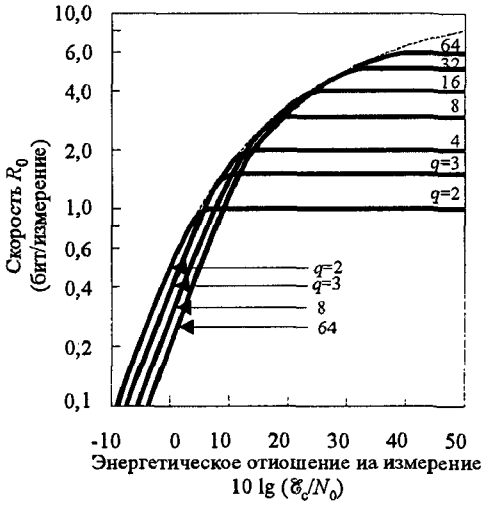


Рис. 7.2.5. Предельная скорость R_0 для равноотстоящих уровней амплитуды с равными вероятностями

7.2.3. Сравнение предельной скорости с пропускной способностью канала с АБГШ

Пропускная способность частотно-ограниченного канала с АБГШ с заданной средней мощностью входного сигнала было приведена в разделе 7.1.2 и определяется так:

$$C' = W\log_2\left(1+\frac{P_{\text{ср}}}{WN_0}\right) \text{ бит/с}, \quad (7.2.36)$$

где $P_{\text{ср}}$ – средняя мощность входного сигнала, а W – полоса канала. Интересно выразить пропускную способность этого канала в бит/измерение, а среднюю мощность в энергии/измерении. С учётом того, что $D = 2W$ и

$$\mathscr{E}_c = \frac{\mathscr{E}}{n} = \frac{P_{\text{ср}}T}{n},$$

имеем

$$P_{\text{ср}} = \frac{n}{T}\mathscr{E}_c = D\mathscr{E}_c. \quad (7.2.37)$$

Определив $C_n = C'/2W = C'/D$ и заменив W и $P_{\text{ср}}$ в (7.2.36), можно записать

$$C_n = \frac{1}{2}\log_2\left(1+2\frac{\mathscr{E}_c}{N_0}\right) = \frac{1}{2}\log_2\left(1+2R_c\gamma_b\right) \text{ бит/измерение}. \quad (7.2.38)$$

Это выражение для нормированной пропускной способности канала можно сравнить с R_0^*, как показано на рис. 7.2.6. Поскольку C_n – безусловный верхний предел скорости передачи R/D, то $R_0^* < C_n$, как и ожидалось. Мы также видим, что для малых значений $\mathscr{E}_c/N_0$ разница меду R_0^* и C_n приближенно равно 3 дБ. Следовательно, использование

случайно, выбираемых, оптимально ограниченных по усреднённой мощности многоуровневых сигналов приводит к функции скорости R_0^*, которая находится внутри 3 дБ зоны пропускной способности канала. Требуется тщательно разработанная техника определения границ, чтобы показать, что вероятность ошибки можно сделать сколь угодно малой, когда $R < DC_n = 2WC_n = C$

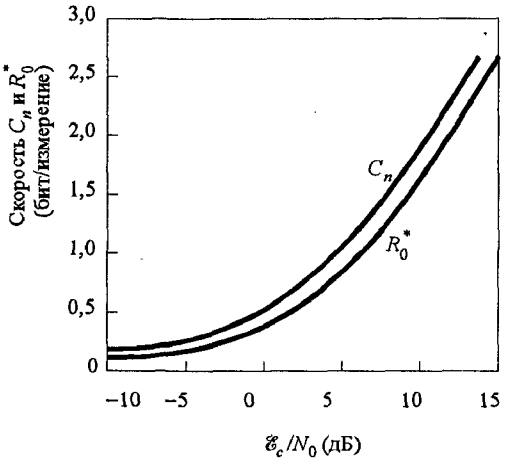


Рис. 7.2.6. Сравнение предельной скорости R_0^* с пропускной способностью для канала с АБГШ

7.3. СИНТЕЗ СИСТЕМЫ СВЯЗИ, ОСНОВАННЫЙ НА ПРЕДЕЛЬНОЙ СКОРОСТИ

В предыдущем обсуждении мы характеризовали качество кодирования и модуляции через вероятность ошибки, которая конечно является определяющей при синтезе системы связи. Однако, во многих случаях, расчет вероятности ошибки чрезвычайно сложен, особенно, если при обработке сигнала в приёмнике используются нелинейные операции, такие, например, как квантование сигнала, или когда аддитивный шум не гауссовский.

Вместо того, чтобы стремиться к расчету точных значений вероятности ошибки для специальных кодов, мы можем использовать усредненную по ансамблю среднюю вероятность ошибки при случайном выборе кодовых слов. Предполагается, что канал имеет q входных символов $\{0, 1, \ldots, q-1\}$ и Q выходных символов $\{0, 1, \ldots, Q-1\}$ и характеризуется переходными вероятностями $P(i|j)$, где $j = 0, 1, \ldots, q-1$ и $i = 0, 1, \ldots, Q-1$, причем $Q \geq q$. Входные символы появляются с вероятностями $\{p_j\}$ и считаются статистически независимыми. Дополнительно, шум в канале считается независимым во времени, так что нет зависимости между принимаемыми соседними символами. При этих условиях, среднюю вероятность ошибки по ансамблю при случайном выборе кодовых слов можно получить с использованием границы Чернова (см. Витерби и Омура, 1979).

Общий результат, который получен для дискретных каналов без памяти, таков:

$$\overline{P_e} < 2^{-n(R_Q - R/D)}, \qquad (7.3.1)$$

где n – длина кодового блока, R – информационная скорость в битах/с, D – число измерений в секунду, а R_Q – предельная скорость для квантователя с Q уровнями, определяемая так:

$$R_Q = \max_{\{p_j\}}\left\{-\log_2 \sum_{i=0}^{Q-1}\left[\sum_{j=0}^{q-1} p_j \sqrt{P(i|j)}\right]^2\right\}. \qquad (7.3.2)$$

С точки зрения синтеза кода, комбинация из модулятора, канала и демодулятора образует канал с дискретным временем с q входами и Q выходами. Переходные вероятности $\{P(i|j)\}$ зависят от характеристик канального шума, числа уровней квантования и типа квантователя, например. равномерного или неравномерного. Для примера, в канале с АБГШ и двоичным входом, условное распределение выхода коррелятора в точке отсчета можно выразить так:

$$p(y|j) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(y-m_j)^2/2\sigma^2}, \qquad j = 0, 1, \qquad (7.3.3)$$

где

$$m_0 = -\sqrt{\mathscr{E}_c}\,,\ m_1 = \sqrt{\mathscr{E}_c} \text{ и } \sigma^2 = \tfrac{1}{2}N_0.$$

Эти две ФПВ показаны на рис. 7.3.1. Рисунок также иллюстрирует схему квантования, которая разделяет вещественную ось на пять областей. При таком разделении мы можем вычислить переходные вероятности и оптимальным образом выбрать пороги, при которых деление на области происходит так, что максимизируется R_Q при заданном Q. Имеем

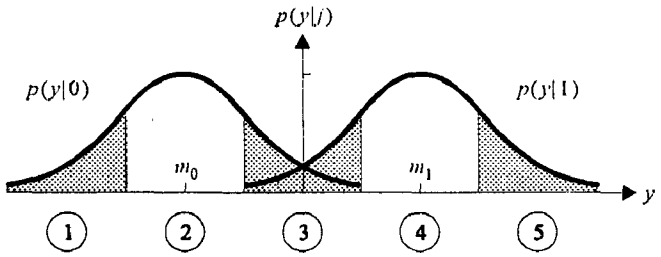


Рис. 7.3.1. Пример квантования выхода демодулятора пятью уровнями

$$p(i|j) = \int_{r_i} p(y|j)dy\,, \qquad (7.3.4)$$

где интеграл от $p(y|j)$ вычисляется по области r_i, которая соответствует переходной вероятности $P(i|j)$.

Величина скорости R_Q в пределе, когда $Q \to \infty$, определяет предельную скорость для неквантованного выхода демодулятора или согласованного фильтра. Можно непосредственно показать, что, когда $Q \to \infty$, первое суммирование (сумма от $i = 0$ до $Q-1$) в (7.3.2) заменяется интегрированием, а переходные вероятности заменяются соответствующими ФПВ. Таким образом, когда канал состоит из q дискретных входов и одного непрерывного выхода y, который представляет неквантованный выход согласованного фильтра или коррелятора в системе связи, которая использует ФМ или многоуровневую (АМ) модуляцию, предельная скорость определяется так

$$R_0 = \max_{\{p_j\}}\left\{-\log_2 \int_{-\infty}^{\infty}\left[\sum_{j=0}^{q-1} p_j \sqrt{p(y|j)}\right]^2 dy\right\}, \qquad (7.3.5)$$

где p_j, $0 \le j \le q-1$ определяет вероятность передачи j-го символа, а $p(y|j)$ – условная

плотность вероятности выхода y согласованного фильтра или взаимокоррелятора, когда передан j-й сигнал. Это и есть выражение для определения качества неквантованного декодирования (декодирования мягких решений).

Пример 7.3.1

Сравним качество двоичной АМ в канале с АБГШ, когда приёмник квантует выход на $Q = 2$, 4 и 8 уровней. Чтобы упростить проблему оптимизации при квантовании сигнала на выходе демодулятора, квантованные уровни располагаются в точках 0, $\pm\tau_h$, $\pm 2\tau_h$, $\ldots \pm\left(2^{b-1}-1\right)\tau_h$, где τ_h – параметр, определяющий шаг квантования, который должен быть выбран, а b – число бит, определяющие уровни квантования. Хорошая стратегия для выбора τ_h – выбрать его так, чтобы минимизировать ОСШ на бит γ_b, которые требуются для работы с кодовой скоростью R_0. Это значит, что параметр шага квантования должен оптимизироваться для любого ОСШ, что означает при практической реализации приёмника, что ОСШ должен измеряться. К сожалению, τ_h не проявляет высокую чувствительность к малым изменениям в ОСШ, так что возможно оптимизировать τ_h для определенного значения ОСШ и получать хорошее качество для широкой области ОСШ вблизи этого номинального значения, используя фиксированное τ_h.

Основываясь на таком подходе, выражение для R_Q, определяемое (7.3.2), было пересчитано для $b = 1$ (декодирование жестких решений), 2 и 3 бита, соответствующие числу уровней квантования $Q = 2$, 4 и 8. Результаты даны кривыми на рис. 7.3.2. Величина R_0 для неквантованного декодирования мягких решений, полученная расчетом по (7.3.5), также дана на рис. 7.3.2. Видим, что двухбитовое квантование с $\tau_h = 1$ выигрывает около 1,4 дБ относительно декодирования жестких решений, а трехбитовое квантование с $\tau_h =$ $=0{,}5$ дает дополнительное улучшение на 0,4 дБ.

При трёхбитовом квантовании мы проигрываем не более, чем на 0,2 дБ, относительно предельного неквантованного декодирования мягких решений. Ясно, что в будущем можно будет лишь немного выиграть путем улучшения точности обработки.

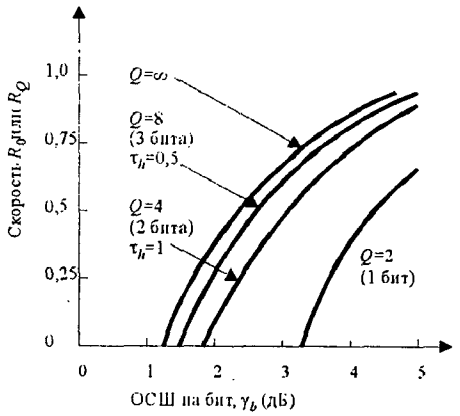


Рис. 7.3.2. Влияние квантования на помехоустойчивость кодовой системы связи при предельной скорости R_0 или R_Q при двоичной ФМ в канале с АБГШ

Если используется недвоичный код совместно с M-ичными $(M = q)$ сигналами, принимаемый сигнал на выходе M согласованных фильтров можно представить вектором $\mathbf{y} = [y_1\ y_2\ \ldots y_M]$. Предельная скорость для этого канала с M-ичным входом и M-ичным (неквантованным) выходом равна

$$R_0 = \max_{\{p_j\}}\left\{-\log_2 \sum_{j=0}^{M-1}\sum_{i=0}^{M-1} p_j p_i \int_{-\infty}^{\infty}\sqrt{p(\mathbf{y}|j)p(\mathbf{y}|i)}\,d\mathbf{y}\right\}, \qquad (7.3.6)$$

где $p(y|j)$ – условная ФПВ вектора y на выходе демодуляторов при передаче j-го сигнала. Заметим, что (7.3.6) похоже по форме на (7.3.5), за исключением того, что теперь мы имеем при обработке M-мерный интеграл, поскольку имеется M выходов демодуляторов.

Предположим, что M сигналов ортогональны, так что M выходов, обусловленных частными входными сигналами, статистически независимы. Как следствие,

$$p(\mathbf{y}|j) = p_{s+n}(y_i)\prod_{\substack{i=0\\ i\neq j}}^{M-1} p_n(y_i), \qquad (7.3.7)$$

где $p_{s+n}(y_i)$ – ФПВ выхода согласованного фильтра, соответствующего переданному сигналу с номером j, а $\{p_n(y_i)\}$ определяется сигналами на выходах остальных $M-1$ согласованных фильтров. Если (7.3.7) подставить в (7.3.6), мы получаем

$$R_0 = \max_{\{p_j\}}\left\{-\log_2\left[\sum_{j=0}^{M-1} p_j^2 + \sum_{j=0}^{M-1}\sum_{\substack{i=0\\ i\neq 1}}^{M-1} p_i p_j\left(\int_{-\infty}^{\infty} dy\sqrt{p_{s+n}(y)p_n(y)}\right)^2\right]\right\}. \qquad (7.3.8)$$

Максимизация R_0 по набору вероятностей входных сигналов приводит к $p_j = 1/M$ для $1 \le j \le M$. Следовательно, (7.3.8) ведет к

$$R_0 = \log_2\left\{\frac{M}{1+(M-1)\left[\int_{-\infty}^{\infty}\sqrt{p_{s+n}(y)p_n(y)}dy\right]^2}\right\} = \qquad (7.3.9)$$

$$= \log_2 M - \log_2\left\{1+(M-1)\left[\int_{-\infty}^{\infty}\sqrt{p_{s+n}(y)p_n(y)}dy\right]^2\right\}$$

Это – искомый результат для предельной скорости для канала с M-ичным входом и M-ичным векторным неквантованным выходом.

При когерентном детектировании M-ичных ортогональных сигналов соответствующие ФПВ равны

$$P_{s+n}(y) = \frac{1}{\sqrt{2\pi}\sigma}e^{-(y-m)^2/2\sigma^2},$$
$$p_n(y) = \frac{1}{\sqrt{2\pi}\sigma}e^{-y^2/2\sigma^2}, \qquad (7.3.10)$$

где $m = \sqrt{\mathscr{E}}$ и $\sigma^2 = \frac{1}{2}N_0$. Подставив эти соотношения в (7.3.9) и вычислив интеграл, получим

$$R_0 = \log_2\left[\frac{M}{1+(M-1)e^{-\mathscr{E}/2N_0}}\right] = \log_2\left[\frac{M}{1+(M-1)e^{-R_w\gamma_b/2}}\right], \qquad (7.3.11)$$

где $\mathscr{E}$ – энергия принимаемого сигнала, R_w – информационная скорость в битах/сигнал, а $\gamma_b = \mathscr{E}_b/N_0$ – ОСШ на бит.

Подчеркнём, что параметр скорости R_w включает в себя кодовую скорость R_c. Для примера если $M = 2$ и код двоичный то $R_w = R_c$. В более общем случае, если код двоичный и $M = 2^\nu$, тогда M-ичные сигналы содержат $R_w = \nu R_c$ бит информации. Также интересно заметить, что если код двоичный и $M = 2$, тогда (7.3.11) преобразуется к

$$R_0 = \log_2\left[\frac{2}{1+e^{-R_c\gamma_b/2}}\right], \qquad M = 2 \text{ (ортогональных сигнала)}, \qquad (7.3.12)$$

что на 3 дБ хуже, чем предельная скорость для противоположных сигналов. Если положим в (7.3.11) $R_w = R_0$ и выполним решение относительно γ_b, то получим

$$\gamma_b = \frac{2}{R_0}\ln\left(\frac{M-1}{2^{-R_0}M-1}\right) \qquad (7.3.13)$$

Зависимость R_0 от γ_b для некоторых значений M иллюстрируется на рис. 7.3.3. Заметим, что кривые насыщаются для любых значениях M при $R_0 = \log_2 M$

Интересно также рассмотреть предельную форму (7.3.11) при $M \to \infty$. Получаем

$$\lim_{M\to\infty} R_0 = \frac{\mathscr{E}}{2N_0 \ln 2}, \text{ бит/сигнал.} \qquad (7.3.14)$$

Поскольку $\mathscr{E} = P_{ср}T$, где T – длительность сигнала, то следует

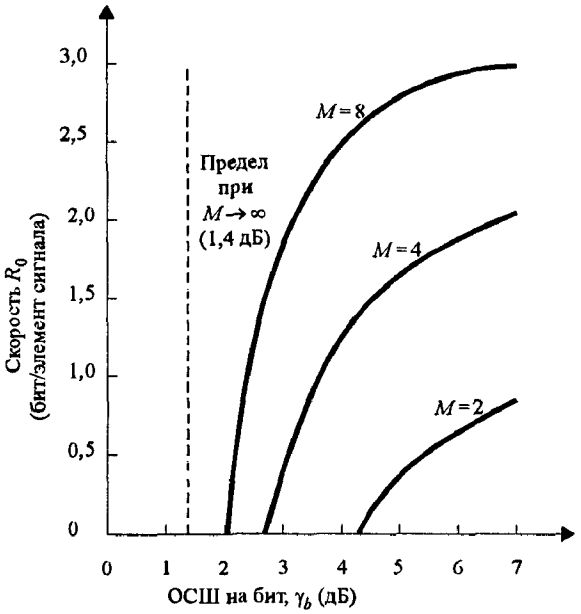


Рис. 7.3.3. Требуемое ОСШ на бит для работы с предельной скоростью R_0 при использовании M-ичных ортогональных сигналов и когерентного детектирования в канале с АБГШ

$$\lim_{M\to\infty} \frac{R_0}{T} = \frac{P_{ср}}{2N_0 \ln 2} = \tfrac{1}{2}C_\infty. \qquad (7.3.15)$$

Таким образом, при $M \to \infty$ предельная скорость равна половине пропускной способности канала с АБГШ с неограниченной полосой. Альтернативно, подставляя $\mathscr{E} = R_0\mathscr{E}_b$ в (7.3.14) даёт $\gamma_b = 2\ln 2$ (1,4 дБ), что является минимальным значением ОСШ, требуемым для работы со скоростью R_0, когда $M \to \infty$. Таким образом, при работе со скоростью R_0 требуется на 3 дБ больше мощности, чем предел Шеннона.

Величина R_0, определённая (7.3.11), базируется на использовании M ортогональных сигналов, которые субоптимальны, когда M мало. Если мы попытаемся максимизировать R_0 путём выбора наилучшего ансамбля из M сигналов, то не будем удивлены, когда найдём, что оптимальным является ансамбль симплексных сигналов. Действительно, R_0 для этих оптимальных сигналов определяется так:

$$R_0 = \log_2\left[\frac{M}{1+(M-1)e^{-M\mathscr{E}/2(M-1)N_0}}\right]. \qquad (7.3.16)$$

Если сравним это выражение с (7.3.11), то заметим, что R_0 в (7.3.16) просто отражает тот факт, что ансамбль симплексных сигналов энергетически более эффективен в $M/(M-1)$ раз.

В случае некогерентного детектирования ФПВ, соответствующая сумме сигнал + шум и одному шуму, можно выразить так:

$$p_{s+n}(y) = y\,e^{-(y^2+a^2)/2} I_0(ay), \quad y \geq 0,$$
$$p_n(y) = y\,e^{-y^2/2}, \quad y \geq 0, \qquad (7.3.17)$$

где, по определению $a = \sqrt{2\mathscr{E}/N_0}$. Расчёт R_0, даваемый (7.3.9), не ведёт к замкнутой форме решения. Вместо этого можно интеграл в (7.3.9) рассчитать численно. Результаты для этого случая были даны Джорданом (1966) и Бухером (1980). Для примера, нормированная предельная скорость R_0 для M-ичных ортогональных сигналов при некогерентном детектировании дана на рис. 7.3.4 для $M = 2,4,8$ и 16.

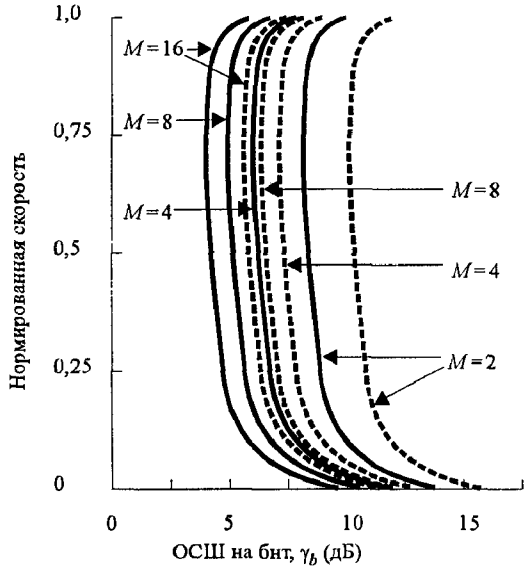


Рис. 7.3.4. Требуемое ОСШ на бит для работы с предельной скоростью R_0 при использовании M-ичных ортогональных сигналов и некогерентного детектирования в канале с АБГШ

С целью сравнения мы также дали кривые для предельной скорости при декодировании мягких решений ($Q = M$) для M-ичных сигналов. В этом случае имеем

$$R_Q = \log_2 \left\{ \frac{M}{\left[\sqrt{(1-P_M)} + \sqrt{(M-1)P_M}\right]^2} \right\}, \quad Q = M, \qquad (7.3.18)$$

где P_M – вероятность ошибки символа. Для относительно широкого диапазона скоростей, разница между декодированием мягких и жёстких решений составляет приблизительно 2 дБ.

Наиболее удивительное в характеристиках качества, отображённых на рис. 7.3.4, – это то, что здесь имеется оптимальная скорость кода при малом M. В отличие от когерентного детектирования, когда ОСШ на бит уменьшается монотонно с уменьшением скорости кода, ОСШ на бит при некогерентном детектировании достигает минимума вблизи нормированной скорости 0,5 и увеличивается как при большей, так и при меньшей скоростях. Минимум довольно широкий, так что здесь имеется область скоростей от 0,2 до 0,9, где ОСШ на бит имеет минимум с разбросом до 1 дБ. Такое характерное поведение характеристики качества при некогерентном детектировании объясняется нелинейной характеристикой детектора.

7.4. БИБЛИОГРАФИЧЕСКИЕ ЗАМЕЧАНИЯ И ССЫЛКИ

Основополагающая работа по описанию канала в терминах его пропускной способности и случайного кодирования принадлежит Шеннону (1948a, b, 1949). Дополнительный вклад был впоследствии сделан Гильбертом (1952), Элиасом (1955), Галлагером (1965), Вайнером (1965), Шенноном и др. (1967), Форни (1968) и Витерби (1969). Все эти ранние публикации собраны в IEEE Press book под заголовком *Key Papers in the Development of Information Theory* под редакцией Слепяна (1974).

Использование параметра предельной скорости в качестве критерия синтеза было предложено и разработано Возенкрафтом и Кеннеди (1966) и Возенкрафтом и Джекобсом (1965). Он был использован Джорданом (1966) при синтезе кодированных сигналов для M-ичных ортогональных сигналов с когерентным и некогерентным детектированием. Следуя этим основополагающим работам, предельная скорость была широко использована в качестве критерия синтеза кодовых сигналов для множества различных условий в канале.

ЗАДАЧИ

7.1. Покажите, что следующие два отношения являются необходимыми и достаточными условиями для ансамбля входных вероятностей $\{P(x_j)\}$, чтобы максимизировать $I(X;Y)$ и, таким образом, достичь пропускной способности ДКБП:

$$I(x_j;Y) \le C \qquad \text{для всех } j \text{ с } P(x_j) > 0,$$
$$I(x_j;Y) = C \qquad \text{для всех } j \text{ с } P(x_j) = 0,$$

где C – пропускная способность канала и

$$I(x_j;Y) = \sum_{i=0}^{Q-1} P(y_i \mid x_j) \log \frac{P(y_i \mid x_j)}{P(y_i)}.$$

7.2. Рис. Р.7.2 иллюстрирует M-ичный симметричный ДКБП с переходными вероятностями $P(y \mid x) = 1 - p$, когда $x = y = k$, для $k = 0,1,2,\ldots,M-1$, и $P(y \mid x) = p/(M-1)$, когда $x \neq y$.

a) Покажите, что этот канал удовлетворяет условиям, данным в задаче 7.1, когда $P(x_k) = 1/M$.

b) Определите и постройте зависимость пропускной способности канала от p.

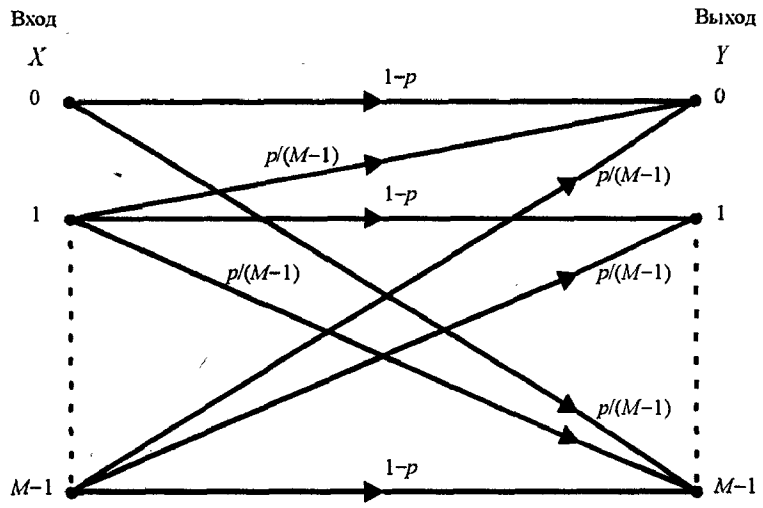


Рис. Р.7.2

7.3. Определите пропускные способности каналов, показанных на рис. Р.7.3.

7.4. Рассмотрите два канала с переходными вероятностями, показанными на рис. Р.7.4. Определите, достигается ли максимизация скорости передачи по каналу при равной вероятности входных символов.

7.5. Телефонный канал имеет полосу частот $W = 3000$ Гц и отношение мощностей сигнала и шума 400 (26 дБ). Предположим, что мы характеризуем канал, как частотно-ограниченный канал с АБГШ с $P_{ср} / WN_0 = 400$.

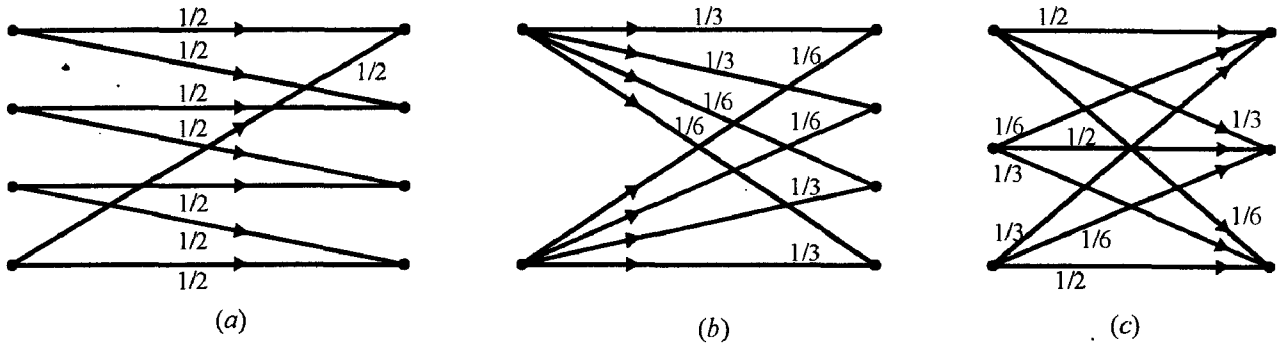


Рис. Р.7.3

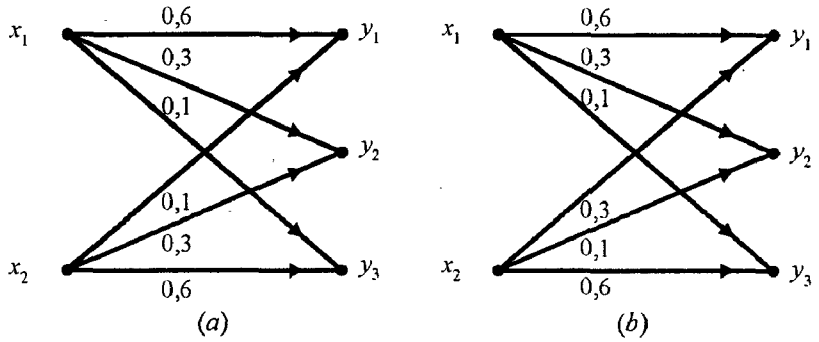


Рис. Р.7.4

a) Определите пропускную способность канала в бит/с.

b) Достаточно ли пропускная способность канала для обеспечения передачи речевого сигнала, который стробируется и кодируется посредством логарифмической ИКМ?

c) Обычно канальные искажения, отличные от аддитивного шума, ограничивают скорость передачи по телефонному каналу меньше, чем пропускная способность эквивалентного частотно-ограниченного канала с АБГШ, рассмотренном в п. (a). Предположим, что на практике достигнута скорость передачи $0{,}7C$ без канального кодирования.

Какие из методов кодирования речевого источника, описанные в разделе 3.5, обеспечат достаточное сжатие, чтобы удовлетворить частотным ограничениям телефонного канала?

7.6. Рассмотрите дискретный канал без памяти (ДКБП) с двоичным входом и четверичным выходом, показанный на рис. Р.7.6.

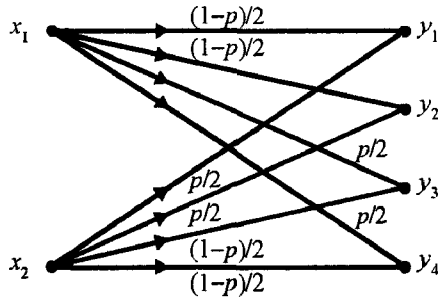


Рис. Р.7.6

a) определите пропускную способность этого канала;

b) покажите, что этот канал эквивалентен ДСК.

7.7. Определите пропускную способность канала, показанного на рис. 7.7.

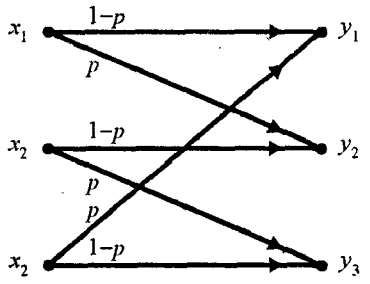


Рис. Р.7.7

7.8. Рассмотрнте ДСК с переходной вероятностью ошибки p. Допустим, что R – это число бит в кодовых словах, которые представляются одним из 2^R возможных уровней на выходе квантователя. Определите:

a) вероятность того, что кодовое слово, переданное по ДСК, принимается без ошибок;

b) вероятность иметь ошибки хотя бы в одном бите (символе?) кодового слова;

c) вероятность иметь n или меньше ошибок в кодовом слове;

d) вычислите вероятность (a), (b) и (c) для $R = 5$, $p = 0{,}01$ и $n_e = 5$.

7.9. Покажите, что для ДСК средняя взаимная информация между последовательностью $X_1X_2...X_n$ канальных входов и соответствующих канальных выходов удовлетворяют условию

$$I(X_1, X_2, ..., X_n; Y_1, Y_2, ..., Y_n) \le \sum_{i=1}^{n} I(X_i; Y_i)$$

с равенством тогда и только тогда, когда символы в последовательности независимы.

7.10. Рис.Р.7.10 иллюстрирует двоичный канал со стиранием с переходными вероятностями $P(0|0) = P(1|1) = 1 - p$ и $P(e|0) = P(e|1) = p$. Вероятности входных символов равны $P(X = 0) = \alpha$ и $P(X = 1) = 1 - \alpha$.

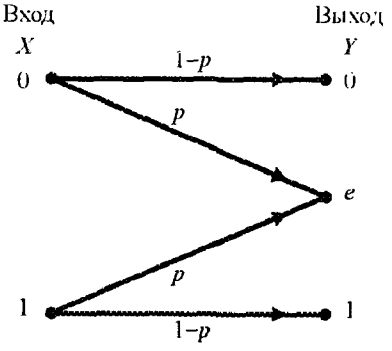


Рис. Р.7.10

a) определите среднюю взаимную информацию $I(X;Y)$ в битах;

b) определите величину α, которая максимизирует $I(X;Y)$, т.е. обеспечивает пропускную способность канала в бит/символ и постройте C как функцию от p для оптимального значеиия α;

c) Для величины α, найденной в (b), определите взаимную информацию $I(x;y) = I(0;0)$, $I(1;1)$, $I(0;e)$ и $I(1;e)$.

7.11. Рассмотрите канал с двоичным входом и троичным выходом с переходными вероятностями, показанными на рис. 7.11, где e означают символы стирания.

Для канала с АБГШ величины α и p определяются так:

$$\alpha = \frac{1}{\sqrt{\pi N_0}} \int_{-\beta}^{\beta} e^{-(x+\sqrt{\mathscr{E}_c})^2 / N_0} dx,$$

$$p = \frac{1}{\sqrt{\pi N_0}} \int_{\beta}^{\infty} e^{-(x+\sqrt{\mathscr{E}_c})^2 / N_0} dx.$$

a) определите R_Q для $Q = 3$ как функцию вероятностей α и p.

b) параметр скорости R_Q зависит от выбора порога в через вероятности α и p. Для определеиных значений $\mathscr{E}_c/N_0$ величину β, которая максимизирует R_Q, можно определить методом проб и ошибок. Например, можно показать, что для $\mathscr{E}_c/N_0$ ниже 0 дБ $\beta_{опт} = 0{,}65\sqrt{N_0/2}$ для $1 \le \mathscr{E}_c / N_0 \le 10$, $\beta_{опт}$ меняется примерно линейно между $0{,}65\sqrt{N_0/2}$ и $1{,}0\sqrt{N_0/2}$ Используя $\beta = 0{,}65\sqrt{N_0/2}$ для всей области $\mathscr{E}_c/N_0$, нарисуйте R_Q как функцию $\mathscr{E}_c/N_0$ и сравните этот результат с R_Q $(\cdot \text{рЦ}\, Q = \infty)$

7.12. Найдите пропускную способность каскадного соединения n двоичных снмметричных каналов с одинаковой вероятностью стирания ε.

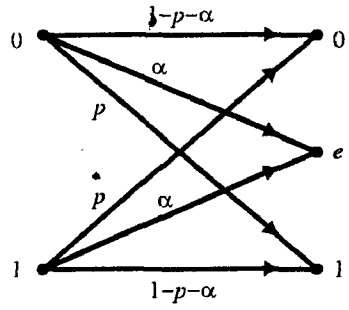


Рис. Р.7.11

Какова пропускная способность, когда число каналов становится неограниченным?

7.13. Каналы 1, 2 и 3 показаны на рис. Р.7.13.

а) найдите пропускную способность канала 1. При каком распределении входа она достигается?

b) найдите пропускную способность канала 2. При каком распределении входа она достигается?

c) Пусть C означает пропускную способность третьего канала, а C_1 и C_2 – пропускные способности первого и второго канала. Какие из следующих соотношений верны и почему?

$$C < \tfrac{1}{2}(C_1 + C_2) \qquad \text{(i)}$$

$$C = \tfrac{1}{2}(C_1 + C_2) \qquad \text{(ii)}$$

$$C > \tfrac{1}{2}(C_1 + C_2) \qquad \text{(iii)}$$

7.14. Пусть C означает пропускную способность дискретного канала без памяти со входным алфавитом $\mathcal{X} = \{x_1, x_2, \ldots x_N\}$ и выходным алфавитом $\mathcal{Y} = \{y_1, y_2, \ldots y_M\}$. Покажите, что $C \le \min\{\log M, \log N\}$.

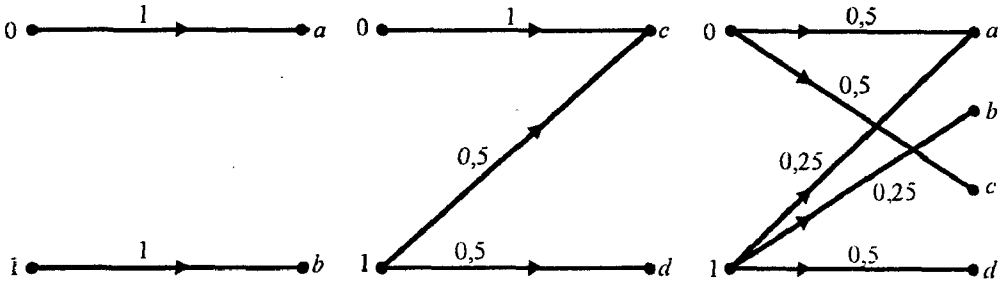


Рис. Р.7.13

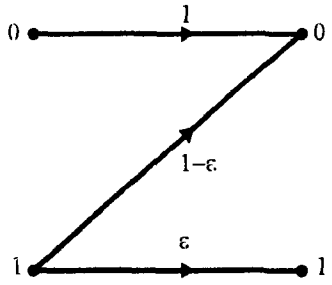


Рис. Р.7.15

7.15. Канал (иазовем его Z) показан на рис. Р.7.15.

а) Найдите распределение входов, при котором достигается пропускная способность;

b) Каково распределение входа и пропускная способность канала для специальных случаев $\varepsilon = 0$, $\varepsilon = 1$ и $\varepsilon = 0{,}5$;

c) Покажите, что если n таких каналов соединены каскадно, то результирующий канал эквивалентен каналу Z с $\varepsilon_1 = \varepsilon^n$

d) Какова пропускная способность эквивалентного Z канала, когда $n \to \infty$.

7.16. Найдите пропускную способность канала с АБГШ с полосой частот 1 МГц, мощностью сигнала 10 Вт и спектральной плотностью мощиости шума $\frac{1}{2}N_0 = 10^{-9}$Вт/Гц

7.17. Канал C_1 – это канал с АБГШ с полосой W, средней мощностью сигнала P и спектральной плотностью шума $\frac{1}{2}N_0$. Каиал C_2 – это канал с аддитивным гауссовским шумом с той же полосой и мощностью сигнала, как в канале C_1, но со спектральной плотностью мощности шума $\Phi_n(f)$. Далее предположим, что суммарная мощность шума для обоих каналов одинакова, т.е.

$$\int_{-W}^{W} \Phi_n(f)df = \int_{-W}^{W} \tfrac{1}{2}N_0 df = N_0 W.$$

Как вы думаете, какой из каналов имеет большую пропускную способность?

Дайте интуитивные объяснения.

7.18. Информация от гауссовского источника без памяти с дискретным временем с нулевым средним и

дисперсией σ^2 должна быть передана по двоичному симметричному каналу с переходной вероятностью ошибки p.

a) Какова минимальная величина достижимого искажения в месте назначения (искажения измеряются среднеквадратичной ошибкой)?

b) Если канал без памяти с АБГШ, с дискретным временем, с мощностью входа P и мощностью шума P_n, каково минимально достижимое искажение?

c) Теперь предположите, что источник имеет те же базовые свойства, но он с памятью. Уменьшится или увеличится искажение при передаче информации по ДСК? Почему?

7.19. X – это двоичный источник без памяти с $P(X=0)=0{,}3$. Информация от источника передается по ДСК с переходной вероятностью $p=0{,}1$.

a) Предположите, что источник непосредственно соединен с каналом, т.е. кодирование не используется. Какова вероятность ошибки в месте назначения?

b) Если в канале произведено кодирование, то какова минимально возможная вероятность ошибки принимаемого сообщения?

c) При каких значениях p возможная надёжная связь (разумеется, с кодированием)?

7.20. Нарисуйте зависимость пропускной способности от $\mathscr{E}_b/N_0$ для канала с АБГШ, который передаёт двоичные противоположные сигналы и использует на приёме оптимальный поэлементный детектор. На тех же осях нарисуйте зависимость пропускной способности того же канала, если передаются двоичные ортогональные сигналы

7.21. В системе связи с кодированием передаются M сообщений $1, 2, \ldots M=2^k$ посредством M *базовых* сигналов $x_1(t)$, $x_2(t)$, ... $x_M(t)$, каждый длительностью nT. Общая форма для сигнала $x_i(t)$ такова:

$$x_i(t)=\sum_{j=0}^{n-1} f_{ij}(t-jT),$$

где $f_{ij}(t)$ может быть одним из двух сигналов $f_1(t)$ или $f_2(t)$, причем $f_1(t)=f_2(t)\equiv 0$ при $t\notin[0,T]$. Кроме того, предположим, что $f_1(t)$ и $f_2(t)$ имеют равную энергию $\mathscr{E}$, а канал идеальный (нет затухания) с АБГШ со спектральной плотностью $\frac{1}{2}N_0$. Это значит, что принимаемый сигнал $r(t)=x(t)+n(t)$, где $x(t)$ – один из сигналов $x_i(t)$, а $n(t)$ – шум.

a) При $f_1(t)=-f_2(t)$ покажите, что размерность пространства сигналов N удовлетворяет условию $N\le n$;

b) Покажите, что в общем $N\le 2n$;

c) При $M=2$ покажите, что при общих $f_1(t)$ и $f_2(t)$

$$p(\text{ошибки}|\text{пер. } x_1(t))\le \int_{R^N}\ldots\int \sqrt{p(\mathbf{r}|\mathbf{x}_1)p(\mathbf{r}|\mathbf{x}_2)}\,d\mathbf{r},$$

где $\mathbf{r}$, $\mathbf{x}_1$ и $\mathbf{x}_2$ – представления для $r(t)$, $x_2(t)$ и $x_2(t)$ в N-мерном пространстве, $d\mathbf{r}$ – элемент объёма пространства.

d) Используя результат (c), покажите, что для любого M

$$p(\text{ошибки}|\text{пер. } x_m(t))\le \sum_{\substack{1\le m'\le M\\ m'\ne m}} \int_{R^N}\ldots\int \sqrt{p(\mathbf{r}|\mathbf{x}_m)p(\mathbf{r}|\mathbf{x}_{m'})}\,d\mathbf{r}\ .$$

e) Покажите, что

$$\int_{R^N}\ldots\int \sqrt{p(\mathbf{r}|\mathbf{x}_m)p(\mathbf{r}|\mathbf{x}_{m'})}\,d\mathbf{r}\ =\exp\left(-\frac{|\mathbf{x}_m-\mathbf{x}_{m'}|^2}{4N_0}\right)$$

и, следовательно,

$$p(\text{ошибки}|\text{пер. } x_m(t))\le \sum_{\substack{1\le m'\le M\\ m'\ne m}} \exp\left(-\frac{|\mathbf{x}_m-\mathbf{x}_{m'}|^2}{4N_0}\right).$$

8

БЛОКОВЫЕ И СВЁРТОЧНЫЕ КОДЫ

В гл. 7 мы рассмотрели кодирование и декодирование в канале с общей точки зрения и показали, что даже случайно выбранные коды в среднем дают качество, близкое к пропускной способности канала Для случая ортогональных сигналов мы показали, что можно достичь предела пропускной способности канала, если число сигналов не ограничивать.

В этой главе мы опишем специальные коды и рассчитаем их качество для канала с АБГШ. В частности, мы рассмотрим два класса кодов, именно линейные блоковые и свёрточные коды. Качество кода рассчитывается как для декодирования жёстких решений, так и для декодирования мягких решений.

8.1. ЛИНЕЙНЫЕ БЛОКОВЫЕ КОДЫ

Блоковый код состоит из набора векторов фиксированной длины, называемых кодовыми словами. Длина кодового слова – это число элементов в векторах, и оно обозначается n. Элементы кодового слова выбираются из алфавита с q элементами. Если алфавит содержит два элемента 0 и 1, код называется двоичным, а элементы любого кодового слова называют битами (двоичными символами). Если элементы кодового слова выбираются из алфавита, имеющего q элементов ($q>2$), код называют не двоичным. Интересно отметить, что если q является степенью 2, т.е. $q = 2^b$, где b – положительное целое число, каждый q-й элемент имеет эквивалентное двоичное представление, состоящее из b битов и, таким образом, недвоичные коды длины N можно отобразить в двоичный код с блоковой длиной $n = bN$.

В двоичном блоковом коде длиной n можно образовать 2^n кодовых слов. Из этих 2^n кодовых слов мы можем выбрать $M = 2^k$ кодовых слов ($k<n$), чтобы сформировать код. Таким образом, блок из k информационных бит отображается в кодовое слово длины n, выбираемое из набора $M = 2^k$ кодовых слов. Мы обозначим результирующий блоковый код, как (n,k) код, а отношение $k/n \equiv R_c$ определим как *скорость* кода. В более общем случае, если код имеет q элементов, можно образовать q^n кодовых слов. Подмножество из $M = 2^k$ кодовых слов можно выбрать для передачи k битовых информационных блоков.

Кроме параметра скорости кода R_c, важным параметром кодового слова является его *вес*, который равен числу ненулевых элементов, слова. В общем, каждое кодовое слово имеет свой собственный вес. Набор всех весов кода образует *распределение весов* кода. Когда все M кодовых слов имеет одинаковый вес, код называется *кодом с фиксированным весом* или *кодом с постоянным весом*. Функции кодирования и декодирования включают арифметические операции суммирования и умножения, выполненные над кодовыми словами. Эти арифметические операции выполняются в соответствии с соотношениями, правилами для алгебраического поля, которое имеет своими элементами символы, содержащиеся в алфавите кода. Для примера, символы в двоичном алфавите равны 0 и 1; следовательно, поле имеет два элемента. В общем, поле F состоит из набора элементов,

над которыми выполняются две арифметические операции, именно, операции сложения и умножения, которые удовлетворяют следующим свойствам (аксиомам):

Сложение

1. Набор F замкнут относительно сложения, т.е. если $a, b \in F$, тогда $a+b \in F$.
2. Сложение ассоциативно, т.е. если a, b и c – элементы F, тогда $a+(b+c)=(a+b)+c$.
3. Сложение коммутативно, т.е. $a+b=b+a$.
4. Набор содержит элемент, называемый *нулевым*, который удовлетворяет условию $a+0=a$.
5. Каждый элемент множества имеет свой собственный отрицательный элемент. Следовательно если b – элемент, то его отрицательный элемент обозначается $-b$. Вычитание двух элементов, такие как $a-b$ определено как $a+(-b)$.

Умножение

1. Набор F замкнут относительно умножения, т.е. если $a, b \in F$, то тогда $ab \in F$.
2. Умножение ассоциативно, т.е. если a, b и c – элементы F, тогда $a(bc)=(ab)c$.
3. Умножение коммутативно, т.е. $ab=ba$.
4. Умножение дистрибутивно со сложением, т.е. $(a+b)c=ac+bc$.
5. Набор содержит элемент, называемый единичным, который удовлетворяет условию $a \cdot 1=a$ для любого элемента $a \in F$.
6. Каждый элемент F, исключая нулевой, имеет обратный элемент. Следовательно если $b \in F$ $(b \neq 0)$, тогда его обратный элемент определён как b^{-1} и $bb^{-1}=1$. Деление двух элементов, обозначаемое как $a \div b$ или a/b, определено как $a \cdot b^{-1}$.

Мы очень свободно обращаемся с полями из вещественных чисел и полями из комплексных чисел. Эти поля могут иметь неограниченное число элементов. Однако, как было указано выше, коды строятся из полей с ограниченным числом элементов. Ограниченное поле с q элементами обычно называют *полем Галуа* и обозначают GF(q).

Каждое поле должно иметь нулевой элемент и единичный элемент – следовательно, простейшее поле – это GF(2). В общем, если q является простым числом, мы можем построить ограниченное поле GF(q), состоящие из элементов $\{0, 1, \ldots, q-1\}$. Операции суммирования и умножения над элементами из GF(q) осуществляются по модулю q и обозначается так (mod q). Например, таблицы сложения и умножения для GF(2) таковы:

+	0	1
0	0	1
1	1	0

•	0	1
0	0	0
1	0	1

Подобным образом, поле GF(5)содержит набор, состоящий из элементов {0, 1, 2, 3, 4}. Таблицы сложения и умножения для GF(5):

+	0	1	2	3	4
0	0	1	2	3	4
1	1	2	3	4	0
2	2	3	4	0	1
3	3	4	0	1	2
4	4	0	1	2	3

•	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

В общем, ограниченное поле GF(q) можно построить, только если q – простое число или степень простого числа. Если q – простое число, умножение и сложение базируются на арифметике по модулю q, как сказано выше. Если $q = p^m$, где p – простое число, а m – какое-либо целое, возможно расширить поле $\mathrm{GF}(p)$ до поля $\mathrm{GF}(p^m)$. Последнее называется *расширенным полем* $\mathrm{GF}(p)$. Умножение и сложение элементов в расширенном поле базируется на арифметике по модулю p.

С этим кратким введением в арифметику операций, которые можно осуществлять над элементами кодовых слов, рассмотрим теперь некоторые базовые характеристики блоковых кодов.

Предположим, что $\mathbf{C}_i$ и $\mathbf{C}_j$ – какие либо два кодовых слова в (n,k) кодовом блоке. Мера разницы между кодовыми словами – число соответствующих элементов или позиций, в которых они различаются. Эта мера называется расстоянием Хемминга между двумя кодовыми словами и обозначается d_{ij}. Ясно, что d_{ij} при $i \neq j$ удовлетворяет условию $0 < d_{ij} \leq n$. Наименьшее значение из набора $\{d_{ij}\}$ для M кодовых слов называется минимальным расстоянием кода и обозначается $d_{\min}$. Поскольку хеммингово расстояние является мерой различия между парами кодовых слов, оно, разумеется, имеет отношение к коэффициенту корреляции между соответствующими парами сигналов, генерируемыми кодовыми словами. Эта связь обсуждается в разделе 8.1.4.

Помимо классификации кодов на двоичные и недвоичные можно также их классифицировать на линейные и нелинейные. Возьмем $\mathbf{C}_i$ и $\mathbf{C}_j$ – два кодовых слова в блоковом коде (n,k) и α_1, α_2 – какие-либо два скалярных элемента из определённого алфавита. Тогда код называют линейным, если и только если $\alpha_1\mathbf{C}_i + \alpha_2\mathbf{C}_j$ тоже является кодовым словом. Это определение подразумевает, что линейный код должен содержать кодовое слово из одних нулей. Как следствие, код с постоянным весом нелинейный.

Предположим, что мы имеем двоичный линейный блоковый код, и пусть $\mathbf{C}_i$, $i = 1, 2, \ldots, M$, где M – число кодовых слов. Для удобства пусть $\mathbf{C}_1$ означает кодовое слово из одних нулей, т.е. $\mathbf{C}_1 = [0\,0\ldots0]$ и пусть w_r означает вес r-го кодового слова. Отсюда следует, что w_r является хемминговым расстоянием между кодовыми словами $\mathbf{C}_r$ и $\mathbf{C}_1$. Т.е. расстояние $d_{1r} = w_r$. В общем, расстояние d_{ij} между парой кодовых слов $\mathbf{C}_i$ и $\mathbf{C}_j$ просто равно весу кодового слова, сформулированного разностью между $\mathbf{C}_i$ и $\mathbf{C}_j$. Поскольку код линейный, разность (что эквивалентно взятию суммы по mod 2 двоичных кодовых слов) между $\mathbf{C}_i$ и $\mathbf{C}_j$ также является кодовым словом с весом, включенным в набор $\{w_r\}$. Следовательно, распределение весов линейного кода полностью характеризует дистанционные свойства кода. Минимальное расстояние кода равно

$$d_{\min} = \min_{r, r \neq 1}\{w_r\} \qquad (8.1.1)$$

Определённое число элементарных понятий из линейной алгебры особенно полезны, когда имеем дело с линейными блоковыми кодами. В частности, набор из всех векторов с n элементами формирует векторное пространство S. Если мы выберем набор из $k < n$ линейно независимых векторов из S и из них сформируем набор из всех линейных комбинаций этих векторов, то результирующий набор образует подпространство в S, назовем его S_c размерности k. Любой набор из k линейно независимых векторов в пространстве S_c образует базис. Теперь рассмотрим набор из векторов S, которые

ортогональны к каждому вектору базиса S_c (и, следовательно, они ортогональны ко всем векторам в S_c). Этот набор векторов также является подпространством S и он называется *нуль-пространством* или *ортогональным пространством* к S_c. Если размерность пространства S_c равна k, то размерность нуль-пространства равна $n-k$.

Если пользоваться терминами, предназначенными для двоичных блоковых кодов, векторное пространство S состоит из 2^n двоичных векторов с n элементами. Линейный (n,k) код является ансамблем 2^k векторов с n элементами, называемыми кодовыми словами, которые формируют подпространство S_c в поле из двух элементов. Поскольку имеется 2^k кодовых слов в S_c, базис для S_c имеет k кодовых слов. Это значит, что для конструирования 2^k линейных комбинаций требуется k линейно независимых кодовых слов, которые формируют весь код. Нуль-пространство для S_c образует другой линейный код, который состоит из 2^{n-k} кодовых слов блока длиной n с $n-k$ информационными битами. Его размерность равна $n-k$. В разделе (8.1.1) мы рассмотрим эти соотношения с большими подробностями.

8.1.1. Порождающая и проверочная матрицы

Пусть x_{m1}, x_{m2}, ..., x_{mk} означает k информационных бит, кодируемых в кодовое слово $\mathbf{C}_m$. В этой главе мы следуем установленным соглашениям о представлении кодовых слов в виде векторов. Так, вектор из k информационных бит на входе кодера обозначается так:

$$\mathbf{X}_m = \left[x_{m1} \;\; x_{m2} \;\; \ldots \; x_{mk}\right],$$

а выходом кодера является вектор из n символов

$$\mathbf{C}_m = \left[c_{m1} \;\; c_{m2} \;\; \ldots \; c_{mn}\right].$$

Операцию кодирования, выполняемую в линейном двоичном блоковом коде, можно представить совокупностью из n уравнений вида

$$c_m = x_{m1}g_{1j} + x_{m2}g_{2j} + \ldots + x_{mk}g_{kj}\;,\;\; j = 1,2,\ldots,n\,, \qquad (8.1.2)$$

где $g_{ij} = 0$ или 1, а $x_{mi}g_{ij}$ представляют произведение x_{mi} и g_{ij}. Линейные уравнения (8.1.2) можно также представить в матричной форме

$$\mathbf{C}_m = \mathbf{X}_m\mathbf{G}, \qquad (8.1.3)$$

где $\mathbf{G}$ – порождающая матрица кода, равная

$$\mathbf{G} = \begin{bmatrix} \leftarrow \mathbf{g}_1 \rightarrow \\ \leftarrow \mathbf{g}_2 \rightarrow \\ \vdots \\ \leftarrow \mathbf{g}_k \rightarrow \end{bmatrix} = \begin{bmatrix} g_{11} & g_{12} & \cdots & g_{1n} \\ g_{21} & g_{22} & \cdots & g_{2n} \\ \vdots & \vdots & & \vdots \\ g_{k1} & g_{k2} & \cdots & g_{kn} \end{bmatrix}. \qquad (8.1.4)$$

Заметим, что произвольное кодовое слово – это просто линейная комбинация векторов $\{\mathbf{g}_i\}$ из $\mathbf{G}$,

$$\mathbf{C}_m = x_{m1}\mathbf{g}_1 + x_{m2}\mathbf{g}_2 + \ldots + x_{mk}\mathbf{g}_k\;. \qquad (8.1.5)$$

Поскольку линейный (n,k) код с 2^k кодовыми словами является подпространством размерности k, векторы $\{\mathbf{g}_i\}$ порождающей матрицы $\mathbf{G}$ должны быть линейно независимыми, т.е. они должны образовывать пространство размерности k. Другими словами, $\{\mathbf{g}_i\}$ должны образовать базис для (n,k) кода. Заметим, что ансамбль базовых

векторов не единственный и, следовательно, **G** не уникальна. Мы также заметим, что, поскольку пространство имеет размерность k, ранг матрицы **G** равен k.

Любую порождающую матрицу (n,k) кода путём проведения операций над строками (и перестановкой столбцов) можно свести к «систематической форме»:

$$\mathbf{G}=\left[\mathbf{I}_k \vdots \mathbf{P}\right]=\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & p_{11} & p_{12} & \cdots & p_{1n-k} \\ 0 & 1 & 0 & \cdots & 0 & p_{21} & p_{22} & \cdots & p_{2n-k} \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 & p_{k1} & p_{k2} & \cdots & p_{kn-k} \end{bmatrix}, \qquad (8.1.6)$$

где $\mathrm{I}_k - k \times k$ единичная матрица, а $\mathbf{P} - k \times (n-k)$ матрица, которая определяет $n-k$ избыточных или проверочных символов. Заметим, что порождающая матрица систематического кода создаёт линейный блоковый код, в котором первые k бит любого кодового слова идентичны информационным битам, а остающиеся $n-k$ бит любого кодового слова являются линейными комбинациями k информационных бит. Эти $(n-k)$ избыточных бита называют паритетными (проверочными) битами. Результирующий (n,k) код называется в этом случае *систематическим кодом.*

Если (n,k) код порождён матрицей, не имеющей систематической формы (8.1.6), он называется несистематическим. Однако . такая матрица эквивалентна матрице в систематической форме в том смысле, что одна может быть получена из другой элементарными операциями над строками и перемещением столбцов. Два (n,k) линейных кода, порожденных двумя эквивалентными порождающими матрицами, называют эквивалентными и один может быть получен из другого перестановкой элементов. Таким образом, каждый линейный (n,k) код эквивалентен линейному систематическому (n,k) коду.

Пример 8.1.1. Рассмотрим код (7, 4) с порождающей матрицей

$$\mathbf{G}=\begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}=\left[\mathbf{I}_4 \vdots \mathbf{P}\right]. \qquad (8.1.7)$$

Типичное кодовое слово можно выразить так:

$\mathbf{C}_m=\left[x_{m1}\ x_{m2}\ x_{m3}\ x_{m4}\ c_{m5}\ c_{m6}\ c_{m7}\right]$, где $\{x_{mj}\}$ представляют четыре информационных бита, а $\{c_{mj}\}$ представляют три паритетных бита, определённых так:

$$\begin{aligned} c_{m5} &= x_{m1}+x_{m2}+x_{m3}, \\ c_{m6} &= x_{m2}+x_{m3}+x_{m4}, \\ c_{m7} &= x_{m1}+x_{m2}+x_{m4}. \end{aligned} \qquad (8.1.8)$$

Линейный систематический (n, k) двоичный блоковый кодер можно реализовать, используя k-битовый регистр сдвига, $n-k$ сумматоров (mod 2), связанных с соответствующими ячейками регистра сдвига и генерирующих проверочные символы, которые потом временно располагаются во втором регистре сдвига длины $n-k$. Затем k информационных бита, а за ними $n-k$ проверочных бита последовательно покидают два регистра и подаются на модулятор. Это кодирование иллюстрируется рис. 8.1.1 для кода (7, 4) из примера (8.1.1)

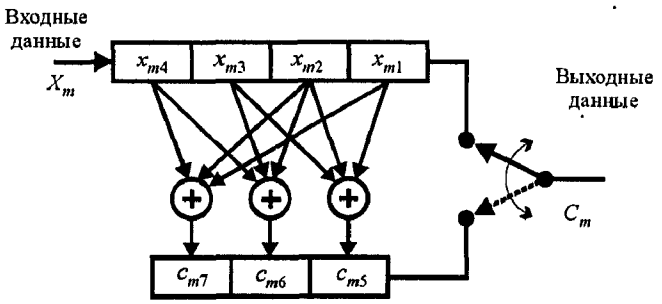


Рис. 8.1.1. Линейный регистр сдвига для получения двоичного кода (7,4)

С любым линейным кодом (n, k) кодом связан дуальный код размерностью $n-k$.

Дуальный код является линейным $(n, n-k)$ кодом с 2^{n-k} кодовыми векторами, которое образуют нуль-пространство по отношению к (n, k) коду. Порождающая матрица для дуального кода, обозначаемая **H**, состоит из $n-k$ линейно независимых кодовых векторов, выбираемых в нуль-пространстве. Любое кодовое слово $\mathbf{C}_m$ из (n, k) кода ортогонально любому кодовому слову дуального кода. Следовательно, любое кодовое слово (n, k) кода ортогонально любой строке матрицы **H**, т.е.

$$\mathbf{C}_m\mathbf{H}^{\mathrm{T}} = \mathbf{0}, \tag{8.1.9}$$

где **0** означает вектор-столбец, состоящий из $n-k$ нулей, а $\mathbf{C}_m$ – кодовое слово (n, k) кода. Поскольку (8.1.9) справедливо для любого кодового слова (n, k) кода, то следует

$$\mathbf{G}\mathbf{H}^{\mathrm{T}} = \mathbf{0}, \tag{8.1.10}$$

где **0**-теперь $k \times (n-k)$ матрица со всеми нулевыми элементами.

Теперь предположим, что линейный (n, k) код является систематическим, и его порождающая матрица дана в систематической форме (8.1.6). Тогда, поскольку $\mathbf{G}\mathbf{H}^{\mathrm{T}} = \mathbf{0}$, следует, что

$$\mathbf{H} = \left[-\mathbf{P}^{\mathrm{T}} \vdots \mathbf{I}_{n-k}\right] \tag{8.1.11}$$

Отрицательный знак в (8.1.11) может быть опущен при работе с двоичными кодами, поскольку вычитание по mod2 идентично сложению по mod2.

Пример (8.1.2). Для систематического кода (7, 4), генерируемого матрицей **G**, определяемой (8.1.7), имеем согласно (8.1.11) матрицу **H** в виде

$$\mathbf{H} = \left[\begin{array}{ccccccc} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{array}\right]. \tag{8.1.12}$$

Теперь уравнение $\mathbf{C}_m\mathbf{H}^{\mathrm{T}} = \mathbf{0}$ распадается на три уравнения

$$\begin{aligned} x_{m1} + x_{m2} + x_{m3} + c_{m5} &= 0, \\ x_{m2} + x_{m3} + x_{m4} + c_{m6} &= 0, \\ x_{m1} + x_{m2} + x_{m4} + c_{m7} &= 0. \end{aligned} \tag{8.1.13}$$

Таким образом, видим, что произведение $\mathbf{C}_m\mathbf{H}^{\mathrm{T}}$ эквивалентно суммированию проверочных символов с соответствующими линейными комбинациями информационных символов, используемых для вычисления c_{mj}, $j = 5, 6, 7$. Это значит, что (8.1.13) эквивалентно (8.1.8).

Матрицу $\mathbf{H}$ можно использовать в декодере для проверки того, удовлетворяет ли принимаемое кодовое слово $\mathbf{Y}$ условию (8.1.13), т.е. $\mathbf{Y}\mathbf{H}^{\mathrm{T}} = \mathbf{0}$. Таким образом, декодер сверяет принятые проверочные символы с соответствующей линейной комбинацией символов y_1, y_2, y_3 и y_4, которые формируют проверочные символы на передаче. Поэтому принято называть $\mathbf{H}$ *проверочной матрицей*, связанной с (n, k) кодом.

Выскажем соображение, касающееся связи минимального расстояния кода и его проверочной матрицы $\mathbf{H}$. Произведение $\mathbf{C}_m\mathbf{H}^{\mathrm{T}}$ с $\mathbf{C}_m \neq 0$ представляют линейную комбинацию n столбцов $\mathbf{H}^{\mathrm{T}}$. Поскольку $\mathbf{C}_m\mathbf{H}^{\mathrm{T}} = \mathbf{0}$, векторы-столбцы $\mathbf{H}$ линейно зависимы. Допустим, что $\mathbf{C}_j$ означает кодовое слово с минимальным весом линейного кода (n, k). Оно должно удовлетворять условию $\mathbf{C}_j\mathbf{H}^{\mathrm{T}} = \mathbf{0}$. Поскольку минимальный вес равен минимальному расстоянию, следует, что $d_{\min}$ столбцов $\mathbf{H}$ линейно зависимы. Альтернативно, мы можем сказать, что не более, чем $d_{\min} - 1$ столбцов $\mathbf{H}$ линейно независимы. Поскольку ранг матрицы $\mathbf{H}$ не больше $n-k$, имеем $n-k \geq d_{\min} - 1$. Следовательно, $d_{\min}$ имеет верхнюю границу

$$d_{\min} \leq n-k+1. \qquad (8.1.14)$$

Задаваясь линейным двоичным кодом (n, k) с минимальным расстоянием $d_{\min}$, мы можем синтезировать линейный двоичный код $(n+1, k)$ путём добавления одного дополнительного проверочного символа к каждому кодовому слову. Проверочный символ обычно выбирается так, чтобы быть проверочным символом по всем символам кодового слова. Таким образом, добавляемый проверочный символ равен 0, если исходное кодовое слово имеет чётное число единиц, и равен 1, если кодовое слово имеет нечетное число единиц. Следовательно, если минимальный вес и, следовательно, минимальное расстояние кода нечётно, добавляемый проверочный символ увеличит минимальное расстояние на 1. Мы называем код $(n+1, k)$ *расширенным кодом*. Его проверочная матрица

$$\mathbf{H}_e = \left[\begin{array}{ccc:c} & & & 0 \\ & \mathbf{H} & & 0 \\ & & & \vdots \\ & & & 0 \\ \hline 1 & 1 & \cdots\ 1 & 1 \end{array}\right], \qquad (8.1.15)$$

где $\mathbf{H}$ – проверочная матрица исходного кода. Систематический (n, k) код может быть также укорочен размещением в начале информационного блока нулевых символов. Это значит, что линейный код (n, k), состоящий из k информационных символов и $n-k$ проверочных может быть укорочен до линейного кода $(n-l, k-l)$, если установить первые l информационных символов (во всех кодовых комбинациях) нулями. Эти l символов не передаются по каналу, а $n-k$ проверочных символа рассчитываются обычным образом, как в исходном коде. Поскольку $\mathbf{C}_m = \mathbf{X}_m\mathbf{G}$, замена первых l символов $\mathbf{X}_m$ нулями эквивалентна сокращению числа строк матрицы $\mathbf{G}$ за счёт удаления первых l строк. Аналогично, вследствие того, что

$$\mathbf{C}_m\mathbf{H}^{\mathrm{T}} = \mathbf{0},$$

мы можем удалить первые l столбцов матрицы $\mathbf{H}$. Укороченный код $(n-l, k-l)$ состоит из 2^{k-l} кодовых слов. Минимальное расстояние этих 2^{k-l} кодовых слов по крайней мере не меньше, чем минимальное расстояние исходного (n, k) кода.

8.1.2. Некоторые известные линейные блоковые коды

В этом подразделе мы подробнее опишем три типа линейных блоковых кодов, которые часто встречаются на практике и характеризуются своими важными параметрами.

Коды Хемминга. Сюда относятся как двоичные, так и недвоичные коды Хемминга. Мы ограничим обсуждение свойствами двоичных кодов Хемминга. Они включают класс кодов со свойствами

$$(n, k) = (2^m - 1, 2^m - 1 - m), \tag{8.1.16}$$

где m – некоторое положительное целое число. Например, если m = 3, мы имеем код (7, 4). Проверочная матрица **H** кода Хемминга имеет особое свойство, которое позволяет нам существенно облегчить описание кода. Напомним, что проверочная матрица (n, k) кода имеет $n-k$ строк и n столбцов. Для двоичного (n, k) кода Хемминга $n = 2^m - 1$ столбцов состоят из всех возможных двоичных векторов с $n-k=m$ элементами, исключая вектор со всеми нулевыми элементами. Для примера, код (7, 4), рассмотренный в примерах 8.1.1 и 8.1.2, является кодом Хемминга. Его проверочная матрица состоит из семи вектор-столбцов (001), (010), (011), (100), (101), (110), (111).

Если необходимо генерировать систематический код Хемминга, то проверочная матрица **H** может легко приводиться к систематической форме (8.1.11). Затем соответствующую порождающую матрицу **G** можно получить из (8.1.11).

Заметим, что любые два столбца матрицы **H** не являются линейно зависимыми, так как иначе эти два столбца были бы идентичными. Однако при $m > 1$ возможно найти три столбца из **H**, которые при сложении дают нуль. Следовательно, для (n, k) кода Хемминга $d_{\min} = 3$.

Путём прибавления ко всем кодовым комбинациям (n, k) кода Хемминга одного проверочного символа получаем расширенный код Хемминга $(n+1, k)$ с $d_{\min} = 4$. С другой стороны, (n, k) код Хемминга можно укоротить до $(n-l, k-l)$ путём удаления l строк порождающей матрицы **G** и, соответственно, удаления l столбцов в проверочной матрице **H**. Распределение весов для класса кодов Хемминга (n, k) известно и выражается в компактной форме посредством следующего весового полинома:

$$A(z) = \sum_{i=0}^{n} A_i z_i = \frac{1}{n+1}\left[(1+z)^n + n(1+z)^{(n-1)/2}(1-z)^{(n+1)/2}\right], \tag{8.1.17}$$

где A_i – число кодовых слов с весом i.

Коды Адамара. Код Адамара получается путём выбора в качестве кодового слова столбцов матрицы Адамара. Матрица Адамара $\mathbf{M}_n$ – это $n \times n$ матрица (n – чётное целое) из единиц и нулей с тем свойством, что один столбец отличается от другого столбца ровно в $\frac{1}{2}n$ позициях.[1] Один столбец матрицы содержит одни нули. Остальные строки имеют $\frac{1}{2}n$ нулей и $\frac{1}{2}n$ единиц.

Для $n = 2$ матрица Адамара имеет вид

$$\mathbf{M}_2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}. \tag{8.1.18}$$

[1] Иногда элементы матрицы Адамара обозначаются +1 и –1. Тогда строки матрицы Адамара взаимно ортогональны. Также заметим, что $M = 2^k$ сигналов, полученных из адамаровских кодовых слов путём отображения каждого бита кодового слова в двоичный ФМ сигнал, взаимно ортогональны.

Затем из матрицы Адамара $\mathbf{M}_n$ можно генерировать матрицу Адамара $\mathbf{M}_{2n}$ согласно соотношению

$$\mathbf{M}_{2n} = \begin{bmatrix} \mathbf{M}_n & \mathbf{M}_n \\ \mathbf{M}_n & \overline{\mathbf{M}_n} \end{bmatrix}, \tag{8.1.19}$$

где $\overline{\mathbf{M}_n}$ означает дополнение матрицы $\mathbf{M}_n$ (нули заменятся единицами и наоборот). Так, подставив (8.1.18) в (8.1.19), получим

$$\mathbf{M}_4 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}. \tag{8.1.20}$$

Дополнение к $\mathbf{M}_4$ равно

$$\overline{\mathbf{M}_4} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}. \tag{8.1.21}$$

Теперь строки из $\mathbf{M}_4$ и $\overline{\mathbf{M}_4}$ формируют линейный двоичный код с длиной блока $n = 4$, имеющий $2n = 8$ кодовых слов. Минимальное расстояние кода $d_{\min} = \frac{1}{2}n = 2$. Путём повторного применения (8.1.19) мы можем генерировать код Адамара с длиной $n = 2^m$, $k = \log_2 2n = \log_2 2^{m+1} = m+1$ и $d_{\min} = \frac{1}{2}n = 2^{m-1}$, где m – положительное целое число. В дополнение к важному частному случаю, когда $n = 2^m$, возможны коды Адамара с другой длиной блока, но эти коды нелинейные.

Код Голея. Код Голея – это двоичный линейный код (23, 12) с $d_{\min} = 7$. Расширенный двоичный линейный код Голея (24, 12) с $d_{\min} = 8$ получается из кода Голея путём добавления ко всем кодовым комбинациям проверочного символа. Таблица (8.1.1) показывает распределение весов кодовых слов кода Голея (23, 12) и расширенного кода Голея (24, 12). Мы обсудим процедуру синтеза кода Голея в разделе 8.1.3.

Таблица 8.1.1. Распределение весов кода Голея (23, 12) и расширенного кода Голея (24, 12)

	Число кодовых слов	
Вес	код (23,12)	код (24,12)
0	1	1
7	253	0
8	506	759
11	1288	0
12	1288	2576
15	506	0
16	253	759
23	1	0
24	0	1

Источник: *Peterson* и *Weldon* (1972)

8.1.3. Циклические коды

Циклические коды – подкласс класса линейных кодов, которые удовлетворяют следующим циклическим свойствам: если $\mathbf{C}=[c_{n-1}\ c_{n-2}\ \ldots c_1\ c_0]$ – кодовое слово циклического кода, тогда $[c_{n-2}\ c_{n-3}\ \ldots c_2\ c_{n-1}]$, полученное циклическим сдвигом элементов кода $\mathbf{C}$, также являются кодовым словом. Все циклические сдвиги $\mathbf{C}$ образуют кодовые слова. Как следствие циклического свойства, эти коды обладают значительным количеством структурных удобств, которые можно использовать при реализации операций кодирования и декодирования. Большое количество алгоритмов эффективных кодеров и декодеров жёстких решений были сделаны посредством циклических кодов, что сделало возможным в практических системах связи строить блоковые коды большой длины с большим количеством кодовых слов. Описание специфических алгоритмов находится вне кругозора этой книги. Наша основная цель – вкратце описать некоторые характеристик циклических кодов.

При работе с циклическими кодами принято связывать с кодовым словом $\mathbf{C}=\left[c_{n-1}\ c_{n-2}\ \ldots c_1\ c_0\right]$ полином $C(p)$ степени $\le n-1$, определённый так

$$C(p)=c_{n-1}p^{n-1}+c_{n-2}p^{n-2}+\ldots+c_1p+c_0. \qquad (8.1.22)$$

Для двоичного кода каждый из коэффициентов полинома является или нулем, или единицей.

Теперь предположим, мы формируем полином

$$pC(p)=c_{n-1}p^{n}+c_{n-2}p^{n-1}+\ldots+c_1p^{2}+c_0p.$$

Этот полином не может представить кодовое слово, так как его степень может быть равна n (если $c_{n-1}=1$). Однако если мы разделим $pC(p)$ на p^n+1, мы получим

$$\frac{pC(p)}{p^n+1}=c_{n-1}+\frac{C_1(p)}{p^n+1}, \qquad (8.1.23)$$

где

$$C_1(p)=c_{n-2}p^{n-1}+c_{n-3}p^{n-2}+\ldots+c_0p+c_{n-1}.$$

Заметим, что полином $C_1(p)$ представляет кодовое слово $\mathbf{C}_1=\left[c_{n-2}\ c_{n-3}\ \ldots c_2\ c_{n-1}\right]$, которое как раз образовано из кодового слова $\mathbf{C}$ циклическим сдвигом на одну позицию. Поскольку $C_1(p)$ представляет собой остаток, полученный делением $pC(p)$ на p^n+1, мы говорим, что

$$C_1(p)=pC(p)\mod(p^n+1). \qquad (8.1.24)$$

Аналогичным образом, если $C(p)$ представляет кодовое слово в циклическом коде, тогда $p^iC(p)\mod(p^n+1)$ также является кодовым словом циклического кода. Так что можно написать

$$p^iC(p)=Q(p)(p^n+1)+C_i(p), \qquad (8.1.25)$$

где остаточный полином $C_i(p)$ представляет кодовое слово циклического кода, а $Q(p)$ – частное.

Мы можем генерировать циклический (n,k) код, используя *порождающий полином* $g(p)$ степени $n-k$ с двоичными коэффициентами, который является множителем при факторизации полинома p^n+1. Порождающий полином в общем виде можно записать так

$$g(p)=p^{n-k}+g_{n-k-1}p^{n-k-1}+g_1p+1. \qquad (8.1.26)$$

Мы также определяем *полином информационного сообщения* $X(p)$

$$X(p) = x_{k-1}p^{k-1} + x_{k-2}p^{k-2} + \ldots + x_1 p + x_0, \qquad (8.1.27)$$

где $[x_{k-1}\ x_{k-2}\ \ldots x_1\ x_0]$ определяет k информационных бит. Ясно, что произведение $X(p)g(p)$ – это полином степени меньшей или равной $n-1$, который может представлять кодовое слово. Заметим, что имеется 2^k полиномов $\{X_i(p)\}$ и, следовательно, имеется 2^k возможных кодовых слов, которые можно формировать при заданном $g(p)$.

Допустим, что мы обозначим эти кодовые слова так

$$C_m(p) = X_m(p)g(p), \quad m = 1, 2, \ldots, 2^k \qquad (8.1.28)$$

Чтобы показать, что кодовые слова в (8.1.28) удовлетворяют циклическому сдвигу, рассмотрим какое-либо кодовое слово $C(p)$ в (8.1.28), Циклический сдвиг $C(p)$ дает

$$C_1(p) = pC(p) + c_{n-1}(p^n + 1) \qquad (8.1.29)$$

и, поскольку и $p^n + 1$, и $C(p)$ делятся на $g(p)$ без остатка, то и $C_1(p)$ делится на $g(p)$ без остатка, т.е. $C_1(p)$ можно представить как

$$C_1(p) = X_1(p)g(p).$$

Следовательно, циклический сдвиг в кодовом слове $C(p)$, генерируемый согласно (8.1.28), порождает другое кодовое слово.

Из вышесказанного мы видим, что кодовые слова, обладающие циклическими свойствами, можно генерировать умножением 2^k сообщений на уникальный полином степени $n–k$ $g(p)$ – называемый порождающим полиномом циклического (n,k) кода, который является множителем при факторизации $p^n + 1$. Циклический код, генерируемый указанным образом, занимает подпространство S_c векторного пространства S. Размерность подпространства S_c равна k.

Пример 8.1.3. Рассмотрим код с длиной блока $n = 7$. Полином $p^7 + 1$ можно разложить на следующие сомножители

$$p^7 + 1 = (p+1)(p^3 + p^2 + 1)(p^3 + p + 1). \qquad (8.1.30)$$

Чтобы синтезировать циклический код (7,4), мы можем взять один из двух порождающих полиномов:

$$\text{или} \quad \begin{aligned} g_1(p) &= p^3 + p^2 + 1 \\ g_2(p) &= p^3 + p + 1. \end{aligned} \qquad (8.1.31)$$

Коды, генерируемые порождающими полиномами $g_1(p)$ и $g_2(p)$, эквивалентны. Кодовые слова кода (7,4), генерируемые полиномом $g_1(p) = p^3 + p^2 + 1$ даны в табл. 8.1.2.

В общем полином $p^n + 1$ можно факторизовать так:

$$p^n + 1 = g(p)h(p),$$

где $g(p)$ – означает порождающий полином для циклического (n,k) кода, $h(p)$ означает проверочный полином степени k. Последний можно использовать для генерирования дуального кода.

Таблица 8.1.2. Циклический код (7,4).
Порождающий полином $g_1(p) = p^3 + p^2 + 1$

Информационные биты				Кодовые слова						
p^3	p^2	p^1	p^0	p^6	p^5	p^4	p^3	p^2	p^1	p^0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	1	1	0	1
0	0	1	0	0	0	1	1	0	1	0
0	0	1	1	0	0	1	0	1	1	1
0	1	0	0	0	1	1	0	1	0	0
0	1	0	1	0	1	1	1	0	0	1
0	1	1	0	0	1	0	1	1	1	0
0	1	1	1	0	1	0	0	0	1	1
1	0	0	0	1	1	0	1	0	0	0
1	0	0	1	1	1	0	0	1	0	1
1	0	1	0	1	1	1	0	0	1	0
1	0	1	1	1	1	1	1	1	1	1
1	1	0	0	1	0	1	1	1	0	0
1	1	0	1	1	0	1	0	0	0	1
1	1	1	0	1	0	0	0	1	1	0
1	1	1	1	1	0	0	1	0	1	1

С этой целью можно использовать полином, *обратный* $h(p)$, определяемый так:

$$p^k h(p^{-1}) = p^k\left(p^{-k} + h_{k-1}p^{-k+1} + \ldots + h_1 p^{-1} + 1\right) = 1 + h_{k-1}p + h_{k-2}p^2 + \ldots + h_1 p^{k-1} + p^k \qquad (8.1.32)$$

Ясно, что $p^n + 1$ делится без остатка и на обратный полином. Следовательно, $p^k h(p^{-1})$ является порождающим полиномом циклического $(n, n-k)$ кода. Этот циклический код дуален коду (n,k), генерируемому порождающим полиномом $g(p)$. Таким образом, $(n, n-k)$ дуальный код образует нуль-пространство циклического (n,k) кода.

Пример 8.1.4. Рассмотрим циклический код, дуальный коду (7,4), генерированному в примере 8.1.3. Этот дуальный циклический код *(7,3)* связан с проверочным полиномом

$$h_1(p) = (p+1)(p^3 + p^2 + 1) = p^4 + p^3 + p^2 + 1. \qquad (8.1.33)$$

Обратный полином равен

$$p^4 h_1(p^{-1}) = 1 + p + p^2 + p^4.$$

Этот полином генерирует дуальный код (7, 3), данный в таблице 8.1.3. Читатель может убедиться в том, что кодовые слова дуального кода (7, 3) ортогональны кодовым словам циклического кода (7, 4) из примера 8.1.3. Заметим, что ни код (7, 4), ни код (7, 3) не являются систематическими.

Желательно показать, как можно получить порождающую матрицу из порождающего полинома циклического кода (n,k). Как указано выше, порождающую матрицу для (n,k) кода можно сконструировать из любого набора k линейно независимых кодовых слов. По заданному порождающему полиному $g(p)$ легко генерировать набор k линейно независимых кодовых слов – это кодовые слова, соответствующие набору k линейно независимых полиномов $p^{k-1}g(p)$, $p^{k-2}g(p)$, ..., $pg(p)$, $g(p)$.

Таблица 8.1.3. Дуальный код (7, 3).
Порождающий полином $p^4 h_1(p^{-1}) = 1 + p + p^2 + p^4$

Информационные биты			Кодовые слова						
p^2	p^1	p^0	p^6	p^5	p^4	p^3	p^2	p^1	p^0
0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	1	0	1	1	1
0	1	0	0	1	0	1	1	1	0
0	1	1	0	1	1	1	0	0	1
1	0	0	1	0	1	1	1	0	0
1	0	1	1	0	0	1	0	1	1
1	1	0	1	1	1	0	0	1	0
1	1	1	1	1	0	0	1	0	1

Поскольку любой из полиномов степени меньшей или равной $n-1$, который делится на $g(p)$, можно выразить как линейную комбинацию этих полиномов, набор образует базис размерностью k. Следовательно, кодовые слова, связанные с этими полиномами, формируют базис размерности k для циклического кода (n, k).

Пример 8.1.5. Четыре строки порождающей матрицы для циклического кода (7, 4) с порождающим полиномом $g_1(p) = p^3 + p^2 + 1$ можно получить из полиномов

$$p^i g_1(p) = p^{3+i} + p^{2+i} + p^i, \qquad i = 3, 2, 1, 0.$$

Легко видеть, что порождающая матрица равна

$$\mathbf{G}_1 = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 \end{bmatrix}. \tag{8.1.34}$$

Аналогично, порождающая матрица для циклического кода (7, 4), образуемая порождающим полиномом $g_2(p) = p^3 + p + 1$, равна

$$\mathbf{G}_2 = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix} \tag{8.1.35}$$

Проверочные матрицы, соответствующие $\mathbf{G}_1$ и $\mathbf{G}_2$, можно сконструировать аналогичным образом, используя соответствующие обратные полиномы (задача 8.8).

Заметим, что порождающие матрицы, полученные таким конструированием, не имеют систематическую форму. Мы можем сконструировать порождающую матрицу циклического кода в систематической форме $\mathbf{G}=[\mathbf{I}_k \vdots \mathbf{P}]$ от порождающего полинома следующим образом. Для начала заметим, что l-ая строка $\mathbf{G}$ соответствует полиному формы $p^{n-l} + R_l(p)$, $l = 1, 2, \ldots, k$, где $R_l(p)$ – полином степени, меньшей чем $n-k$. Эту форму можно получить делением p^{n-l} на $g(p)$. Таким образом, имеем

$$\frac{p^{n-1}}{g(p)} = Q_l(p) + \frac{R_l(p)}{g(p)}, \quad l = 1, 2, \ldots, k$$

или, что эквивалентно

$$p^{n-1} = Q_l(p)g(p) + R_l(p), \quad l = 1, 2, \ldots, k, \tag{8.1.36}$$

где $Q_l(p)$ – частное. Но $p^{n-l} + R_l(p)$ представляет кодовое слово в циклическом коде, поскольку $p^{n-1} + R_l(p) = Q_l(p)g(p)$. Следовательно, желательный полином, соответствующий l-й строке $\mathbf{G}$, равен $p^{n-l} + R_l(p)$.

Пример 8.1.6. Для циклического кода (7, 4) с порождающим полиномом $g_2(p) = p^3 + p + 1$, ранее рассмотренного в примере 8.1.5, имеем

$$p^6 = (p^3 + p + 1)g_2(p) + p^2 + 1,$$
$$p^5 = (p^2 + 1)g_2(p) + p^2 + p + 1,$$
$$p^4 = pg_2(p) + p^2 + p,$$
$$p^3 = g_2(p) + p + 1.$$

Следовательно, порождающая матрица кода в систематической форме

$$\mathbf{G}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix} \tag{8.1.37}$$

и соответствующая проверочная матрица

$$\mathbf{H}_2 = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}. \tag{8.1.38}$$

Оставляем в качестве упражнения для читателя показать, что порождающая матрица $\mathbf{G}_2$, даваемая (8.1.35) и матрица $\mathbf{G}_2$ в систематической форме (8.1.37) генерируют одинаковые наборы кодовых слов (задача 8.2).

Метод конструирования порождающей матрицы $\mathbf{G}$ в систематической форме согласно (8.1.36) также подразумевает, что систематический код можно генерировать непосредственно из порождающего полинома $g(p)$. Предположим, что мы умножим полином сообщения $X(p)$ на p^{n-k}. Тогда получим

$$p^{n-k}X(p) = x_{k-1}p^{n-1} + x_{k-2}p^{n-2} + \ldots + x_1 p^{n-k+1} + x_0 p^{n-k}.$$

В систематическом коде этот полином представляет первые k бит слова $C(p)$. К этому полиному мы должны прибавить полином степени меньшей, чем $n-k$, представляющей проверочные символы. Теперь, если $p^{n-k}X(p)$ разделить на $g(p)$, результат равен

$$\frac{p^{n-k}X(p)}{g(p)} = Q(p) + \frac{r(p)}{g(p)}$$

или

$$p^{n-k}X(p) = Q(p)g(p) + r(p), \tag{8.1.39}$$

где $r(p)$ имеет степень меньшую, чем $n-k$. Ясно, что $Q(p)g(p)$ – это кодовое слово циклического кода. Следовательно, суммируя (по mod 2) $r(p)$ с обеими частями (8.1.39), мы получаем желаемый систематический код.

Подытоживая, скажем, что систематический код можно генерировать так:

1. Умножаем полином сообщения $X(p)$ на p^{n-k};
2. Делим $p^{n-k}X(p)$ на $g(p)$, чтобы получить остаток $r(p)$; и
3. Добавляем $r(p)$ к $p^{n-k}X(p)$.

Ниже мы продемонстрируем, как эти вычисления можно выполнить, используя сдвиговые регистры с обратной связью.

Поскольку $p^n+1=g(p)h(p)$ или, что эквивалентно, $g(p)h(p)=0 \bmod(p^n+1)$, мы видим, что полиномы $g(p)$ и $h(p)$ *ортогональны*. Далее, полиномы $p^i g(p)$ и $p^j h(p)$ также ортогональны для всех i и j. Однако, векторы, соответствующие полиномам $g(p)$ и $h(p)$, ортогональны только, если порядок следования элементов одного из этих векторов реверсировать. То же утверждение применимо к векторам, соответствующим полиномам $p^i g(p)$ и $p^j h(p)$. Действительно, если проверочный полином $h(p)$ используется как порождающий для $(n, n-k)$ дуального кода, ансамбль кодовых слов, полученный так, включают в себя те же кодовые слова, которые генерируются обратным полиномом, за исключением того, что кодовые векторы реверсированы. Это подразумевает, что порождающая матрица для дуального кода, полученная от обратного полинома $p^k h(p^{-1})$, может быть также получить непосредственно от $h(p)$. Поскольку проверочная матрица **H** для циклического кода (n, k) является порождающей матрицей для дуального кода, следует, что **H** также можно получить от $h(p)$. Следующий пример иллюстрируют это соотношение.

Пример 8.1.7. Дуальный код для циклического кода (7, 4), порождаемого полиномом $g_1(p)=p^3+p^2+1$, – это код (7, 3), который порождается обратным полиномом $p^4 h_1(p^{-1})=p^4+p^2+p+1$. Однако, мы можем также использовать $h_1(p)$ для получения порождающей матрицы для дуального кода. Матрица, соответствующая полиному $p^i h_1(p)$, $i=2, 1, 0$, равна

$$\mathbf{G}_{h1}=\begin{bmatrix}1&1&1&0&1&0&0\\0&1&1&1&0&1&0\\0&0&1&1&1&0&1\end{bmatrix}.$$

Порождающая матрица для дуального кода (7, 3), которая является проверочной матрицей для циклического кода (7, 4), состоит из строк $\mathbf{G}_{h1}$, взятых в инверсном порядке.

Таким образом, $\mathbf{H}_1=\begin{bmatrix}0&0&1&0&1&1&1\\0&1&0&1&1&1&0\\1&0&1&1&1&0&0\end{bmatrix}$.

Читатель может убедиться, что $\mathbf{G}_1\mathbf{H}_1^{\mathrm{T}}=0$.

Заметим, что векторы $\mathbf{H}_1$ состоит из всех семи двоичных вектор-столбцов длины 3, исключая вектор из одних нулей. Но это как раз описание проверочной матрицы для кода Хемминга (7, 4). Следовательно, циклический код (7, 4) эквивалентен коду Хемминга (7, 4), рассмотренному ранее в примерах 8.1.1 и 8.1.2.

Кодеры для циклических кодов. Операции кодирования при создании циклических кодов можно выполнить при помощи линейных сдвигающих регистров с обратной связью, с использованием порождающего или проверочного полинома. Сначала рассмотрим использование $g(p)$.

Как указано выше, генерирование систематического циклического кода включает три ступени, а именно: умножение полинома сообщения $X(p)$ на p^{n-k}, деление этого произведения на $g(p)$ и в заключение, прибавление остатка $r(p)$ к $p^{n-k}X(p)$. Из этих трёх ступеней только деление является нетривиальным.

Деление полинома $A(p) = p^{n-k}X(p)$ степени $n-1$ на полином

$$g(p) = g_{n-k}p^{n-k} + g_{n-k-1}p^{n-k-1} + \ldots + g_1 p + g_0$$

можно выполнить посредствам $(n-k)$ ячеек регистра сдвига с обратной связью, показанного на рис. 8.1.2.

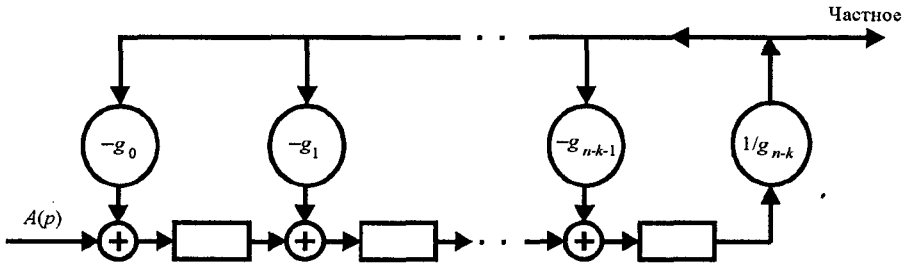


Рис. 8.1.2. Регистр сдвига с обратной связью для деления полинома $A(p)$ на $g(p)$

Первоначально ячейки сдвига регистра содержит одни нули. Коэффициенты $A(p)$ поступают и продвигаются по регистру сдвига по одному коэффициенту за такт, начиная с коэффициентов более высокого порядка, т.е. с a_{n-1}, затем a_{n-2} и так далее. После $(n-k-1)$-го сдвига, первый ненулевой выход частного равен $q_1 = (g_{n-k})^{-1}a_{n-1}$. Последующие выходы генерируются так, как показано на рис. 8.1.2. Для образования каждого выходного коэффициента линии мы должны вычесть полином $g(p)$, умноженный на этот коэффициент, как при обычном «длинном» делении. Это вычитание производится посредством обратной связи. Таким образом, регистр сдвига на рис. 8.1.2 обеспечивает деление двух полиномов.

В нашем случае $g_{n-k} = g_0 = 1$, и для двоичных кодов арифметические операции выполняются по mod 2. Следовательно, операция вычитания сводится к сложению по mod 2. Далее мы будем только интересоваться генерированием проверочных символов для каждого кодового слова, поскольку код систематический. Как следствие, кодер циклического кода принимает вид, показанный на рис. 8.1.3.

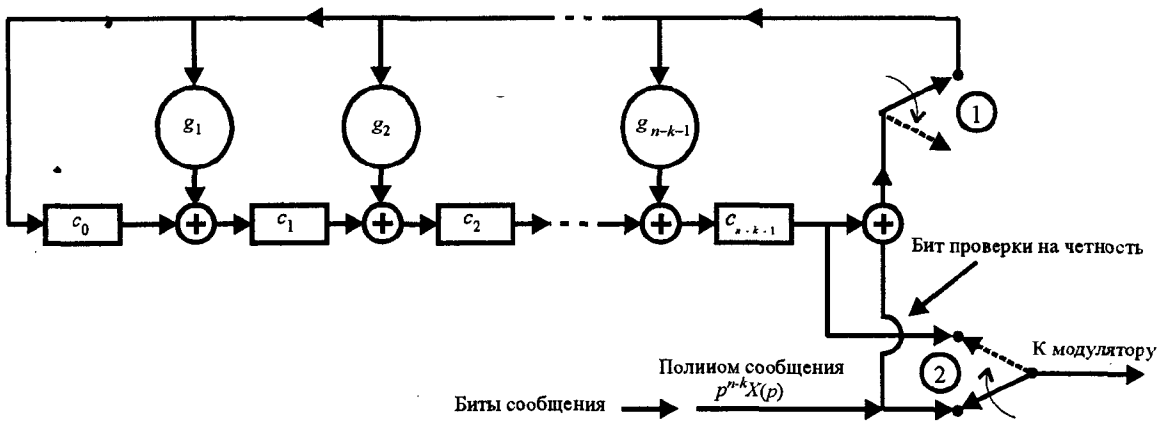


Рис. 8.1.3. Циклический кодер с использованием порождающего полинома $g(p)$

Первые k бит на выходе кодера просто равны k информационным битам. Эти k бит одновременно поступают на регистр сдвига, поскольку ключ 1 замкнут. Заметим, что умножение полиномов p^{n-k} и $X(p)$ явно не производится. После того как все k информационных бита попали на вход кодера (и к модулятору), положения двух ключей на рис. 8.1.3 меняются на обратные. Начиная с этого времени, содержимое регистра сдвига просто даёт $n-k$ проверочных символов, которые соответствуют коэффициентам полинома остатка. Эти $n-k$ бита последовательно отправляются на модулятор.

Пример 8.1.8. Регистр сдвига для кодирования циклическим кодом (7, 4) с порождающим полиномом $g(p)=p^3+p+1$ показан на рис. 8.1.4.

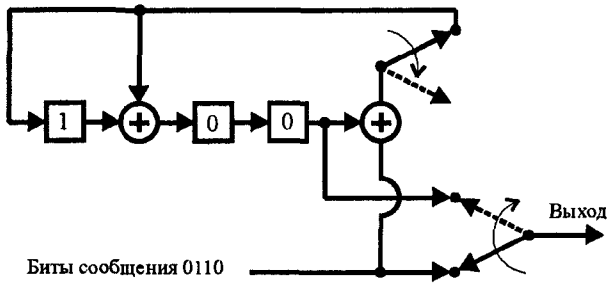


Рис. 8.1.4. Циклический кодер (7,4) с использованием порождающего полинома $g(p)=p^3+p+1$

Предположим, что сообщением является цепочка 0110. Содержание сдвигового регистра дано таблицей:

Вход	Шаг сдвига	Содержимое регистра
	0	0 0 0
0	1	0 0 0
1	2	1 1 0
1	3	1 0 1
0	4	1 0 0

Таким образом, три проверочных символа: 100. Они соответствуют кодовым символам $c_5 = 0$, $c_6 = 0$ и $c_7 = 1$.

Вместо использования порождающего полинома $g(p)$, мы можем выполнить кодер циклического кода, используя проверочный полином

$$h(p) = p^k + h_{k-1}p^{k-1} + \ldots + h_1 p + 1.$$

Соответствующий кодер показан на рис. 8.1.5. Первоначально k информационных символа (бита) передвигаются по регистру сдвига и одновременно поступает на модулятор. После того, как все k информационных символов пройдут по регистру, ключ переводится в положение 2, и сдвиговой регистр работает ещё $n-k$ такта, генерируя $n-k$ проверочных символов, как показано на рис. 8.1.5.

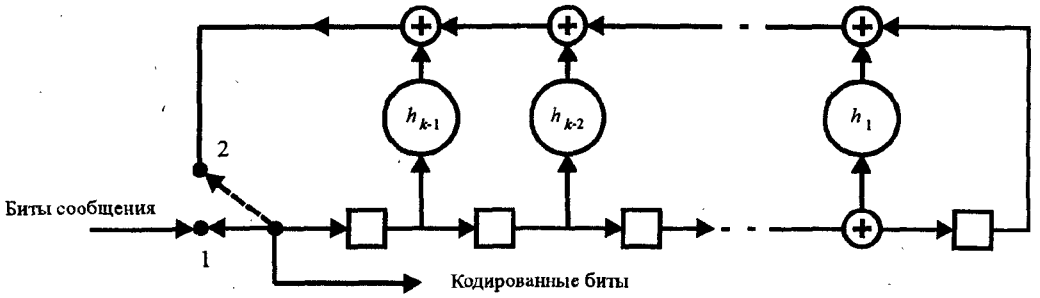


Рис. 8.1.5. Циклический кодер (n,k) с использованием проверочного полинома $h(p)$

Пример 8.1.9. Проверочный полином для циклического кода (7,4), генерируемый порождающим полиномом $g(p) = p^3 + p + 1$, равен $h(p) = p^4 + p^2 + p + 1$. Кодер для этого кода, основанный на проверочном полиноме, иллюстрируется на рис. 8.1.6. Если на входе кодера действует сообщение 0110, то проверочными символами являются $c_5 = 0$, $c_6 = 0$, $c_7 = 1$, как легко проверить.

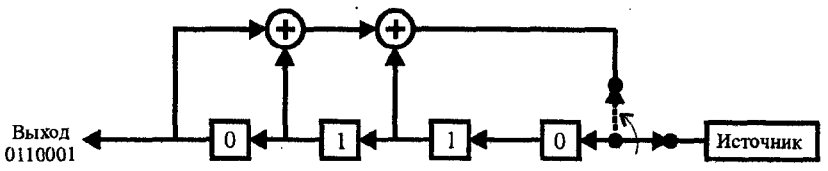


Рис. 8.1.6. Циклический кодер (7,4) с использованием проверочного полинома $h(p)=p^4+p^2+1$

Следует заметить, что кодер, базирующийся на порождающем полиноме, проще, когда $n-k < k$ $(k > \frac{1}{2}n)$, т.е. для больших скоростей кода $(R_c > \frac{1}{2})$. В то же время кодер, базирующийся на проверочном полиноме, проще, когда $k < n-k$ $(k < \frac{1}{2}n)$, что соответствует низкоскоростным кодам $(R_c < \frac{1}{2})$.

Циклические коды Хемминга. Класс циклических кодов включает коды Хемминга, которые имеют длину блока $n = 2^m - 1$ и $n-k = m$ проверочных символов, где m – любое положительное целое число. Циклические коды Хемминга эквивалентны кодам Хемминга, описанным в разделе 8.1.2.

Циклический код Голея (23, 12). Линейный код Голея (23, 12), описанный в разделе (8.1.2), можно генерировать как циклический код посредством порождающего полинома

$$g(p) = p^{11} + p^9 + p^7 + p^6 + p^5 + p + 1. \qquad (8.1.40)$$

Кодовые слова имеют минимальное расстояние $d_{\min} = 7$.

Коды сдвигового регистра максимальной длины. Коды сдвигового регистра максимальной длины – это класс циклических кодов

$$(n, k) = (2^m - 1, m), \qquad (8.1.41)$$

где m – положительное целое. Кодовые слова обычно генерируются посредством m-каскадного цифрового регистра сдвига с обратной связью, основанного на проверочном полиноме. Для каждого передаваемого кодового слова m информационных бит вводятся в регистр сдвига, а ключ перемещается с позиции 1 на позицию 2. Содержание регистра сдвигается влево в общей сложности $2^m - 1$ раз. Такая операция генерирует систематический код с требуемой длиной $n = 2^m - 1$. Для примера, кодовые слова, генерируемые 3-каскадным регистром сдвига по рис. 8.1.7, сведены в таблицу 8.1.4.

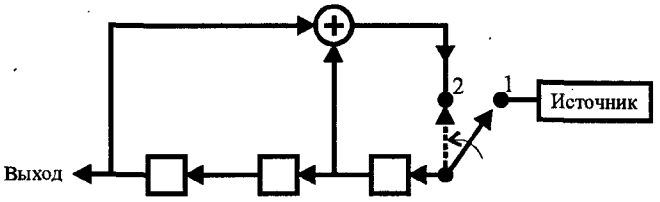


Рис. 8.1.7. Трёхкаскадный регистр сдвига (m=3) с обратной связью

Таблица 8.1.4. Код сдвигового регистра максимальной длины для m=3

Информационные биты			Кодовые слова						
0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	1	1	1	0	1
0	1	0	0	1	0	0	1	1	1
0	1	1	0	1	1	1	0	1	0
1	0	0	1	0	0	1	1	1	0
1	0	1	1	0	1	0	0	1	1
1	1	0	1	1	0	1	0	0	1
1	1	1	1	1	1	0	1	0	0

Заметим, что, за исключением кодовых слов из одних нулей, все кодовые слова, генерируемые регистром сдвига, представляют собой циклические сдвиги единственного кодового слова. Это легко увидеть из диаграммы состояний регистра сдвига, которая иллюстрируется на рис. 8.1.8 для m=3. Когда регистр сдвига после первоначального заполнения сдвигается $2^m - 1$ раз, он проходит через все возможные $2^m - 1$ состояния. Таким образом, регистр сдвига возвращается обратно к начальному состоянию за $2^m - 1$ шагов сдвига.

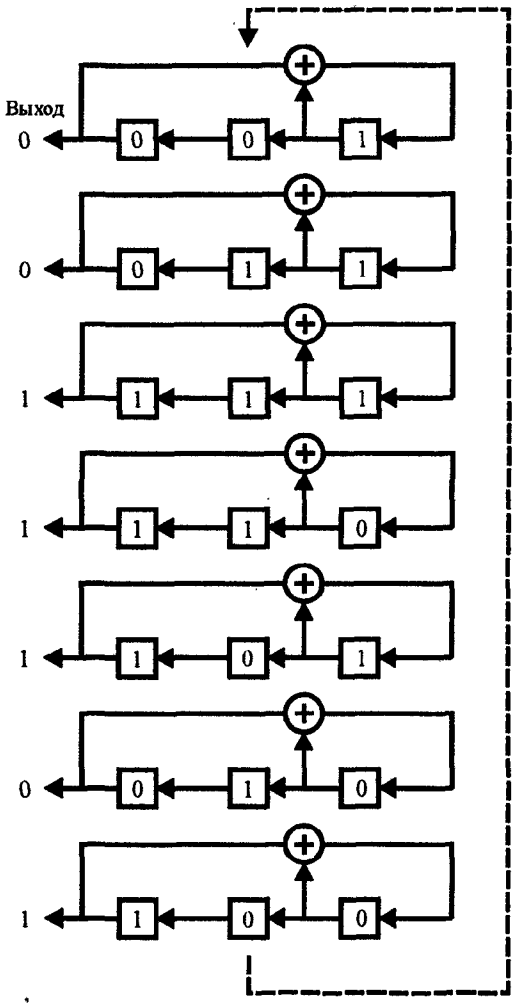


Рис. 8.1.8. Семь шагов последовательности максимальной длины, генерируемой регистром сдвига с m=3

Следовательно, выходная последовательность периодическая, и её период равен $n = 2^m - 1$. Поскольку имеется $2^m - 1$ возможных состояний, такая длина соответствует максимально возможному периоду. Это и объясняет, что $2^m - 1$ кодовых слов являются различными циклическими сдвигами единственного кодового слова. Коды регистра сдвига максимальной длины существуют для любой положительной величины m. Таблица 8.1.5. показывает номера ячеек, которые объединяются в сумматоре по mod 2, в регистре сдвига максимальной длины для $2 \le m \le 34$.

Другая особенность кодовых слов в регистре сдвига максимальной длины заключается в том, что каждое кодовое слово, за исключением слова из одних нулей, содержит 2^{m-1} единиц и 2^{m-1} нулей. Поскольку код линейный, его вес является также минимальным расстоянием кода, т.е.

$$d_{\min} = 2^{m-1}.$$

В заключение заметим, что код регистра сдвига максимальной длины (7,3), показанный в табл. 8.1.4, идентичен коду (7,3), данному в табл. 8.1.3 и является дуальным коду Хемминга (7,4), данному в табл. 8.1.2. Это не совпадение. Коды регистра сдвига максимальной длины являются дуальными кодами для циклических кодов Хемминга $\left(2^m - 1,\ 2^m - 1 - m\right)$.

Таблица 8.1.5. Соединения в сдвиговом регистре для генерирования последовательности максимальной длины

m	№№ отводов к сумматору по модулю 2	m	№№ отводов к сумматору по модулю 2	m	№№ отводов к сумматору по модулю 2	m	№№ отводов к сумматору по модулю 2
2	1, 2	11	1, 10	20	1, 18	29	1, 28
3	1, 3	12	1, 7, 9, 10	21	1, 20	30	1, 8, 29, 30
4	1, 4	13	1, 10, 11, 13	22	1, 22	31	1, 29
5	1, 4	14	1, 5, 9, 14	23	1, 19	32	1, 11, 31, 32
6	1, 6	15	1, 15	24	1, 18, 23, 24	33	1, 21
7	1, 7	16	1, 5, 14, 16	25	1, 23	34	1, 8, 33, 34
8	1, 5, 6, 7	17	1, 15	26	1, 21, 25, 26		
9	1, 6	18	1, 12	27	1, 23, 26, 27		
10	1, 8	19	1, 15, 18, 19	28	1, 26		

Источник: *Forney* (1970)

Регистр сдвига для генерирования кода максимальной длины можно также использовать для генерирования периодической двоичной последовательности с периодом $n = 2^m - 1$. Двоичная периодическая последовательность имеет периодическую автокорреляционную функцию $\phi(m)$ со значениями $\phi(m) = n$ для $m = 0, \pm n, \pm 2n, \ldots$ и $\phi(m) = -1$ для других сдвигов, как будет описано в разделе 8.2.4. Эта импульсно-подобная автокорреляционная функция подразумевает, что спектр мощности близок к равномерному и, следовательно, m-последовательность проявляет свойства белого шума. Поэтому последовательности максимальной длины называют псевдошумовыми (ПШ) последовательностями и они находят применение для скремблирования данных и для генерации широкополосных сигналов с рассеянным спектром.

Коды Боуза-Чоудхури-Хоквингема (БЧХ). БЧХ коды представляют большой класс циклических кодов как с двоичным, так и с недвоичным алфавитами.

Двоичные БЧХ коды можно построить с параметрами

$$n = 2^m - 1$$
$$n - k \le mt$$
$$d_{\min} = 2t + 1,$$

где m ($m \ge 3$) и t – произвольные положительные целые числа. Этот класс двоичных кодов предоставляет разработчику систем связи большой выбор длин блока и скоростей кода. Недвоичные БЧХ коды включают в себя мощные коды Рида–Соломона, которые будут описаны ниже.

Порождающие полиномы для БЧХ кодов можно конструировать из множителей полинома $p^{2^m-1} + 1$.

В таблице 8.1.6. приведены коэффициенты порождающих полиномов для БЧХ кодов длины $7 \le n \le 255$, соответствующие $3 \le m \le 8$. Коэффициенты даны в восьмеричной форме, причём самая левая цифра соответствует слагаемому полинома с наивысшей степенью.

Таблица 8.1.6. Коэффициенты порождающих полиномов (в восьмеричной форме) для кодов БЧХ длиной $7 \le n \le 255$

n	k	t	$g(p)$
7	4	1	13
15	11	1	23
	7	2	721
	5	3	2467
31	26	1	45
	21	2	3551
	16	3	107657
	11	5	5423325
	6	7	313365047
63	57	1	103
	51	2	12471
	45	3	1701317
	39	4	166623567
	36	5	1033500423
	30	6	157464165547
	24	7	17323260404441
	18	10	1363026512351725
	16	11	6331141367235453
	10	13	472622305527250155
	7	15	5231045543503271737
127	120	1	211
	113	2	41567
	106	3	11554743
	99	4	3447023271
	92	5	624730022327
	85	6	130704476322273
	78	7	26230002166130115
	71	9	6255010713253127753
	64	10	1206534025570773100045
	57	11	335265252505705053517721
	50	13	544465125233140124215014 21
	43	14	177217722136512275212205743 43
	36	15	314607466652207504476457472 1735
	29	21	403114461367670603667530141176155
	22	23	123376070404722522435445626637647043
	15	27	220570424456045547705230137622176043 53
	8	31	704726405275103065147622427156773313302 17
255	247	1	435
	239	2	267543
	231	3	156720665
	223	4	75626641375
	215	5	23157564726421
	207	6	16176560567636227
	199	7	7633031270420722341
	191	8	2663470176115333714567

Таблица 8.1.6. (продолжение)

n	k	t	$g(p)$
255	187	9	52755313540001322236351
	179	10	22634710717340432416300455
	171	11	154162142123423560770616 3067
	163	12	75004155100756025515747245 14601
	155	13	37575130054076650157225064646 77633
	147	14	16421301735371655253041653054410 11711
	139	15	46140173206017556157072273024745 3567445
	131	18	21571333147151015126125027744214202416 5471
	123	19	12061405224206600371721032651614122627250 6267
	115	21	60526665572100247263636404600276352556313 472737
	107	22	22205772322066256312417300235347420176574750154441
	99	23	10656667253473174222741416201574332252411076432303431
	91	25	67502650303274441727236317247325110755507627207243 44561
	87	26	11013676341474323643523163430717204620672254527331 1721317
	79	27	66700035637657500020270344207366174621015326711766 541342355
	71	29	24024710520644321515554172112331163205444250362557 643221706035
	63	30	10754475055163544325315217357707003666111726455267 613656702543301
	55	31	73154252035011001330152753060320543254143267550105 57044426035473617
	47	42	25335420170626465630330413774062331751233341454460 4500506602455254317 3
	45	43	15202056055234161131101346376423701563670024470762 37303320215702505154 1
	37	45	51363302550670074141774472454375304207357061743234 323476443547374030440 03
	29	47	30257155366730714655270640123613771153422423242011 74114060254757410403565037
	21	55	12562152570603326560017731536076121032273414056530 74542521153121614466513473725
	13	59	46417320050525645444265737142500660043306774454765 61403174677213570261344605005 47
	9	63	15726025217472463201031043255355134614162367212044 07454511276611554770556167751 6057

Источник: *Stenbit* (1964), © 1964 *IEEE*

Так, коэффициентами полинома для кода (15,5) является 2467, что соответствует двоичной форме 10 100 110 111. Следовательно, порождающий полином равен $g(p) = p^{10} + p^8 + p^5 + p^4 + p^2 + p + 1$.

Более общий список порождающих полиномов для БЧХ кодов дан Питерсоном и Уэлдоном (1972), которые дали таблицы множителей полиномов $p^{2^m-1} + 1$ для $m \leq 34$.

8.1.4. Оптимальное декодирование мягких решений для линейных блоковых кодов

В этом подразделе мы рассмотрим качество линейных блоковых кодов в канале с АБГШ, когда на приёме используется оптимальное (без квантования) декодирование мягких решений. Символы кодового слова могут быть переданы посредством произвольных двоичных сигналов одним из методов, описанных в главе 5. Для наших целей мы рассмотрим двоичную (или четверичную) когерентную ФМ, что является наиболее эффективным методом, и двоичную ортогональную ЧМ с когерентным или некогерентным детектированием.

Пусть $\mathscr{E}$ означает энергию переданного сигнала на кодовое слово и пусть $\mathscr{E}_c$ означает энергию сигнала, требуемую для передачи отдельного элемента (бита) кодового слова. Поскольку в кодовом слове n бит, то $\mathscr{E}=n\mathscr{E}_c$. Так как кодовые слова содержат k бит информации, то энергия на информационный бит

$$\mathscr{E}_b = \frac{\mathscr{E}}{k} = \frac{n}{k}\mathscr{E}_c = \frac{\mathscr{E}_c}{R_c}. \tag{8.1.43}$$

Считается, что все кодовые слова равновероятны с априорной вероятностью $1/M$.

Допустим, что символы кодового слова передаются посредством двоичной ФМ. Каждое кодовое слово отображается одним из M сигналов. Из главы 5 мы знаем, что приёмник, оптимальный по критерию минимума средней вероятности ошибки на кодовое слово, можно реализовать в канале с АБГШ в виде параллельного банка M фильтров, согласованных с M возможными сигналами. Выходы M согласованных фильтров в конце каждого сигнального интервала, который определяется передачей n символов кодового слова, сравниваются и выбирается кодовое слово, которому соответствует максимальный выход согласованного фильтра. Альтернативно можно использовать M взаимных корреляторов. В любом случае, реализацию приёмника можно упростить. Это значит, что эквивалентный оптимальный приёмник можно реализовать, используя единственный фильтр (или коррелятор), согласованный с двоичным ФМ сигналом, использованным для передачи каждого бита кодового слова, а за ним декодер, который формирует M величин для решения, соответствующих M кодовым словам.

Для конкретности, пусть r_j, $j = 1, 2, \ldots, n$, представляют n последовательных отсчётов выхода единственного согласованного фильтра для конкретного кодового слова. Поскольку используется сигнал двоичной ФМ, выход r_j можно выразить или так:

$$r_j = \sqrt{\mathscr{E}_c} + n_j, \tag{8.1.44}$$

когда j-й разряд кодового слова содержит «1», или так:

$$r_j = -\sqrt{\mathscr{E}_c} + n_j, \tag{8.1.45}$$

когда j-й разряд содержит «0». Величины $\{n_j\}$ представляют АБГШ в отсчётных точках. Каждое n_j имеет нулевое среднее и дисперсию $\frac{1}{2}N_0$. Зная M возможных к передаче кодовых слов и принятые значения $\{r_j\}$, оптимальный декодер формирует M корреляционных метрик

$$CM_i = C(\mathbf{r}, \mathbf{C}_i) = \sum_{j=1}^{n}(2c_{ij} - 1)r_j, \quad i = 1, 2, \ldots, M, \tag{8.1.46}$$

где c_{ij} означает символ на j-ой позиции i-го кодового слова. Так, если $c_{ij} = 1$, взвешивающий множитель $2c_{ij} - 1 = 1$, а если $c_{ij} = 0$, взвешивающий множитель

$2c_{ij}-1=\pm 1$. Такое взвешивание приводит к тому, что корреляционная метрика, соответствующая действительно переданному кодовому слову, будет иметь среднюю величину $\sqrt{\mathscr{E}_c}n$, в то время как другие $M-1$ метрик будет иметь меньшее значение.

Хотя вычисления, требуемые для формирования корреляционных метрик для мягкого декодирования согласно (8.1.46), относительно простые, всё же затруднительно вычислять (8.1.46) для всех возможных кодовых слов, когда число кодовых слов велико, например $M>2^{10}$. В этих случаях всё же возможно реализовать декодирование мягких решений, используя алгоритмы, которые применяют технику для отбрасывания неправдоподобных кодовых слов без вычисления всего набора их корреляционных метрик, определяемых (8.1.46). Несколько типов такого декодирования мягких решений были описаны в литературе. Интересующихся этим читателей отошлём к статьям Форни (1966), Уэлдона (1971), Чейза (1972), Вайнберга и Вольфа (1973), Вольфа (1978) и Матиса и Модестино (1982).

Для определения вероятности ошибки блокового кода заметим, что, когда такой код применяется в двоичном симметричном канале, каким является канал с АБГШ, и когда осуществляется оптимальное декодирование мягких решений, то вероятность ошибки при передаче m-го кодового слова одинакова для всех m. Поэтому для простоты предположим, что передаётся кодовое слово $\mathbf{C}_1$, состоящее из одних нулей. Для правильного декодирования $\mathbf{C}_1$ корреляционная метрика CM_1 должна превышать все остальные $M-1$ корреляционные метрики CM_m, $m=2, 3, \ldots, M$. Все метрики распределены по Гауссу. Среднее значение CM_1 равно $\sqrt{\mathscr{E}_c}n$, в то время как средние значения для CM_m, $m=2, 3, \ldots, M$, равны $\sqrt{\mathscr{E}_c}n\left(1-2w_m/n\right)$.

Дисперсия для каждой величины, участвующей в решении, равна $\frac{1}{2}N_0$. Нахождение точного выражения для вероятности правильного декодирования или, что эквивалентно, нахождение вероятности ошибки в кодовом слове усложняется наличием корреляции между M корреляционными метриками. Коэффициенты взаимной корреляции между $\mathbf{C}_1$ и другими $M-1$ кодовыми словами равны

$$\rho_m=\left(1-2w_m/n\right),\ m=2, 3, \ldots, M, \tag{8.1.47}$$

где w_m означает вес m-го кодового слова .

Вместо того, чтобы пытаться получить точную формулу для вероятности ошибки, мы обратимся к объединённой верхней границе. Вероятность того, что $CM_m>CM_1$, равна

$$P_2(m)=Q\left(\sqrt{\frac{\mathscr{E}}{N_0}\left(1-\rho_m\right)}\right), \tag{8.1.48}$$

где $\mathscr{E}=k\mathscr{E}_b$ - энергия сигнала кодового слова. Подставив для ρ_m значение из (8.1.47), а для $\mathscr{E}$ значение из (8.1.43.), получаем

$$P_2(m)=Q\left(\sqrt{\frac{2\mathscr{E}_b}{N_0}R_c w_m}\right)=Q\left(\sqrt{2\gamma_b R_c w_m}\right), \tag{8.1.49}$$

где γ_b – это ОСШ на бит, а R_c – скорость кода.

Средняя вероятность ошибки кодового слова ограничена сверху суммой вероятностей ошибок двоичных событий, определяемых (8.1.49.), т.е.

$$P_M\le\sum_{m=2}^{M}P_2(m)\le\sum_{m=2}^{M}Q\left(\sqrt{2\gamma_b R_c w_m}\right). \tag{8.1.50}$$

Вычисление вероятности ошибки декодирования мягких решений согласно (8.1.50)

требует знания распределения весов кода. Распределения весов для многих кодов даны во многих изданиях по теории кодирования, например, Берлекэмпа (1968) и МакВильямса и Слоэна (1977).

Некоторую свободную (неплотную) границу можно получить, заметив, что

$$Q\left(\sqrt{2\gamma_b R_c w_m}\right) \le Q\left(\sqrt{2\gamma_b R_c d_{\min}}\right) < \exp\left(-\gamma_b R_c d_{\min}\right). \qquad (8.1.51)$$

Следовательно,

$$P_M \le (M-1)Q\left(\sqrt{2\gamma_b R_c d_{\min}}\right) < \exp\left(-\gamma_b R_c d_{\min} + k\ln 2\right). \qquad (8.1.52)$$

Эта граница особенно полезна, поскольку она не требует знания распределения весов кода. Если верхнюю границу в (8.1.52.) сравнить с характеристикой качества двоичных систем ФМ без кодирования, которые ограничены сверху как $\frac{1}{2}\exp\left(-\gamma_b\right)$, находим, что кодирование даёт выигрыш примерно на $10\lg\left(R_c d_{\min} - k\ln 2/\gamma_b\right)$ дБ .Мы можем это назвать *выигрыш от кодирования.* Заметим, что величина выигрыша зависит от параметров кода, а также от ОСШ на бит γ_b.

Выражение для вероятности ошибки ансамбля сигналов с одинаковой взаимной корреляцией любой пары, что выполняется для симплексной системы сигналов, описанной в разделе 5.2, позволяет получить третью аппроксимацию для вероятности ошибки передаваемых сигналов. Мы знаем, что максимальное значение коэффициента взаимной корреляции между парой кодируемых сигналов равно

$$\rho_{\max} = 1 - \frac{2}{n} d_{\min}. \qquad (8.1.53)$$

Если предположить (в наихудшем случае), что все M кодовых слов имеют коэффициент взаимной корреляции, равный $\rho_{\max}$, то вероятность ошибки кодового слова можно легко вычислить. Поскольку некоторые кодовые слова разделены больше, чем на минимальное расстояние, вероятность ошибки, вычисляемая при $\rho_r = \rho_{\max}$, имеет фактически верхнюю границу. Таким образом,

$$P_M \le 1 - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{\upsilon^2}{2}}\left(\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\upsilon+\sqrt{4\gamma_b R_c d_{\min}}} e^{-\frac{x^2}{2}} dx\right)^{M-1} d\upsilon. \qquad (8.1.54)$$

Границы качества для линейных блоковых кодов, данные выше, даны для вероятности ошибки блока или вероятности ошибки кодового слова.

Получить оценку эквивалентной вероятности ошибки на бит P_b существенно более сложно. В общем, если произошла ошибка в блоке, некоторые из k информационных бит в блоке будут приняты без ошибки, а некоторые останутся с ошибкой. Для ортогональных сигналов множитель конверсии, на который надо умножить P_M, чтобы получить P_b, равен $2^{k-1}/\left(2^k - 1\right)$. Этот множитель равен единице для $k = 1$ и приближается к $\frac{1}{2}$ по мере увеличения k, что эквивалентно предположению, что, в среднем, половина от k информационных бит будут с ошибкой, когда произошла ошибка в блоке. Множитель конверсии для кодированных сигналов зависит сложным образом от дистанционных характеристик кода, но он, конечно, не хуже, чем в предположении, что, в среднем, половина из k информационных бита будут с ошибкой, если произошла ошибка в блоке. Следовательно, $P_b \le \frac{1}{2} P_M$.

Границы качества, даваемые (8.1.50), (8.1.52) и (8.1.54), также годятся для случая, когда пара двоичных символов кодового слова передаётся четверичной ФМ, поскольку четверичную ФМ можно рассматривать как эквивалент двух независимых двоичных сигналов ФМ, переданных в квадратуре. Более того, границы (8.1.52) и (8.1.54), которые

зависят только от минимального расстояния кода, приемлемы также к нелинейным двоичным кодам.

Если для передачи каждого символа кодового слова через канал с АБГШ используется двоичная ортогональная ЧМ, оптимальный приёмник можно реализовать посредством двух согласованных фильтров, один согласованный с частотой, соответствующей передаче 0, а другой согласованный с частотой, соответствующей передаче 1. За ними следует декодер, который формирует M корреляционных метрик, соответствующих M возможным кодовым словам. В любом случае, пусть r_{0j} и r_{1j} – отсчёты на входе устройство сложения. Корреляционные метрики, сформированные декодером, можно выразить так

$$CM_i = \sum_{j=1}^{n}\left[c_{ij}r_{1j} + \left(1-c_{ij}\right)r_{0j}\right], \quad i = 1, 2, \ldots, M, \tag{8.1.55}$$

где c_{ij} представляет j-й символ в i-м кодовом слове. Кодовое слово, соответствующее наибольшому из $\{CM_i\}$, выбирается в качестве переданного кодового слова.

Если осуществляется когерентное детектирование сигналов двоичной ЧМ, случайные величины $\{r_{0j}\}$ и $\{r_{1j}\}$ являются гауссовскими и, следовательно, корреляционные метрики $\{CM_i\}$ также гауссовские. В этом случае границы для качества кода легко найти. Для конкретности, предположим, что передается кодовое слово $\mathbf{C}_1$, состоящее из одних нулей.

Тогда

$$\left.\begin{aligned} r_{0j} &= \sqrt{\mathscr{E}_c} + n_{0j} \\ r_{1j} &= n_{1j} \end{aligned}\right\} \quad j = 1, 2, \ldots, n, \tag{8.1.56}$$

где $\{n_{ij}\}$, $i = 0,1$; $j = 1, 2, \ldots, n$-взаимно статистически независимые гауссовские случайные величины с нулевым средним и дисперсией $\frac{1}{2}N_0$. Следовательно, CM_1 является гауссовской величиной со средним $\sqrt{\mathscr{E}_c}n$ и дисперсией $\frac{1}{2}N_0$. С другой стороны, корреляционная метрика CM_m, соответствующая кодовому слову с весом w_m, является гауссовской случайной величиной со средним $\sqrt{\mathscr{E}_c}n\left(1-w_m/n\right)$ и дисперсией $\frac{1}{2}nN_0$. Поскольку величины $\{CM_m\}$ коррелированны, мы снова обращаемся к объединенной границе. Коэффициенты корреляции определяются так:

$$\rho_m = 1 - w_m/n. \tag{8.1.57}$$

Следовательно, вероятность того, что $CM_m > CM_1$, равна

$$P_2(m) = Q\left(\sqrt{\gamma_b R_c w_m}\right). \tag{8.1.58}$$

Сравнение этого результата с тем, который дается (8.1.49) для когерентной ФМ показывает, что когерентная ФМ требует на 3 дБ меньше ОСШ для достижения того же качества. Это не удивительно с учетом того факта, что некодированная двоичная ФМ на 3 дБ лучше, чем двоичная ортогональная ЧМ при когерентном детектировании. Таким образом, преимущество ФМ относительно ЧМ сохраняется для кодированных сигналов. Затем мы заключаем, что границы, данные (8.1.50), (8.1.52), (8.1.54), приемлемы для кодированных сигналов, передаваемых посредством двоичной ортогональной когерентной ЧМ, если γ_b заменить на $\frac{1}{2}\gamma_b$.

Если в приёмнике используется квадратичное детектирование двоичных ортогональных ЧМ сигналов, качество дополнительно ухудшается за счёт потерь при некогерентном сложении, как будет указано в гл. 12. Предположим снова, что передаётся кодовое слово из одних нулей. Тогда корреляционные метрики определяются (8.1.55), где величины на входе декодера теперь равны

$$\left.\begin{aligned} r_{0j} &= \left|\sqrt{\mathscr{E}_c} + N_{0j}\right|^2 \\ r_{1j} &= \left|N_{1j}\right|^2 \end{aligned}\right\} \quad j = 1, 2, \ldots, n, \tag{8.1.59}$$

где $\{N_{0j}\}$ и $\{N_{1j}\}$ представляют комплексные статистически взаимно независимые гауссовские случайные величины с нулевыми средними и дисперсией N_0. Корреляционная метрика CM_1 равна

$$CM_1 = \sum_{j=1}^{n} r_{0j}, \tag{8.1.60}$$

в то время как корреляционные метрики, соответствующие кодовым словам с весом w_m, статистически эквивалентны корреляционным метрикам кодовых слов, в которых $c_{mj} = 1$ для $1 \le j \le w_m$ и $c_{mj} = 0$ для $w_m + 1 \le j \le n$. Таким образом, CM_m можно выразить так

$$CM_m = \sum_{j=1}^{w_m} r_{1j} + \sum_{j=w_m+1}^{n} r_{0j}. \tag{8.1.61}$$

Разность между CM_1 и CM_m равна

$$CM_1 - CM_m = \sum_{j=1}^{w_m} \left(r_{0j} - r_{1j}\right), \tag{8.1.62}$$

а вероятность ошибки равна вероятности того, что $CM_1 - CM_m < 0$. Но эта разность является частным случаем общей квадратичной формы комплексных гауссовских случайных величин, рассматриваемых в гл. 12 и в приложении В. Выражение для вероятности ошибки при различении CM_1 и одного из сигналов CM_m имеет вид (см. раздел 12.1.1):

$$P_2(m) = \frac{1}{2^{2w_m - 1}} \exp\left(-\tfrac{1}{2}\gamma_b R_c w_m\right) \sum_{i=0}^{w_m - 1} K_i \left(\tfrac{1}{2}\gamma_b R_c w_m\right)^i, \tag{8.1.63}$$

где, по определению,

$$K_i = \frac{1}{i!} \sum_{r=0}^{w_m - 1 - i} \binom{2w_m - 1}{r}. \tag{8.1.64}$$

Объединённая граница, получаемая путём суммирования $P_2(m)$ по $2 \le m \le M$, даёт нам верхнюю границу для вероятности ошибочного декодирования кодового слова.

Как альтернативу, мы можем использовать минимальное расстояние вместо распределения весов, и получить менее плотную верхнюю границу

$$P_M \le \frac{M-1}{2^{2d_{\min} - 1}} \exp\left(-\tfrac{1}{2}\gamma_b R_c w_m\right) \sum_{i=0}^{d_{\min} - 1} K_i \left(\tfrac{1}{2}\gamma_b R_c w_m\right)^i. \tag{8.1.65}$$

Меру потерь из-за некогерентного сложения, при квадратичном детектировании и сложении n элементарных двоичных сигналов в кодовом слове, можно получить из рис. 12.1.1, если $d_{\min}$ используется вместо L. Полученные потери соответствуют случаю, когда n элементарных двоичных ФМ сигналов сначала детектируются когерентно и складываются согласно (8.1.55), а затем суммы подвергаются квадратичному детектированию или детектируется огибающая для того, чтобы получить M величин для решения. Вероятность ошибки при двоичном различении для последнего случая равна

$$P_2(m) = \frac{1}{2} \exp\left(-\tfrac{1}{2}\gamma_b R_c w_m\right), \tag{8.1.66}$$

и, таким образом,

$$P_M \le \sum_{m=2}^{M} P_2(m)$$

Если вместо распределения весов использовать $d_{\min}$, то объединённая граница для вероятности ошибочного декодирования кодового слова в последнем случае определяется так:

$$P_M \le \tfrac{1}{2}(M-1)\exp\left(-\tfrac{1}{2}\gamma_b R_c w_m\right). \qquad (8.1.67)$$

Требуемую полосу канала для передачи кодированных сигналов можно определить следующим образом. Если для передачи каждого символа кодового слова используется двоичная ФМ, то требуемая полоса примерно равна величине, обратной интервалу времени, требуемого для передачи каждого кодового символа. При скорости передачи R бит/с время, требуемое для передачи k информационных символов и $n-k$ избыточных (проверочных) символов (всего n символов) равно $T = k/R$. Следовательно,

$$W = \frac{1}{T/n} = \frac{n}{k/R} = \frac{R}{R_c}. \qquad (8.1.68)$$

Таким образом, коэффициент расширения полосы B_c для кодированного сигнала

$$B_c = \frac{W}{R} = \frac{n}{k} = \frac{1}{R_c}. \qquad (8.1.69)$$

С другой стороны, если для передачи кодового символа в кодовом слове используется двоичная ЧМ и некогерентное детектирование, то $W \approx 2n/T$ и, следовательно, множитель расширения полосы увеличивается примерно в 2 раза относительно ФМ. В любом случае B_c увеличивается обратно скорости кода или, что эквивалентно, он увеличивается линейно с увеличением размера кода n.

Мы теперь в состоянии сравнить и характеристики качества, и требования по полосе для кодированных сигналов при использовании ортогональных сигналов для передачи кодовых элементов. Сравнение выражения для P_M, даваемого (5.2.21) для ортогональных сигналов, и (8.1.54) для кодированных сигналов с ФМ и когерентным приёмом показывает, что кодирование сигналов ведёт к энергетической потере самое большее $10\lg\left(n/2d_{\min}\right)$ дБ относительно ортогональных сигналов, при том же числе сигналов. С другой стороны, если мы скомпенсируем потери в ОСШ, обусловленные кодированием ФМ, увеличением числа кодовых слов, так что передача кодированных сигналов потребует $M_c = 2^{k_c}$ сигналов, а ортогональных сигналов потребуется $M_0 = 2^{k_0}$, тогда [исходя из объединенных границ (5.2.27) и (8.1.52)] качество, полученное двумя наборами сигналов при больших ОСШ примерно равно, если

$$k_0 = 2R_c d_{\min}. \qquad (8.1.70)$$

Исходя из этих условий, множитель расширения полосы для ортогональных сигналов можно выразить так:

$$B_{e0} = \frac{M_0}{2\log_2 M_0}, \qquad (8.1.71)$$

в то время как для кодированных сигналов имеем $B_{ec} = 1/R_c$. Отношение B_{e0}, даваемое (8.1.71), к B_{ec}

$$\frac{B_{e0}}{B_{ec}} = \frac{2^{2R_c d_{\min}}}{4d_{\min}} \qquad (8.1.72)$$

определяет меру отношения полос между ортогональными сигналами и кодированными сигналами при использовании ФМ и когерентного приёма.

Например, предположим, что мы используем двоичный циклический код (63, 33), который имеет минимальное расстояние $d_{\min} = 12$. Отношение полос ортогональных сигналов и кодированных сигналов ФМ, определяемое (8.1.72), равно 127. Это указывает на частотную эффективность, получаемую кодированием применительно к ортогональным сигналам.

8.1.5. Декодирование жёстких решений

Границы, данные в разделе 8.1.4 для качества кодированных сигналов в канале с АБГШ, основывались на предпосылке, что отсчёты на выходе согласованного фильтра или корреляторa не квантованы. Хотя такая обработка обеспечивает наилучшее качество, принципиальным ограничением являются вычислительные затраты по формированию M корреляционных метрик и их сравнению для определения наибольшей. Количество вычислений получается чрезмерным, когда число кодовых слов M велико.

Для сокращения вычислительных затрат аналоговые отсчеты на выходе согласованных фильтров можно квантовать и операции декодирования выполнить как цифровые. В этом подразделе мы рассмотрим крайний случай, когда каждый отсчёт, соответствующий одному символу кодового слова, квантуется на два уровня: нуль и единица. Это значит, что принято решение (жёсткое решение) о том, передан ли с каждым кодовым символом кодового слова «0» или «1». Результирующий канал с дискретным временем (состоящий из модулятора, канала с АБГШ и демодулятора) образует ДСК с вероятностью ошибки p. Если для передачи кодового символа кодового слова используется двоичная ФМ и осуществляется когерентный приём, то

$$p = Q\left(\sqrt{\frac{2E_c}{N_0}}\right) = Q\left(\sqrt{2\gamma_b R_c}\right). \qquad (8.1.73)$$

С другой стороны, если для передачи кодового символа кодового слова используется двоичная ЧМ, тогда

$$p = Q\left(\sqrt{\gamma_b R_c}\right) \text{ – при когерентном детектировании} \qquad (8.1.74)$$

и

$$p = \tfrac{1}{2}\exp\left(-\tfrac{1}{2}\gamma_b R_c\right) \text{ – при некогерентном детектировании.} \qquad (8.1.75)$$

Декодирование по минимальному расстоянию (правило максимального правдоподобия). Жёсткие решения демодулятора для каждого принятого кодового слова поступают на декодер, который сравнивает принятое кодовое слово с M возможными переданными кодовыми словами и принимает решение в пользу кодового слова, которое ближе всего по Хеммингу к принятому кодовому слову, т.е. отличается от него в наименьшем числе разрядов. Это правило декодирования по минимальному расстоянию оптимально в том смысле, что оно обеспечивает минимальную вероятность ошибочного декодирования кодового слова в двоичном симметричном канале.

Концептуально простой, но в вычислительном отношении неэффективный, метод для декодирования жёстких решений сводится к тому, чтобы сначала суммировать (mod 2) принятый вектор кодового слова со всеми M возможными к передаче кодовыми словами $\mathbf{C}_i$, чтобы получить векторы ошибок $\mathbf{e}_i$. Таким образом, $\mathbf{e}_i$ представляет ошибочное событие, которое должно произойти в канале для того, чтобы превратить кодовое слово $\mathbf{C}_i$ в данное принятое кодовое слово. Число ошибок при превращении $\mathbf{C}_i$ в принятое кодовое слово как раз равно числу единиц в $\mathbf{e}_i$. Таким образом, если мы просто сосчитаем

вес каждого их M векторов $\{\mathbf{e}_i\}$ и примем решение в пользу кодового слова, которое привело к наименьшему весу вектора ошибки, мы фактически имеем реализацию правила декодирования по минимуму расстояния.

Более эффективный метод для декодирования жёстких решений сводится к использованию проверочной матрицы $\mathbf{H}$. Для детальной разработки вопроса, предположим, что $\mathbf{C}_m$ – это переданное кодовое слово, а $\mathbf{Y}$ – принятое кодовое слово на выходе демодулятора. В общем случае $\mathbf{Y}$ можно выразить так

$$\mathbf{Y} = \mathbf{C}_m + \mathbf{e},$$

где $\mathbf{e}$ означает произвольный вектор ошибок. Произведение $\mathbf{Y}\mathbf{H}^{\mathrm{T}}$ даёт с учётом (8.1.9)

$$\mathbf{Y}\mathbf{H}^{\mathrm{T}} = (\mathbf{C}_m + \mathbf{e})\mathbf{H}^{\mathrm{T}} = \mathbf{C}_m\mathbf{H}^{\mathrm{T}} + \mathbf{e}\mathbf{H}^{\mathrm{T}} = \mathbf{e}\mathbf{H}^{\mathrm{T}} = \mathbf{S}, \qquad (8.1.76)$$

где $(n-k)$-мерный вектор $\mathbf{S}$ называется *синдромом образца ошибки.* Другими словами, вектор $\mathbf{S}$ имеет нулевые компоненты для тех уравнений проверочных символов, которые выполняются, и ненулевые компоненты для тех уравнений проверочных символов, которые не выполняются. Таким образом, $\mathbf{S}$ содержит образец неудач проверок.

Подчеркнём, что синдром $\mathbf{S}$ является характеристикой образцов ошибок, но не переданных кодовых слов. Далее заметим, что имеется 2^n возможных образцов ошибок, но только 2^{n-k} синдромных. Следовательно, различные образцы ошибок приводят к одинаковым синдромам.

Предположим, мы сконструируем таблицу декодирования $2^k \times 2^{n-k}$, в заголовке которой запишем все 2^k возможных кодовых слов, начиная кодовым словом из одних нулей в первом (самом левом) столбце. Это кодовое слово из одних нулей представляет также образец ошибки, состоящий из одних нулей. Мы заполним первый столбец путём записи сначала всех образцов ошибок $\{\mathbf{e}_i\}$ с весом 1 (их число равно n+1). Если $n < 2^{n-k}$, мы можем затем записать образцы двойных ошибок, затем образцы тройных ошибок и т.д., пока не получим общее 2^{n-k} записей в первом столбце. Таким образом, число строк, которые можно иметь в таблице, равно 2^{n-k}, что равно числу возможных синдромов. Далее мы суммируем каждый образец ошибки из первого столбца с соответствующими кодовыми словами, расположенными в верхней строке таблицы. Так мы заполняем остаток таблицы следующим образом

$$\begin{array}{ccccc}
\mathbf{C}_1 & \mathbf{C}_2 & \mathbf{C}_3 & \dots & \mathbf{C}_{2^k} \\
\mathbf{e}_2 & \mathbf{C}_2+\mathbf{e}_2 & \mathbf{C}_3+\mathbf{e}_2 & \dots & \mathbf{C}_{2^k}+\mathbf{e}_2 \\
\mathbf{e}_3 & \mathbf{C}_2+\mathbf{e}_3 & \mathbf{C}_3+\mathbf{e}_2 & \dots & \mathbf{C}_{2^k}+\mathbf{e}_2 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
\mathbf{e}_{2^{n-k}} & \mathbf{C}_2+\mathbf{e}_{2^{n-k}} & \mathbf{C}_3+\mathbf{e}_{2^{n-k}} & \dots & \mathbf{C}_{2^k}+\mathbf{e}_{2^{n-k}}.
\end{array}$$

Эту таблицу называют *стандартным расположением* для заданного кода. Каждая строка, включая первую, состоит из k принимаемых кодовых слов, которые образуются из соответствующих образцов ошибок первого столбца. Каждую строку называют смежным классом, а первые (самые левые) кодовые слова (или образцы ошибок) называют лидерами смежных классов. Следовательно, смежный класс состоит из всевозможных принимаемых слов, получающихся от частного образца ошибки (лидера смежного класса).

Пример 8.1.10. Сконструируем стандартное расположение для систематического кода (5, 2) с порождающей матрицей

$$G = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \end{bmatrix}.$$

Этот код имеет минимальное расстояние $d_{\min} = 3$. Стандартное расположение дано в табл. (8.1.7). Заметим, что в этом коде среди 8 лидеров смежных классов один состоит из одних нулей, пять лидеров имеют вес, равный 1 и два имеют вес, равный 2. Хотя имеется много больше образцов двойных ошибок, здесь достаточно двух, чтобы заполнить таблицу.

Табл. 8.1.7. Стандартное расположение для кода (5, 2)

Кодовые слова			
00000	01011	10101	11110
00001	01010	10100	11111
00010	01001	10111	11100
00100	01111	10001	11010
01000	00011	11101	10110
10000	11011	00101	01110
11000	10011	01101	00110
10010	11001	00111	01100

Они были выбраны так, чтобы соответствующие им синдромы отличались от тех, которые соответствуют одиночным ошибкам.

Теперь предположим, что $\mathbf{e}_i$ являются лидером смешанного класса, и что было передано кодовое слово $\mathbf{C}_m$. Тогда образец ошибки $\mathbf{e}_i$ приводит к принимаемому кодовому слову

$$\mathbf{Y} = \mathbf{C}_m + \mathbf{e}_i .$$

Синдром равен

$$\mathbf{S} = \left(\mathbf{C}_m + \mathbf{e}_i\right)\mathbf{H}^{\mathrm{T}} = \mathbf{C}_m\mathbf{H}^{\mathrm{T}} + \mathbf{e}_i\mathbf{H}^{\mathrm{T}} = \mathbf{e}_i\mathbf{H}^{\mathrm{T}} .$$

Ясно, что все принимаемые кодовые слова в тех же смежных классах, образованные из тех же $\mathbf{e}_i$, имеют одинаковые синдромы, поскольку последние зависит только от образцов ошибки. Далее, каждый смежный класс, образованный различным $\mathbf{e}_i$ имеет свой синдром. Установив эти характеристики из стандартного расположения, мы можем просто сконструировать синдромную таблицу декодирования, в которой запишем 2^{n-k} синдромов и соответствующие 2^{n-k} лидеров смешанных классов, которые представляют образцы ошибок с минимальным весом. Затем, для данного принимаемого кодового вектора $\mathbf{Y}$ мы вычисляем синдром

$$\mathbf{S} = \mathbf{Y}\mathbf{H}^{\mathrm{T}} .$$

Для вычисленного синдрома $\mathbf{S}$ мы находим соответствующий (наиболее правдоподобный) вектор ошибки, скажем $\hat{\mathbf{e}}_m$. Этот вектор ошибки суммируется с $\mathbf{Y}$, чтобы получить декодированное слово

$$\hat{\mathbf{C}}_m = \mathbf{Y} \oplus \hat{\mathbf{e}}_m .$$

Пример 8.1.11. Рассмотрим код (5, 2) со стандартным расположением, данным в таблице (8.1.7). Синдромы, соответствующие наиболее правдоподобным образцам ошибок, даны в таблице (8.1.8). Теперь предположим, что действительный вектор ошибок в канале

$$\mathbf{e} = [1\,0\,1\,0\,0].$$

Синдром, вычисленный для этой ошибки, равен $\mathbf{S} = \left[001\right]$. Ошибка, определяемая по этому синдрому из таблицы, равна $\hat{\mathbf{e}} = \left[00001\right]$. Когда $\hat{\mathbf{e}}$ суммируется с $\mathbf{Y}$, результат

приведёт к ошибке декодирования. Другими словами, код (5, 2) корректирует все одиночные ошибки и только две двойные ошибки, именно $[11000]$ и $[10010]$.

Табл. 8.1.8. Таблица синдромов для кода (5, 2)

Синдром	Вектор ошибки
0 0 0	0 0 0 0 0
0 0 1	0 0 0 0 1
0 1 0	0 0 0 1 0
1 0 0	0 0 1 0 0
0 1 1	0 1 0 0 0
1 0 1	1 0 0 0 0
1 1 0	1 1 0 0 0
1 1 1	1 0 0 1 0

Синдромное декодирование циклических кодов. Как описано выше, декодирование жёстких решений для линейного блокового кода можно выполнить, вычислив сначала синдром $\mathbf{S} = \mathbf{Y}\mathbf{H}^{\mathrm{T}}$, затем найдя по таблице синдромов наиболее вероятный образец ошибки $\mathbf{e}$, соответствующий вычисленному синдрому $\mathbf{S}$, и, наконец, суммируя образец ошибки $\mathbf{e}$ с принятым вектором $\mathbf{Y}$, чтобы получить в качестве решения наиболее правдоподобное кодовое слово $\hat{\mathbf{C}}_m$. Когда код циклический, вычисление синдрома можно выполнить с помощью регистра сдвига, подобного тому, который использовался в кодере.

Для более подробного обсуждения рассмотрим систематический циклический код и представим принимаемый кодовый вектор $\mathbf{Y}$ полиномом $Y(p)$. В общем $\mathbf{Y} = \mathbf{C} + \mathbf{e}$, где $\mathbf{C}$ – переданное кодовое слово, а $\mathbf{e}$ – вектор ошибки. Таким образом, имеем

$$Y(p) = C(p) + e(p) = X(p)g(p) + e(p). \qquad (8.1.77)$$

Теперь разделим $Y(p)$ на порождающий полином $g(p)$. Это деление даёт

$$\frac{Y(p)}{g(p)} = Q(p) + \frac{R(p)}{g(p)}$$

или, что эквивалентно,

$$Y(p) = Q(p)g(p) + R(p). \qquad (8.1.78)$$

Остаток $R(p)$ – полином степени, меньшей или равной $n-k-1$. Если объединить (8.1.77) и (8.1.78), получим

$$e(p) = [X(p) + Q(p)]g(p) + R(p). \qquad (8.1.79)$$

Это соотношение показывает, что остаток $R(p)$, полученный от деления $Y(p)$ на $g(p)$, зависит только от полинома ошибки $e(p)$ и, следовательно, $R(p)$ – это просто синдром $S(p)$, связанный с образцом ошибки $\mathbf{e}$. Таким образом,

$$Y(p) = Q(p)g(p) + S(p), \qquad (8.1.80)$$

где $S(p)$ – полином синдрома степени меньшей или равной $n-k-1$. Если $Y(p)$ делится на $g(p)$ точно (без остатка), тогда $S(p) = 0$ и принятое слово определяет решение декодера: $\hat{\mathbf{C}}_m = \mathbf{Y}$.

Деление $Y(p)$ на порождающий полином $g(p)$ можно выполнить посредством регистра сдвига, выполняющего деление, как описано ранее. Сначала принимаемый вектор $\mathbf{Y}$ передвигается по $(n-k)$-каскадному регистру сдвига, как показано на рис. 8.1.9. Первоначально все ячейки регистра сдвига обнулены, а ключ находится в положении 1. После того как весь принятый n-символьный вектор вошёл в регистр, содержимое $n-k$ ячеек образует синдром в порядке следования символов, указанном на рис. 8.1.9. Эти символы поочерёдно извлекаются из регистра при установке ключа в положение 2. По заданному синдрому от $(n-k)$-каскадного регистра сдвига и справочной таблице можно идентифицировать наиболее вероятный вектор ошибки.

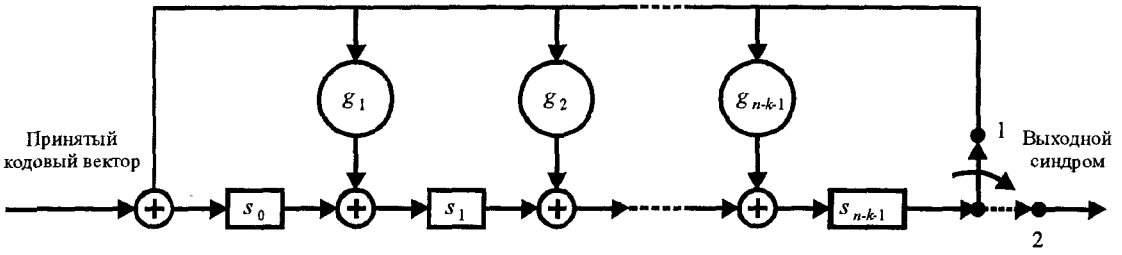
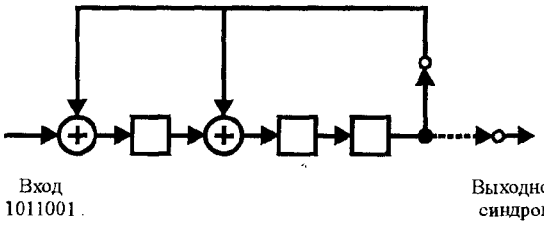


Рис. 8.1.9. (n-k)-каскадный регистр сдвига для вычисления синдрома

Пример 8.1.12. Рассмотрим вычисление синдрома для циклического кода Хемминга (7,4) с порождающим полиномом $g(p)=p^3+p+1$. Предположим, что принят вектор $\mathbf{Y}=[1001101]$. Он подаётся на трехкаскадный регистр, показанный на рисунке 8.1.10.



Сдвиг	Содержимое регистра
0	000
1	100
2	010
3	001
4	010
5	101
6	100
7	110

Рис. 8.1.10. Вычисление синдрома для циклического кода (7,4) с порождающим полиномом $g(p)=p^3+p+1$ при принимаемом векторе $\mathbf{Y}=[1\,0\,0\,1\,1\,0\,1]$

После семи сдвигов содержимое регистров сдвига 110, что соответствуют синдрому $\mathbf{S}=[011]$. Этому синдрому соответствует наиболее правдоподобный вектор ошибки $\mathbf{e}=[0001000]$ и, следовательно,

$$\hat{\mathbf{C}}_m=\mathbf{Y}+\mathbf{e}=[1000101].$$

Информационные символы 1000.

Метод справочной таблицы, использующей синдром, практически используется тогда, когда $n-k$ мало, например, $n-k<10$. Этот метод не подходит для многих интересных мощных кодов. Например, если $n-k=20$, таблица имеет 2^{20} (примерно 1 миллион) записей. Необходимость хранения большого количества данных и затраты времени, требуемые для обнаружения нужных записей в такой большой таблице, делает метод декодирования со справочной таблицей непрактичным для длинных кодов, имеющих большое число проверочных символов.

Для класса циклических кодов и более сложного класса БЧХ кодов были разработаны более эффективные алгоритмы декодирования жёстких решений. Описание этих алгоритмов требует дальнейшей разработки вычислительных методов в конечных полях, которые находятся вне нашей области охвата теории кодирования. Достаточно указать, что существует алгоритмы эффективного декодирования, что делает возможным реализовать длинные БЧХ коды с большой избыточностью в практике цифровых систем связи.

Интересующего читателя отправляем к книгам Питерсона и Уэлдона (1972), Лина и Костелло (1983), Блейхута (1983), Берлекэмпа (1968) и к статье Форни (1965).

Способность обнаружения и исправления ошибок. Из вышеизложенного ясно, что когда синдром содержит одни нули, принимаемое кодовое слово являются одним из 2^k возможных к передаче кодовых слов. Поскольку минимальное расстояние между парой кодовых слов равно $d_{\min}$, образец ошибки весом $d_{\min}$ может трансформировать один из этих 2^k кодовых слов кода в другое кодовое слово. Если это случается, мы имеем *необнаруженную ошибку*. С другой стороны, если действительное число ошибок меньше, чем $d_{\min}$, синдром будет иметь ненулевой вес. Если это случится, мы обнаружили наличие одной или больше ошибок в канале. Ясно, что блоковый (n,k) код способен обнаружить $d_{\min}-1$ ошибок. Обнаружение ошибок можно использовать совместно со схемой автоматического запроса – повторения (АЗП) для повторной передачи кодового слова.

Способность кода исправлять ошибки также зависит от минимального расстояния. Однако, число исправляемых образцов ошибок ограничено числом возможных синдромов или лидеров смежных классов в стандартном расположении. Чтобы определить способность (n,k) кода исправлять ошибки, удобно рассматривать 2^k кодовых слов как точки в n-мерном пространстве. Если каждое кодовое слово рассматривать как центр сферы с радиусом (расстоянием Хемминга) t, то наибольшее значение, которое может принять t без пересечения (или касания) с любой парой из 2^k сфер, равно $t=\left\lfloor \frac{1}{2}\left(d_{\min}-1\right)\right\rfloor$, где $\lfloor x \rfloor$ означает наибольшее целое, содержащееся в x. Внутри каждой сферы лежат всевозможные принимаемые кодовые слова с расстоянием меньшим или равным t от действительно переданного кодового слова. Следовательно, любой принимаемый кодовый вектор, который попадает внутрь сферы, декодируется в разрешённое кодовое слово в центре сферы. Это подразумевает, что (n,k) код с минимальным расстоянием $d_{\min}$ способен исправить $t=\left\lfloor \frac{1}{2}\left(d_{\min}-1\right)\right\rfloor$ ошибок. Рис. 8.1.11 является двумерной моделью кодовых слов и сфер.

Как описано выше, код может быть использован для обнаружения $d_{\min}-1$ ошибок или для исправления $t=\left\lfloor \frac{1}{2}\left(d_{\min}-1\right)\right\rfloor$ ошибок. Ясно, что для исправления t ошибок подразумевается, что мы обнаружили t ошибок. Однако, возможно также обнаружить больше чем t ошибок, если мы пойдём на компромисс по исправляющей способности кода. Например, код с $d_{\min}=7$ может исправить $t=3$ ошибки. Если мы желаем обнаружить четыре ошибки, мы можем это сделать, уменьшив радиус сферы вокруг каждого кодового слова с 3 до 2. Таким образом, обнаруживаются образцы с четырьмя ошибками, но только образцы с двумя ошибками исправляются. Другими словами, если возникнут только две ошибки, они исправятся, а если возникнут три или четыре ошибки, приёмник может запросить повторение. Если возникают более чем четыре ошибки, они останутся необнаруженными, если кодовые слова попадают внутри сферы радиуса 2. Аналогично,

при $d_{\min}=7$ можно обнаружить пять ошибок и одну исправить. В общем, код с минимальным расстоянием $d_{\min}$ может обнаружить e_d ошибок и исправить e_c ошибок, если

$$e_d + e_c \le d_{\min} - 1 \quad \text{и} \quad e_c \le e_d.$$

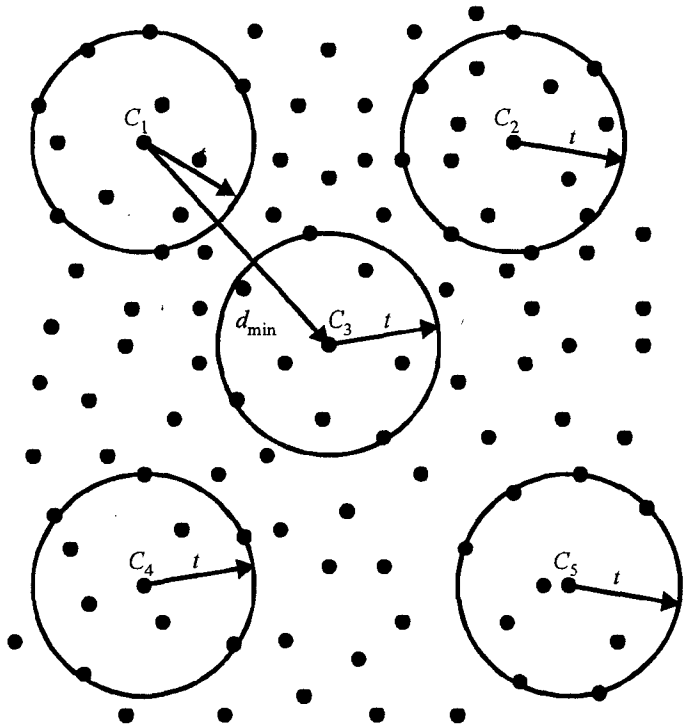


Рис. 8.1.11. Представление кодовых слов в виде центров сфер радиуса $t=\lfloor (d_{\min}-1)/2 \rfloor$

Вероятность ошибки при декодировании с исправлением ошибок. Мы завершим этот раздел определением вероятности ошибки при декодировании жёстких решений для линейных двоичных блоковых кодов, основываясь только на исправлении ошибок.

Из вышеизложенного ясно, что оптимальный декодер в симметричном двоичном канале будет декодировать правильно, если (но не обязательно только если) число ошибок в кодовом слове меньше, чем половина минимального расстояния $d_{\min}$ кода. Это значит, число ошибок не выше $t=\left\lfloor \frac{1}{2}(d_{\min}-1) \right\rfloor$ всегда исправится. В двоичном симметричном канале без памяти ошибки в отдельных символах возникают независимо. Следовательно, вероятность m ошибок в блоке из n символов равна

$$P(m,n) = \binom{n}{m} p^m (1-p)^{n-m}, \tag{8.1.81}$$

и, следовательно, вероятность ошибки кодового слова ограничена сверху выражением

$$P_M \le \sum_{m=t+1}^{n} P(m,n). \tag{8.1.82}$$

Равенство в (8.1.82) имеет место, если линейный блоковый код *совершенный.* Чтобы описать базовые характеристики совершенного кода, предположим, что мы размещаем сферу радиусом t вокруг каждого возможного кодового слова. Каждая сфера вокруг кодового слова содержит набор всех кодовых слов с расстоянием Хемминга относительно центра, меньшим или равным t. Теперь число кодовых слов в сфере радиуса

$t = \left\lfloor \frac{1}{2}(d_{\min} - 1) \right\rfloor$ равно $1 + \binom{n}{1} + \binom{n}{2} + \ldots + \binom{n}{t} = \sum_{i=0}^{t} \binom{n}{i}$.

Поскольку имеется $M = 2^k$ возможных передаваемых кодовых слов, имеется 2^k непересекающихся сфер, каждая с радиусом t. Общее количество кодовых слов, включённых в 2^k сферах, не может превышать 2^n возможных на приёме кодовых слов. Таким образом, код, исправляющий t ошибок, должен удовлетворять неравенству $2^k \sum_{i=0}^{t} \binom{n}{i} \le 2^n$ или, что эквивалентно,

$$2^{n-k} \ge \sum_{i=0}^{t} \binom{n}{i}. \qquad (8.1.83)$$

Совершенный код имеет свойство, что все сферы с радиусом $t = \left\lfloor \frac{1}{2}(d_{\min} - 1) \right\rfloor$ вокруг $M = 2^k$ возможных к передаче кодовых слов разъединены и каждое принимаемое кодовое слово попадает в одну из сфер. Таким образом, каждое принимаемое кодовое слово самое большее находится на расстоянии t от возможных к передаче кодовых слов, и в (8.1.82) выполняется знак равенства. Для такого кода все образцы ошибки с весом, меньшим или равным t, исправляются оптимальным декодером. С другой стороны, образцы ошибки весом $t+1$ или большим не могут быть исправлены. Следовательно, выражение для вероятности ошибки, даваемое (8.1.82), выполняется со знаком равенства. Код Голея (23, 12), имеющий $d_{\min} = 7$ и $t = 3$, является совершенным кодом. Коды Хемминга, которые имеют три параметра $n = 2^{n-k} - 1$, $d_{\min} = 3$ и $t = 1$, также совершенные коды. Эти два нетривиальных кода и тривиальный $(n,1)$ код, состоящий из двух кодовых слов с нечётным числом символов n единиц и нулей, причём $d_{\min} = n$, являются единственными совершенными двоичными блоковыми кодами. Эти коды оптимальны в ДСК в том смысле, что они приводят к меньшей вероятности ошибки декодирования среди всех кодов, имеющих ту же длину блока и то же число информационных символов.

Оптимальные свойства, определённые выше, имеют место и для *квазисовершенных* кодов. Квазисовершенный код характеризуется свойством, что все сферы с радиусом Хемминга t вокруг M возможных к передаче кодовых слов не пересекаются, и каждое принимаемое кодовое слово находится самое большее на расстоянии $t+1$ от возможных к передаче кодовых слов. Для такого кода все образцы ошибок с весом, меньшим или равным t и некоторые образцы ошибок весом $t+1$ исправляются, но ни один образец ошибки весом $t+2$ или больше не исправляются в кодовом слове. Ясно, (8.1.82) является для этих кодов верхней границей для вероятности ошибки, а

$$P_M \ge \sum_{m=t+2}^{n} P(m,n) \qquad (8.1.84)$$

является нижней границей.

Более точную меру качества для квазисовершенных кодов можно получить, используя неравенство в (8.1.83). Общее число кодовых слов, исключая 2^k сфер радиусом t, равно

$$N_{t+1} = 2^n - 2^k \sum_{i=0}^{t} \binom{n}{i}.$$

Если эти кодовые слова поровну разделить на 2^k наборов и каждый набор связать с одной из 2^k сфер, тогда число кодовых слов в каждой сфере увеличивается на

$$\beta_{t+1} = 2^{n-k} - \sum_{i=0}^{t} \binom{n}{i} \qquad (8.1.85)$$

кодовых слов, имеющих расстояние $t+1$ от переданных кодовых слов. Как следствие, от $\binom{n}{t+1}$ образцов ошибок с расстоянием $t+1$ от каждого кодового слова мы можем исправить β_{t+1} образцов ошибок. Таким образом, вероятность ошибочного декодирования квазисовершенного кода можно выразить так:

$$P_M = \sum_{m=t+2}^{n} P(m,n) + \left[\binom{n}{t+1} - \beta_{t+1}\right] p^{t+1}(1-p)^{n-t-1}. \qquad (8.1.86)$$

Имеется много известных квазисовершенных кодов, хотя они не существуют для всех наборов n и k. Поскольку такие коды оптимальны для двоичного симметричного канала, то любой (n,k) линейный блоковый код должен иметь вероятность ошибки, которая, по крайней мере, принимает значение (8.1.86). Следовательно, (8.1.86) является нижней границей вероятности ошибки декодирования для любого линейного блокового кода (n,k), где t – наибольшее целое, так чтобы $\beta_{t+1} \geq 0$.

Другую пару значений для верхней и нижней границ можно получить, рассматривая два кодовых слова, которые отличаются на минимальное расстояние. Сначала заметим, что P_M не может быть меньше, чем вероятность ошибочного декодирования переданного кодового слова в ближайшее кодовое слово, которое имеет расстояние $d_{\min}$ от переданного кодового слова. Т.е.

$$P_M \geq \sum_{m=\lfloor d_{\min}/2 \rfloor+1}^{d_{\min}} \binom{d_{\min}}{m} p^m (1-p)^{d_{\min}-m}. \qquad (8.1.87)$$

С другой стороны, P_M не может быть больше, чем умноженная в $M-1$ раз вероятность ошибочного декодирования переданного кодового слова в ближайшее кодовое слово, которое имеет расстояние $d_{\min}$ от переданного кодового слова:

$$P_M \leq (M-1) \sum_{m=\lfloor d_{\min}/2 \rfloor+1}^{d_{\min}} \binom{d_{\min}}{m} p^m (1-p)^{d_{\min}-m}. \qquad (8.1.88)$$

Когда M велико, нижняя граница (8.1.88) и верхняя граница (8.1.87) весьма далеки одна от другой. Плотную верхнюю границу для P_M можно получить, используя границу Чернова, представленную раньше в разделе 2.1.6. Предположим снова, что передается кодовое слово из одних нулей. При сравнении принимаемого кодового слова с кодовым словом из одних нулей и с кодовым словом веса w_m вероятность ошибки декодирования, полученная из границы Чернова (задача 8.22), ограничена сверху выражением

$$P_2(w_m) \leq \left[4p(1-p)\right]^{w_m/2}. \qquad (8.1.89)$$

Объединение вероятностей двоичных решений ведет к верхней границе

$$P_M \leq \sum_{m=2}^{M} \left[4p(1-p)\right]^{w_m/2}. \qquad (8.1.90)$$

Более простую версию (8.1.90) можно получить, если взять $d_{\min}$ вместо распределения весов. Т.е.

$$P_M \leq (M-1)\left[4p(1-p)\right]^{w_m/2}. \qquad (8.1.91)$$

Конечно, (8.1.90) – более плотная граница, чем (8.1.91).

В разделе (8.1.6) мы сравним различные границы, данные выше, для конкретного кода, а именно кода Голея (23, 12). Дополнительно мы сравним характеристики качества при декодировании жёстких и мягких решений.

8.1.6. Сравнение качества декодирования жёстких и мягких решений

Интересно и поучительно сравнить границы характеристик качества линейных блоковых кодов в канале с АБГШ при декодировании мягких и жёстких решений.

Для иллюстрации мы используем код Голея (23, 12), который имеет относительно простое распределение весов, данных в таблице 8.1.1. Как было констатировано раньше, этот код имеет минимальное расстояние $d_{\min} = 7$. Сначала мы рассчитаем и сравним границы вероятности ошибки при декодировании жёстких решений. Поскольку код Голея (23, 12) является совершенным, точное выражение для вероятности декодирования жёстких решений равно

$$P_M = \sum_{m=4}^{23} \binom{23}{m} p^m (1-p)^{23-m} = 1 - \sum_{m=0}^{3} \binom{23}{m} p^m (1-p)^{23-m}, \qquad (8.1.92)$$

где p – вероятность ошибки двоичного элемента в двоичном симметричном канале. Предполагается, что для передачи двоичных элементов кодового слова используется двоичная (или четверичная) ФМ и осуществляется когерентная обработка в месте приёма.

При этих условиях соответствующее выражение для p дано в (8.1.13). Дополнительно к точному выражению для вероятности ошибки, даваемое (8.1.92), мы имеем нижнюю границу, даваемую (8.1.81) и три верхние границы, даваемые (8.1.88), (8.1.90) и (8.1.91).

Численные результаты, полученные из этих границ, сравниваются с точным значением вероятности ошибки на рис. 8.1.12.

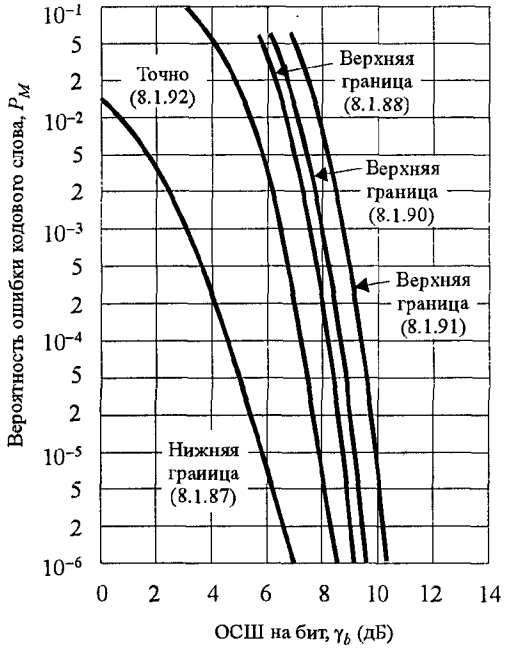


Рис. 8.1.12. Сравнение граничных и точных значений вероятности ошибки для декодирования жёстких решений для кода Голея (23,12)

Видим, что нижняя граница очень свободная. При $P_M = 10^{-5}$ нижняя граница отличается примерно на 2 дБ по сравнению с точным значением вероятности ошибки. При $P_M = 10^{-2}$ разница увеличивается примерно до 4 дБ. Из трёх верхних границ наиболее плотная та, которая определяется (8.1.88); она отличается меньше, чем на 1 дБ по сравнению с точным значением вероятности ошибки при $P_M = 10^{-5}$. Граница Чернова (8.1.90), которая использует распределения весов, также относительно плотная. Наконец,

граница Чернова, которая использует только минимальное кодовое расстояние, наихудшая из трёх. При $P_M = 10^{-5}$ она отличается от точного значения вероятности ошибки примерно на 2 дБ. Все три верхние границы очень неточные для значений вероятности ошибки выше $P_M = 10^{-2}$.

Интересно также сравнить характеристики качества при декодировании мягких и жёстких решений. Для этого сравнения мы используем верхние границы для вероятности ошибки при декодировании мягких решений, даваемые (8.1.52), и точное выражение для вероятности ошибки для декодирования жёстких решений, определяемое (8.1.92). Рис. 8.1.13 иллюстрирует эти характеристики качества.

Мы видим, что две границы для декодирования мягких решений отличаются примерно на 0,5 дБ при $P_M = 10^{-6}$ и примерно на 1 дБ при $P_M = 10^{-2}$. Мы также видим, что разница в качестве декодирования жёстких и мягких решений примерно равна 2 дБ в области $10^{-2} \le P_M \le 10^{-6}$. В области $P_M > 10^{-2}$ кривая вероятности ошибки при декодировании с жёстким решением пересекает кривые для границ вероятности ошибки при декодировании мягких решений. Такое поведение указывает на то, что границы для декодирования мягких решений неточные, когда $P_M > 10^{-2}$.

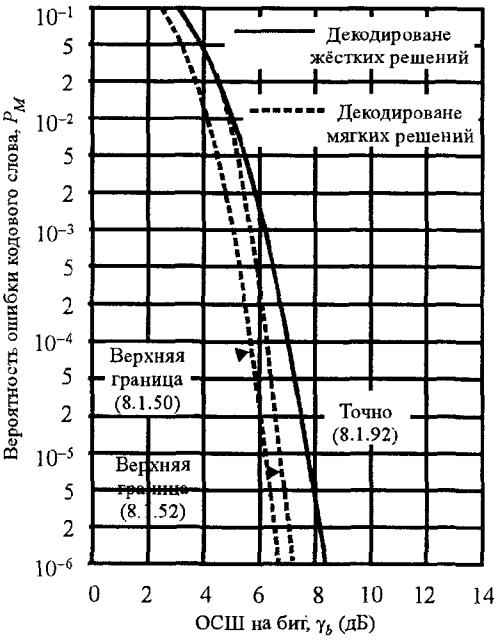


Рис. 8.1.13. Сравнение декодирования мягких и жёстких решений для кода Голея (23,12)

Разница в 2 дБ (по ОСШ на бит) между декодированием жёстких и мягких решений применима не только для кода Голея, но является фундаментальным результатом, который применим в общем к кодированию в системах цифровой связи в каналах с АБГШ. Этот результат получен ниже при расчёте пропускной способности канала с АБГШ при декодировании жёстких и мягких решений.

Пропускная способность ДСК в битах/символ, полу́ченная в разделе 7.1.2, равна

$$C = 1 + p\log_2 p + (1-p)\log_2(1-p), \qquad (8.1.93)$$

где вероятность ошибочного приёма двоичного элемента p при использовании когерентной ФМ в канале с АБГШ определяется (8.1.13). Предположим, что мы используем (8.1.13) для p, и пусть $C = R_c$ в (8.1.93), и тогда определим γ_b, которое удовлетворяет уравнению. Результат показан на рис. 8.1.14 в форме кривой зависимости R_c от γ_b. Например, предположим, что мы хотим использовать код со скоростью $R_c = \frac{1}{2}$. Для этой скорости кода видим, что минимальное значение ОСШ на бит, требуемое для достижения пропускной способности канала при декодировании жёстких решений, равно примерно 1,6 дБ.

Каков предел ОСШ на бит, когда скорость кода стремится к нулю? Для многих значений R_c вероятность p можно аппроксимировать так:

$$p \approx \tfrac{1}{2} - \sqrt{\gamma_b R_c/\pi} \qquad (8.1.94)$$

Если выражение для p подставить в (8.1.93), а логарифм в (8.1.93) аппроксимировать как

$$\log_2(1+x) \approx \left(x - \tfrac{1}{2}x^2\right)/\ln 2,$$

формула для пропускной способности канала сводится к

$$C = \frac{2}{\pi \ln 2}\gamma_b R_c. \qquad (8.1.95)$$

Теперь положим $C = R_c$. Тогда, в пределе, когда $R_c \to 0$, получаем результат

$$\gamma_b = \tfrac{1}{2}\pi \ln 2 \quad (0{,}37 \text{ дБ}). \qquad (8.1.96)$$

Пропускную способность канала с АБГШ с двоичным входом при декодировании мягких решений можно вычислить аналогичным образом. В разделе 7.1.2 мы привели выражение для пропускной способности (в битах на кодовый символ) для этого случая

$$C = \tfrac{1}{2}\sum_{k=0}^{1}\int_{-\infty}^{\infty} p(y|k)\log_2 \frac{p(y|k)}{p(y)}dy, \qquad (8.1.97)$$

где $p(y|k)$, $k = 0, 1$, означает ФПВ для выхода демодулятора при условии, что передан, соответственно, символ 0 или 1. Для канала с АБГШ имеем

$$p(y|k) = \frac{1}{\sqrt{2\pi}}e^{-(y-m_k)^2/2\sigma^2}, \quad k = 0, 1, \qquad (8.1.98)$$

где $m_0 = -\sqrt{\mathscr{E}_c}$, $m_1 = \sqrt{\mathscr{E}_c}$, $\sigma^2 = \tfrac{1}{2}N_0$ и $\mathscr{E}_c = R_c\mathscr{E}_b$. Безусловная плотность вероятности $p(y)$ определяется половиной суммы $p(y|1)$ и $p(y|0)$. Поскольку R_c стремится к нулю, выражение (8.1.97) для пропускной способности канала можно аппроксимировать формулой

$$C \approx \gamma_b R_c/\ln 2. \qquad (8.1.99)$$

Опять предположим $C = R_c$. Таким образом, когда $R_c \to 0$, минимальное значение ОСШ на бит, требуемое для достижения пропускной способности, равно

$$\gamma_b = \ln 2 \quad (-1{,}6 \text{ дБ}). \qquad (8.1.100)$$

Используя (8.1.98) в (8.1.97) и положив $C = R_c$ можно получить численное решение для скорости кода в области $0 \le R_c \le 1$. Результат этих расчётов также показан на рис. 8.1.14.



Рис. 8.1.14. Скорость кода как функция минимального ОСШ на бит при декодировании мягких и жёстких решений

Из вышеизложенного следует, что в пределе, когда R_c стремится к нулю, разница в ОСШ γ_b между декодированием жёстких и мягких решений равна $\frac{1}{2}\pi$, что приближённо равно 2 дБ. С другой стороны, по мере увеличения R_c до единицы, разница в γ_b для двух разновидностей декодирования уменьшается.

Например, при $R_c = 0{,}8$ разница примерно равна 1,5 дБ. Кривые на рис. 8.1.14 дают больше информации, чем только разницу в качестве между декодированием мягких и жёстких решений. Эти кривые подробно определяют требуемые значения минимума ОСШ для данной скорости кода. Например, скорость кода $R_c = 0{,}8$ может обеспечить сколь угодно малую вероятность ошибки при ОСШ на бит 2 дБ, если используется декодирование мягких решений.

Для сравнения заметим, что двоичная ФМ при отсутствии кодирования требует 9,6 дБ для достижения вероятности ошибки 10^{-5}. Следовательно, возможен выигрыш в 7,6 дБ при использовании скорости кода $R_c = \frac{4}{5}$. К сожалению, для достижения такого большого выигрыша за счёт кодирования обычно требуется применение очень длинных кодовых блоков, которые ведут к очень сложному приёмнику. Тем не менее, кривые рис. 8.1.14 дают оценку для сравнения выигрыша кодирования, достигаемого практически реализуемыми кодами с основными ограничениями при декодировании мягких или жёстких решений.

Вместо сравнения различия между декодированием жёстких и мягких решений, основанного на соотношениях пропускной способности канала, мы можем делать такие же простые сравнения, основанные на параметрах скорости при случайном кодировании. В главе 7 мы показали, что средняя вероятность ошибки по ансамблю случайно выбранных двоичных кодовых слов имеет верхнюю границу

$$\overline{P_e} \le 2^{-n(R_0 - R_c)}, \qquad (8.1.101)$$

где $R_c = k/n$ – скорость кода, а предельная скорость R_0 связывает верхнюю границу с R_c так, что $P_e \to 0$ при $n \to \infty$. Для неквантованного декодирования (мягких решений) R_0 определяется так

$$R_0 = \log_2 \frac{2}{1 + e^{-\mathscr{E}_c/N_0}}, \qquad (8.1.102)$$

где $\mathscr{E}_c/N_0 = R_c\gamma_b$ – ОСШ на измерение. Этот результат был получен в разделе 7.2.

С другой стороны, если выход демодулятора квантуется на Q уровней до декодирования, то границу Чернова можно использовать в качестве верхней границы для усреднённой по ансамблю вероятностей двоичной ошибки $\overline{P_2(\mathbf{s}_l, \mathbf{s}_m)}$, данной в разделе 7.2. Результат такого расчёта – та же верхняя граница для $\overline{P_e}$, определённая по (8.1.101), но с заменой R_0 на R_Q, где

$$R_Q = \max_{p_j}\left\{-\log_2 \sum_{i=0}^{Q-1}\left[\sum_{j=0}^{1} p_j \sqrt{P(i|j)}\right]^2\right\}. \qquad (8.1.103)$$

В (8.1.103) $\{p_j\}$ – априорные вероятности двух сигналов на входе канала и $\{P(i|j)\}$ означают переходные вероятности канала. Например, для случая ДСК и $p_1 = p_0 = \frac{1}{2}$, $P(0|0) = P(1|1) = 1 - p$ и $P(0|1) = P(1|0) = p$ следует

$$R_Q = \log_2 \frac{2}{1 + \sqrt{4p(1-p)}}, \qquad Q = 2, \qquad (8.1.104)$$

где

$$p = Q\left(\sqrt{2\gamma_b R_c}\right). \qquad (8.1.105)$$

Кривые R_Q в зависимости от $10\lg(\mathscr{E}_c/N_0)$ иллюстрируется на рис. 8.1.15 для $Q = 2$ и $Q = \infty$ (декодирование мягких решений).

Заметим, что разница в качестве декодирования между неквантованным декодированием мягких решений и декодированием жёстких решений приблизительно равно 2 дБ. Фактически, снова можно легко показать, что при $\mathscr{E}_c/N_0 \to 0$ потеря в качестве, обусловленная декодированием жёстких решений, равна $10\lg(\pi/2) \approx 2$ дБ, что является той же разницей в децибелах, которая была получена в нашем сравнении при

использовании соотношений для пропускной способности канала. Напомним, что около 1 дБ этих потерь можно восполнить квантованием выхода демодулятора на трёх уровнях вместо двух (см. задачу 7.11). Дополнительное улучшение возможно путём квантования выхода демодулятора на число уровней, большее трёх, как показано в разделе 7.3.

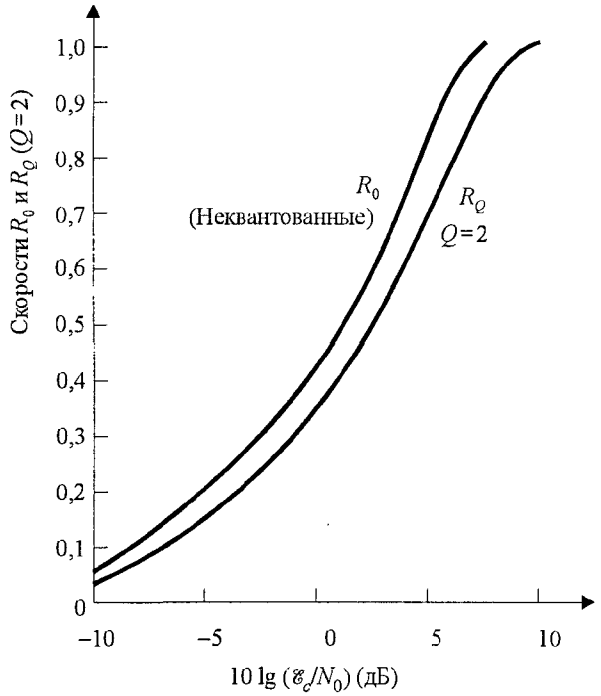


Рис. 8.1.15. Сравнение R_0 (декодирование мягких решений) с R_Q (декодирование жёстких решений) в функции от ОСШ на измерение

8.1.7. Границы для минимальных расстояний линейных блоковых кодов

Выражения для вероятности ошибки, полученные в этой главе для декодирования мягких и жёстких решений линейных двоичных блоковых кодов, ясно указывает на важное значение параметра минимальное кодовое расстояние для качества кода. Если мы, например, рассмотрим декодирование мягких решений, верхняя граница вероятности ошибки, представленная в (8.1.52), указывает на то, что для заданной скорости кода $R_c = k/n$ вероятность ошибки в канале с АБГШ уменьшается экспоненциально с $d_{\min}$. Если эту границу использовать в соединении с нижней границей для $d_{\min}$, данной ниже, мы получаем верхнюю границу для P_M, которую можно достичь многими известными кодами. Аналогично, мы можем использовать верхнюю границу данную в (8.1.82) для вероятности ошибки при декодировании жёстких решений в соединении с нижней границей для $d_{\min}$ для получения верхней границы для вероятности ошибочного декодирования линейных двоичных блоковых кодов в ДСК.

С другой стороны, верхнюю граница для $d_{\min}$ можно использовать для определения нижней границы вероятности ошибки, достигаемой наилучшими кодами. Для примера, предположим, что используется декодирование жёстких решений. В этом случае мы имеем две нижние границы для P_M, даваемые (8.1.86) и (8.1.87), причём первая более плотная. Если хотя бы одна из этих границ использовалась совместно с верхней границей для $d_{\min}$, то результатом будет нижняя граница для P_M для наилучшего (n,k) кода. Таким образом, верхние и нижние границы с $d_{\min}$ очень важны для оценки эффективности кодов.

Простая верхняя граница для минимального расстояния двоичного или недвоичного линейного блокового кода (n,k) была дана в (8.1.14) как $d_{\min} \le n-k+1$. Удобно нормировать это выражение через длину блока n. Это даёт

$$\frac{d_{\min}}{n} \le \left(1-R_c\right)+\frac{1}{n}, \tag{8.1.106}$$

где R_c – скорость кода. Для больших n слагаемым $1/n$ можно пренебречь. Если код имеет наибольшее возможное расстояние, т.е. $d_{\min}=n-k+1$, его называют *разделимым кодом с максимальным расстоянием*. Исключая случаи тривиального кода $(n, 1)$ для передачи двоичных сообщений с повторением, не существует двоичных разделимых кодов с максимальным расстоянием. Фактически верхняя граница в (8.1.106) для двоичных кодов весьма неточная. С другой стороны, существуют недвоичные коды с $d_{\min}=n-k+1$. Например, коды Рида–Соломона, которые представляют подкласс БЧХ кодов, являются разделимыми кодами с максимальным расстоянием. В дополнение к верхней границе, данной выше, имеется несколько плотных границ для минимального расстояния линейных блоковых кодов. Мы вкратце опишем четыре важные границы, три из них верхние границы, а четвертая нижняя. Доказательство этих границ сложное и не представляет особого интереса в нашем последующем обсуждении. Интересующемуся читателю можно порекомендовать главу 4 книги Питерсона и Уэлдона (1972) с этими доказательствами.

Одна верхняя граница для минимальных расстояний может получиться из неравенства (8.1.83). Взяв логарифм от обеих частей (8.1.83) и разделив на n, мы получим

$$1-R_c \ge \frac{1}{n}\log_2 \sum_{i=0}^{t}\binom{n}{i} \tag{8.1.107}$$

Поскольку эффективность кода, измеряемая параметром t, связана с минимальным расстоянием, (8.1.107) является верхней границей для минимального расстояния. Ее называют *верхней границей Хемминга.*

Получена асимптотическая форма (8.1.107) при $n \to \infty$. Теперь для любого n пусть t_0 является наибольшим целым t, при котором выполняется (8.1.107). Тогда можно показать (Питерсон и Уэлдон, 1972) что при $n \to \infty$ отношение t/n для каждого (n,k) блокового кода не может превысить t_0/n, где t_0/n удовлетворяет условию

$$1-R_c = H\left(t_0/n\right), \tag{8.1.108}$$

а $H(\cdot)$ – двоичная энтропийная функция, определяемая (3.2.10).

Обобщение границы Хемминга на недвоичные коды с основанием q простое:

$$1-R_c \ge \frac{1}{n}\log_q\left[\sum_{i=0}^{t}\binom{n}{i}(q-1)^i\right] \tag{8.1.109}$$

Другую верхнюю границу, открытую Плоткиным (1960), можно определить так. Число проверочных символов, требуемое для достижения минимального расстояния $d_{\min}$ в линейном блоковом коде (n,k), удовлетворяет неравенству

$$n-k \ge \frac{qd_{\min}-1}{q-1}-1-\log_q d_{\min}. \tag{8.1.110}$$

Для двоичного кода (8.1.110) можно выразить так

$$\frac{d_{\min}}{n}\left(1-\frac{1}{2d_{\min}}\log_2 d_{\min}\right) \le \frac{1}{2}\left(1-R_c+\frac{2}{n}\right).$$

В пределе, когда $n \to \infty$, при $d_{\min}/n \le \frac{1}{2}$ (8.1.110) приводит к

$$d_{\min}/n \le \tfrac{1}{2}\left(1 - R_c\right). \qquad (8.1.111)$$

Наконец, имеется другая плотная верхняя граница для минимального расстояния, полученная Элиасом (Берлекэмп, 1968). Её можно выразить в асимптотической форме:

$$d_{\min}/n \le 2A(1-A), \qquad (8.1.112)$$

где параметр A связан со скоростью кода уравнением

$$R_c = 1 + A\log_2 A + (1-A)\log_2(1-A), \quad 0 \le A \le \tfrac{1}{2}. \qquad (8.1.113)$$

Существуют также нижние границы для минимального расстояния линейного блокового кода (n,k). В частности, существует двоичный блоковый код, имеющий нормированное минимальное расстояние, которое асимптотически удовлетворяет неравенству

$$d_{\min}/n \ge \alpha, \qquad (8.1.114)$$

где α связано со скоростью кода уравнением

$$R_c = 1 - H(\alpha) = 1 + \alpha\log_2 \alpha + (1-\alpha)\log_2(1-\alpha), \quad 0 \le \alpha \le \tfrac{1}{2}. \qquad (8.1.115)$$

Эта нижняя граница – частный случай нижней границы, открытой Гильбертом (1952) и Варшамовым (1957), которая применима для недвоичных и двоичных блоковых кодов.

Асимптотические границы, данные выше, приведены на рис. 8.1.16 для двоичных кодов. С целью сравнения на рисунке даны также кривые минимального расстояния, как функции скорости кода для БЧХ кодов с длиной блоков $n = 31$ и 63.

Видно, что для $n = 31$ и 63 нормированное минимальное расстояние хорошо ложится над нижней границей Варшамова–Гильберта. По мере увеличения длины блока n, эффективность БЧХ кодов ослабевает. Например, когда $n = 1023$, кривая нормированного минимального расстояния ложится близко к границе Варшамова–Гильберта. Если n возрастает выше $n = 1023$, нормированное минимальное расстояние БЧХ кода продолжает уменьшается и падает ниже границы Варшамова–Гильберта. Это значит, что $d_{\min}/n$ приближается к нулю по мере того как n стремится к бесконечности. Следовательно, БЧХ коды, которые являются наиболее важным классом циклических кодов, не очень эффективны при больших длинах блоков.

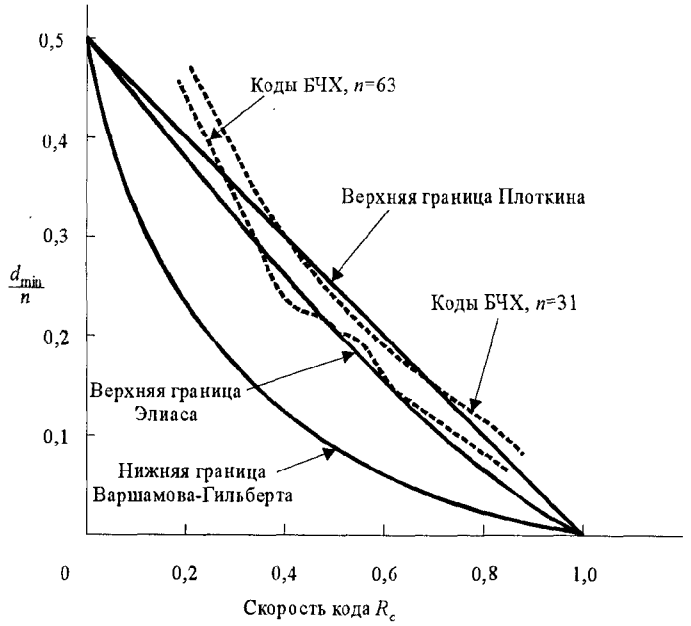


Рис. 8.1.16. Верхняя и нижняя границы нормированного минимального расстояния в функции от скорости кода

8.1.8. Недвоичные блоковые коды и каскадные блоковые коды

Недвоичные блоковые коды состоят из набора кодовых слов фиксированной длины, в которых каждый элемент кодового слова выбирается из алфавита, содержащего q символов, обозначаемых $\{0, 1, 2,\ldots,q-1\}$. Обычно $q=2^k$, так что k информационных бит отображается одним из q символов. Длина недвоичного кодового слова обозначается через N, а число информационных символов, закодированных блоком из N символов, обозначается K. Минимальное расстояние недвоичного кода обозначается $D_{\min}$. Систематический блоковый код (N,K) содержит K информационных символов и $N-K$ проверочных символов.

Среди различных типов недвоичных линейных блоковых кодов коды Рида–Соломона являются одними из самых важных для практических приложений. Как было указано раньше, они составляют подкласс БЧХ кодов, которые, в свою очередь, являются классом циклических кодов. Эти коды описываются параметрами

$$N = q-1 = 2^k-1$$
$$k = 1, 2, 3,\ldots N-1$$
$$D_{\min} = N-K+1$$
$$R_c = K/N$$

Такой код гарантированно исправляет до

$$t=\left\lfloor \tfrac{1}{2}\left(D_{\min}-1\right)\right\rfloor=\left\lfloor \tfrac{1}{2}(N-K)\right\rfloor \qquad (8.1.117)$$

ошибок символов. Конечно, эти коды могут быть расширены и укорочены так, как было описано ранее для двоичных блоковых кодов.

Распределение весов $\{A_i\}$ класса кодов Рида–Соломона известно. Коэффициенты во взвешивающем полиноме определяются так:

$$A_i=\binom{N}{i}(q-1)\sum_{j=0}^{i-D}(-1)^j\binom{i-1}{j}q^{i-j-D}, \quad i\geq D_{\min}, \qquad (8.1.118)$$

где $D\equiv D_{\min}$ и $q=2^k$.

Одно объяснение важности кодов Рида–Соломона – их хорошие дистанционные свойства. Второе объяснение их важности – существование эффективных алгоритмов декодирования жёстких решений, которые делают возможным реализовать относительно длинные коды во многих практических приложениях, где требуется кодирование.

Недвоичный код хорошо согласован с техникой M-ичной модуляции для передачи 2^k возможных символов. В частности, часто используется M-ичная ортогональная система сигналов, например, M-ичная ЧМ. Каждый из 2^k символов в q-ичном алфавите отображается в один из $M=2^k$ ортогональных сигналов. Таким образом, передача кодового слова связана с передачей N ортогональных сигналов, где каждый сигнал выбирается из набора $M=2^k$ возможных сигналов.

Оптимальный демодулятор для такого сигнала, искажённого в канале с АБГШ, состоит из M согласованных фильтров (или взаимных корреляторов) чьи выходы подаются к декодеру в форме мягких или жёстких решений. Если демодулятор вынес жёсткие решения, то вероятность ошибки символа P_M и параметры кода достаточны для характеристики качества декодера. Действительно, модулятор, канал с АБГШ и демодулятор формируют эквивалентный симметричный канал без памяти с дискретным (M-ичным) входом и дискретным (M-ичным) выходом, характеризуемый переходными

вероятностями $P_c = 1 - P_M$ и $P_M/(M-1)$. Эта модель канала, которая иллюстрируется на рис. 8.1.17, является обобщением ДСК.

Качество декодера жёстких решений можно характеризовать следующей верхней границей для вероятности ошибки кодового слова:

$$P_e \le \sum_{i=t+1}^{N} \binom{N}{i} P_M^t \left(1 - P_M\right)^{N-t}, \qquad (8.1.119)$$

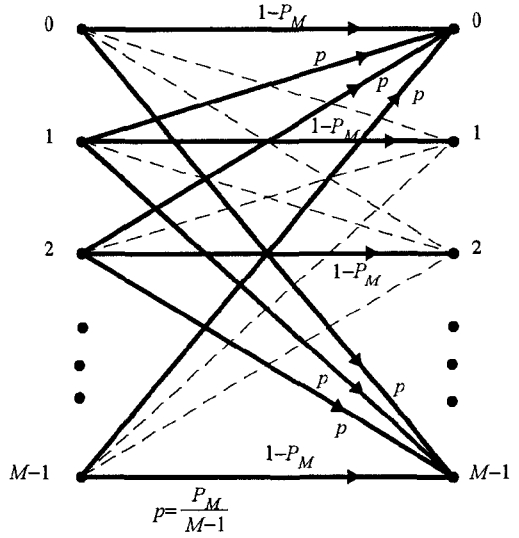


Рис. 8.1.17. M-ичный по входу, M-ичный по выходу симметричный канал без памяти

где t – число ошибок, гарантированно исправляемое кодом.

Ошибке кодового слова соответствует вероятность ошибки символа

$$P_{es} = \frac{1}{N} \sum_{i=t+1}^{N} i \binom{N}{i} P_M^t \left(1 - P_M\right)^{N-t}. \qquad (8.1.120)$$

Далее, если символы отображаются битами, то вероятность ошибки на бит, исходя из (8.1.120), равна

$$P_{eb} = \frac{2^{k-1}}{2^k - 1} P_{es}. \qquad (8.1.121)$$

Пример 8.1.13. Рассчитаем качество кода Рида–Соломона при $N = 2^5 - 1 = 31$, $D_{\min} = 3, 5, 9, 17$. Соответствующие значения K равны 29, 27, 23 и 15. При модуляции используются $M = q = 32$ ортогональных сигнала ЧМ и некогерентное детектирование на приёме.

Вероятность ошибочного приёма символа определяется (5.4.46), и её можно выразить в виде

$$P_M = \frac{1}{M} e^{-\gamma} \sum_{n=2}^{M} (-1)^n \binom{M}{n} e^{-\gamma/n}, \qquad (8.1.122)$$

где γ – ОСШ на кодовый символ. Используя (8.1.122) в (8.1.120) и объединяя результат с (8.1.121), мы получим вероятность ошибки на бит. Результаты этих расчётов даны на рис. 8.1.18.

Заметим, что более мощные коды (больше $D_{\min}$) дают худшее качество при низких ОСШ на бит, чем слабые коды. С другой стороны, при высоких ОСШ более мощные коды дают лучшее качество. Таким образом, имеется пересечение среди различных кодов, как показано для примера на рис. 8.1.18 для значений $t = 1$ и $t = 8$.

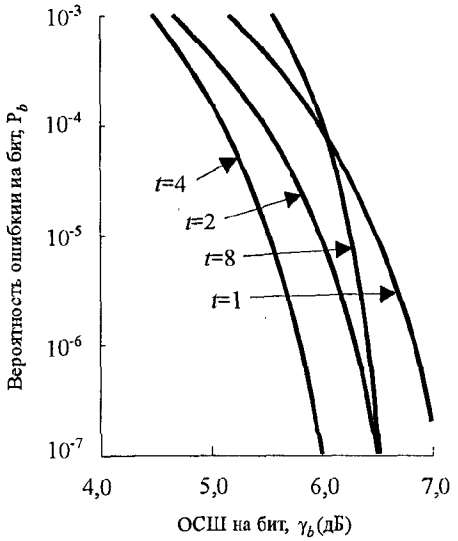


Рис. 8.1.18. Характеристики некоторых кодов Рида-Соломона, исправляющих t ошибок, с N=31, 32-позиционной ЧМ в канале с АБГШ (некогерентная демодуляция)

Пересечение также возникает для кодов с $t = 1$, 2 и 4 при малых значениях ОСШ на бит. Аналогично, кривые для $t = 4$, 8 и $t = 8$, 2 пересекается в области больших значений ОСШ. Это характерное поведение кодированных сигналов при некогерентном детектировании.

Если демодулятор не выносит жёсткое решение по каждому кодовому символу, но, вместо этого, отправляет неквантованные выходы согласованных фильтров к декодеру, можно использовать декодирование мягких решений. Такое декодирование включает в себя формирование $q^k = 2^{kK}$ корреляционных метрик, где каждая метрика соответствует одному из q^k кодовых слов и состоит из суммы выходов N согласованных фильтров, соответствующих N кодовым символам. Выходы согласованных фильтров можно (1) суммировать когерентно, (2) детектировать по огибающей, а затем суммировать или (3) квадратировать и затем суммировать. Если используется когерентное детектирование, а в канале действует АБГШ, расчёт вероятности ошибки является простым обобщением двоичного случая, рассмотренного в разделе 8.1.4. С другой стороны, если используется детектирование огибающей или квадратичное детектирование и некогерентное сложение для формирования величин, по которым принимается решение, расчёт качества декодера значительно более сложен.

Каскадные блоковые коды. Каскадный код состоит из двух отдельных кодов, которые объединяются для образования большего кода. Обычно один из кодов выбирается недвоичным, а второй двоичным. Они соединяются каскадно, как показано на рис. 8.1.19.

Недвоичный (N, K) код образует внешний код, а двоичный – внутренний код. Кодовые слова формируются путём подразделения блока на Kk информационных бита по K группам, называемым символами, причём каждый такой символ состоит из k бит. K символов (с k битами каждый) кодируются в N символов внешним кодом, как это обычно делается при недвоичном кодировании. Внутренний кодер берет каждый k-битовый символ и кодирует его в двоичный блоковый код длины n. Таким образом, мы получаем каскадный блоковый код, имеющий длину Nn бита и содержащий Kk информационных

бита. Это значит, мы создали эквивалентный (Nn, Kk) длинный двоичный код. Биты в каждом кодовом слове передаются по каналу посредствам ФМ или, возможно, ЧМ.

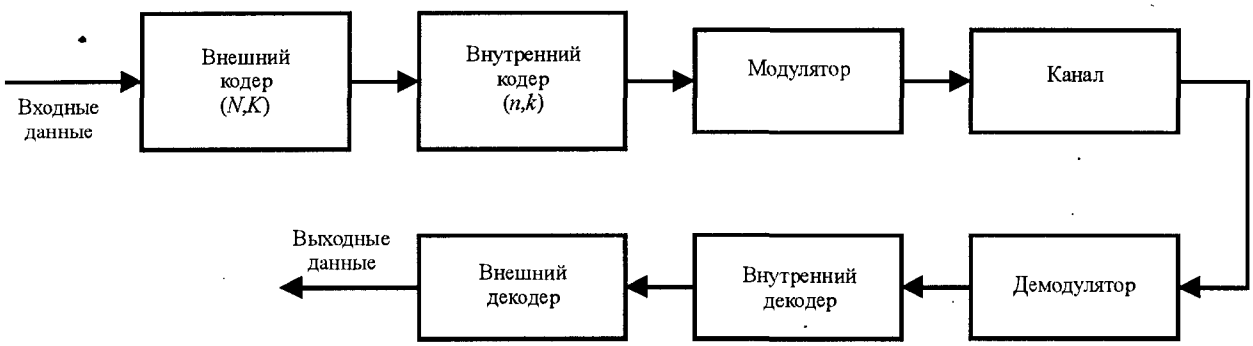


Рис. 8.1.19. Блок-схема системы связи, использующей каскадный код

Также укажем, что минимальное расстояние для каскадного кода равно $d_{\min}D_{\min}$, где $D_{\min}$ – это минимальное расстояние для внешнего кода, а $d_{\min}$ – минимальное расстояние для внутреннего кода. Далее, скорость каскадного кода равна Kk/Nn, что равно произведению скоростей двух кодов.

Декодер жёстких решений для каскадного кода удобно разделить на внутренний декодер и внешний декодер. Внутренний декодер выполняет жёсткое решение по каждой группе из n бита, соответствующие кодовому слову внутреннего кода, и выносит решение о k информационных битах, основываясь на алгоритме максимального правдоподобия (минимума расстояния). Эти k бит представляют один символ внешнего кода. Когда принят блок из N k-битовых символов от внутреннего декодера, внешний декодер принимает жёсткое решение по K k-битовым информационным символам, основываясь на декодирование по правилу максимального правдоподобия.

При каскадном кодировании возможно и декодирование мягких решений. Обычно оно выполняется по внутреннему коду, если он выбран так, что имеет немного кодовых слов, т.е. 2^k не очень велико. Внешний код обычно декодируется посредствам декодера жёстких решений, особенно если длина блока велика и имеется много кодовых слов. С другой стороны, можно достичь достаточный выигрыш в качестве при использовании декодирования мягких решений по внутреннему и внешнему кодам, чтобы оправдать дополнительную сложность декодирования. Этот случай интересен для цифровых систем связи в радиоканалах с замираниями, как мы покажем в главе 14.

Мы завершим этот подраздел следующим примером.

Пример 8.1.14. Предположим, что код Хемминга (7, 4), описанный в примерах 8.1.1 и 8.1.2, используется как внутренний код при каскадном кодировании, причём в качестве внешнего кода используется код Рида–Соломона. Поскольку K=4, выберем длину кода Рида–Соломона $N = 2^4 - 1 = 15$. Число информационных символов K в кодовых словах внешнего кода можно выбрать в области $1 \le K \le 14$ для того, чтобы достичь желательную скорость кода.

8.1.9. Перемежение кодовых символов в каналах с пакетами ошибок

Большинство хорошо известных кодов, которые были разработаны для увеличения надёжности при передачи информации, являются эффективными, когда ошибки, вызванные каналом, статистически независимы. Это случай канала с АБГШ. Однако, имеются каналы, в которых появляются пакеты ошибок. Один пример – это класс каналов,

характеризуемых многопутёвостью и замираниями, которые в подробностях описаны в главе 14. Замирания сигнала, обусловленные меняющимся во времени многопутевым распространением волн, часто вызывают снижение уровня сигнала ниже уровня шума, что приводит к большому числу ошибок. Второй пример – это класс каналов магнитной записи (ленточной или дисковой), в которых дефекты в записывающей среде приводят к пачкам ошибок. Такие группы ошибок обычно не исправляются кодами, оптимально рассчитанными для статистически независимых ошибок.

Значительные работы были выполнены по синтезу кодов, которые способны исправить пакеты ошибок. Вероятно, наиболее известными кодами, исправляющими пакеты ошибок, является подкласс циклических кодов, называемых кодами Файра, в честь П.Файра (1959), который их открыл. Другой класс циклических кодов для исправления пакетов ошибок был позднее открыт Буртоном (1969).

Пакет ошибок длины b определяется как последовательность из b символов (бит) ошибок, первым и последним из которых является «1». Способность кода *исправить пачки ошибок* определяется длиной наиболее короткой пачки ошибок, которую он не может исправить. Относительно просто можно показать, что систематический (n,k) код, который имеет $n-k$ проверочных символов, может корректировать пачки ошибок длиной $b \le \lfloor \frac{1}{2}(n-k) \rfloor$.

Эффективный метод работы в каналах с пачками ошибок заключается в перемежении кодовых посылок таким путём, что канал с пачками ошибок трансформируется в канал, имеющий независимые ошибки. Затем используется код, рассчитанный на независимые ошибки в канале (короткие пакеты).

Блок-схема системы связи, которая использует перемежение символов, показана на рис. 8.1.20.

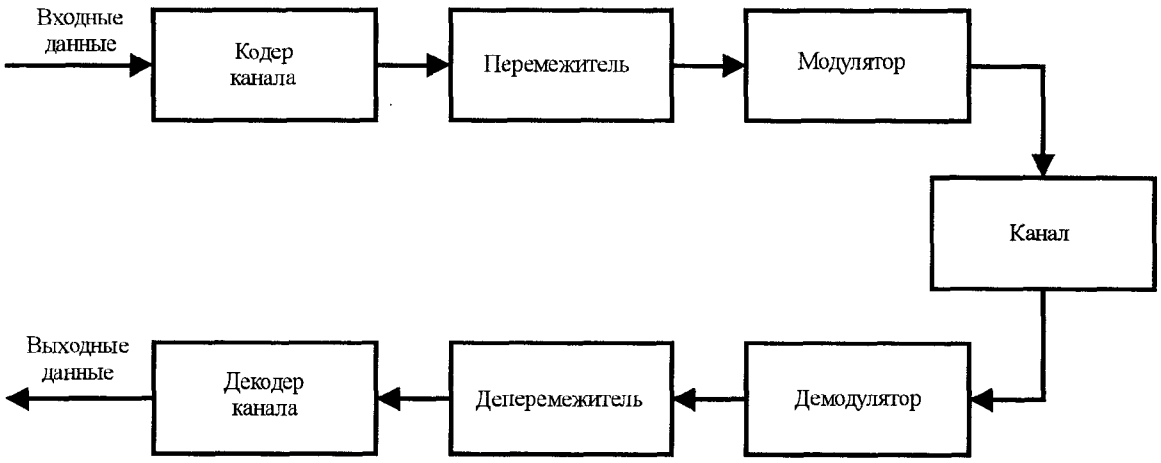


Рис. 8.1.20. Блок-схема системы связи, использующей перемежение в канале с группированием ошибок

Кодированные данные перегруппируются перемежителем и передаются по каналу. На приёме, после (жёстких или мягких решений) демодулятора деперемежитель восстанавливает символы в нужной последовательности и направляет их к декодеру. Как результат перемежения-деперемежения, пачки ошибок рассеиваются во времени так, что ошибки внутри кодовых слов становятся независимыми.

Перемежитель может принять одну из двух форм: блоковая структура или свёрточная структура. *Блоковый перемежитель* формирует кодированные данные в прямоугольный

массив из m строк и n столбцов. Обычно, каждая строка массива состоит из кодового слова длин n. Перемежитель степени m состоит из m столбцов (m кодовых слов), как показано на рис. 8.1.21.

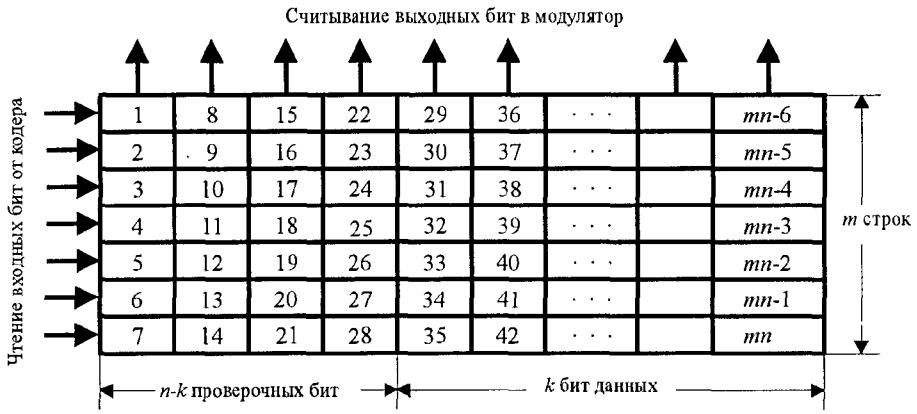


Рис. 8.1.21. Перемежитель кодированных бит

Биты считываются по столбцам и передаются по каналу. В приёмнике деперемежитель располагает данные в тот же прямоугольный формат, но теперь они считываются по строкам, одно кодовое слово за раз. Результат такой перегруппировки данных при передаче по каналу сводится к тому, что пачка ошибок длины $l = mb$ разбивается на m пачек длиной b каждая. Таким образом, код (n,k), который может справляться с пачкой ошибок длины $b \le \lfloor \frac{1}{2}(n-k) \rfloor$ можно соединить с перемежителем степени m для того, чтобы образовывать блоковый код (mn,mk), который может справляться с пачками ошибок длиной $l = mb$.

Свёрточный перемежитель можно использовать вместо блокового перемежителя таким же путём. Свёрточные перемежители лучше согласованы к использованию совместно со свёрточными кодами, которые описываются в следующем разделе. Структура свёрточного перемежителя была описана Рамсеем (1970) и Форни (1971).

8.2. СВЁРТОЧНЫЕ КОДЫ

Свёрточный код создаётся прохождением передаваемой информационной последовательности через линейный сдвиговый регистр с конечным числом состояний. В общем, регистр сдвига состоит из K (k-битовых) ячеек и линейного преобразователя, состоящего из n функциональных генераторов и выполняющего алгебраические функции, как показано на рис. 8.2.1.

Входные данные к кодеру, которые считаются двоичными, продвигаются вдоль регистра сдвига по k бит за раз. Число выходных битов для каждой k-битовой входной последовательности равно n. Следовательно, кодовая скорость, определённая как $R_c = k/n$, согласуется с определением скорости блокового кода. Параметр K называется кодовым ограничением[1] свёрточного кода[2].

---

[1] Чаще всего при k=1 кодовым ограничением называют число K–1 (прп)

[2] Во многих случаях кодовое ограничение кода задается скорее в битах, чем в k-битовых блоках. Следовательно, регистр сдвига можно назвать L-ячеечным регистром сдвига, где $L = Kk$. Более того, в общем случае L может и не быть кратным K.

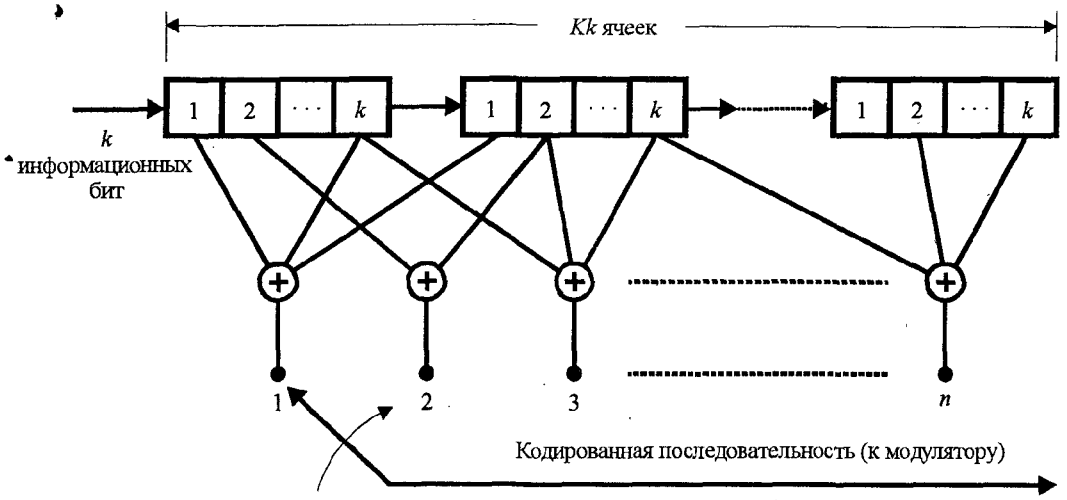


Рис. 8.2.1. Свёрточный кодер

Один метод для описания свёрточного кода сводится к заданию его порождающей матрицы, так же, как мы это делали для блоковых кодов. В общем, порождающая матрица для свёрточного кода полубесконечная, поскольку входная последовательность полубесконечная. Можно использовать эквивалентное представление кода, в котором мы определяем набор из n векторов, один вектор для каждого из n сумматоров по *mod* 2. Каждый вектор имеет Kk измерений и содержит в себе информацию о соединениях кодера с сумматорами по mod 2. «1» в i-ой позиции вектора указывает на то, что соответствующая ячейка регистра сдвига подсоединена к сумматору по *mod* 2, а 0 в данной позиции указывает на то, что такого соединения нет.

Для конкретности рассмотрим двоичный свёрточный кодер с кодовым ограничением $K = 3$, $k = 1$ и $n = 3$, показанный на рис. 8.2.2.

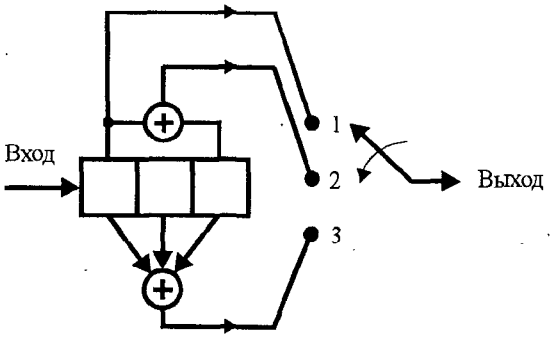


Рис. 8.2.2. Свёрточный кодер с K=3, k=1, n=3

Считается, что первоначально все ячейки регистра сдвига находятся в нулевом состоянии. Допустим, что первый входной бит «1». Он без задержки появляется на выходе первой (левой) ячейки регистра и, соответственно, на всех трёх входах выходного ключа (мультиплексора). Ключ поочерёдно выдаёт содержимое входов, и выходная последовательность из 3 бит 111. Допустим, что второй входной бит «0». Он записывается в первую ячейку регистра, проталкивает предыдущий бит («1») во вторую ячейку – и на входах мультиплексора (сверху вниз) появляются 0, 0 и 1. Тогда вторая выходная последовательность 001. Если третий входной бит 1, выходная последовательность 100 – и так далее. Таким образом, в ответ на каждый входной бит ($k = 1$) свёрточный кодер

откликается тремя битами, по числу функциональных генераторов ($n = 3$). Пронумеруем функциональные генераторы и их выходы в соответствии с порядком считывания мультиплексором (1, 2 и 3). Затем, поскольку с первым функциональным генератором соединена только первая ячейка (сумматор по $mod\,2$ не нужен) генератор можно отобразить вектором

$$\mathbf{g}_1 = [100].$$

Второй функциональный генератор соединен с ячейками 1 и 3. Следовательно,

$$\mathbf{g}_2 = [101].$$

Наконец,

$$\mathbf{g}_3 = [111].$$

Генераторы (порождающие полиномы) для этого кода обычно дают в восьмеричной форме как **g**=(4, 5, 7). Мы заключаем, что, когда $k = 1$, требуются n генераторов, каждый размерностью K, чтобы описать кодер.

В общем случае при $k > 1$ и кодовом ограничении K эти n генераторов отображаются Kk -мерными векторами, как оговорено выше. Следующий пример иллюстрирует случай, когда $k = 2$ и $n = 3$.

Пример 8.2.1. Рассмотрим свёрточный кодер со скоростью кода 2/3, показанный на рис. 8.2.3. В этом кодере каждый раз два бита поступают на вход регистров сдвига, а на выходе генерируется три бита.

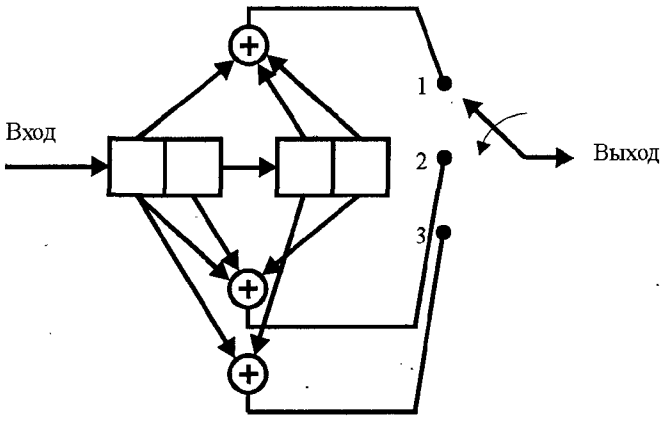


Рис. 8.2.3. Свёрточный кодер с K=2, k=2, n=3

Генераторы определяются векторами

$\mathbf{g}_1 = [1011]$, $\mathbf{g}_2 = [1101]$, $\mathbf{g}_3 = [1010]$.

В восьмеричной форме **g**=(13, 15, 12)

Имеются три альтернативных метода, которые часто используются для описания свёрточного кода. Это древовидная диаграмма, решётчатая диаграмма и диаграмма состояний. Для примера, древовидная диаграмма для свёрточного кодера, показанного на рис. 8.2.2, иллюстрируется на рис. 8.2.4.

Предположим, что кодер находится первоначально в нулевом состоянии (во всех ячейках нули). Диаграмма показывает, что, если первый вход 0 – выходная последовательность 000, а если первый вход 1 – выходная последовательность 111. Теперь, если первый вход 1, а второй 0 – второй набор выходных битов 001. Продвигаясь по дереву видим, что если третий входной бит 0, тогда выходной 011, если же третий выходной бит 1, то выход 100. Видим, что частная последовательность обуславливает выбор узла дерева, а правило движения по ветвям дерева такое – надо двигаться к верхней

ветви, если следующий бит 0 и к нижней, если следующий бит 1. Таким образом, траектория частного пути по дереву определяется входной последовательностью.

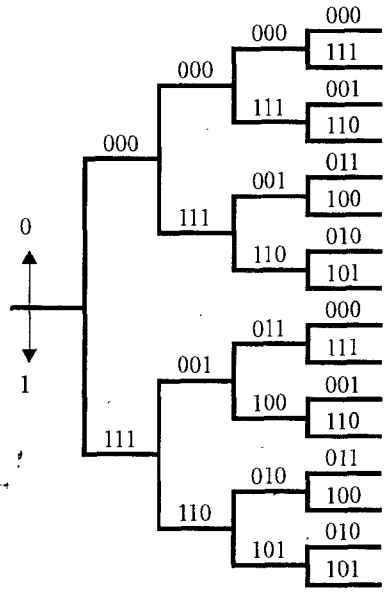


Рис. 8.2.4. Древовидная диаграмма для свёрточного кода, имеющего скорость 1/3, K=3

Внимательное наблюдение за деревом, показанном на рис. 8.2.4, обнаруживает, что структура повторяет себя после третьего такта. Правый столбец выходных «троек» бит распадается на две одинаковые совокупности по 8 «троек». Это поведение согласуется с тем фактом, что кодовое ограничение $K = 3$. Это значит, трёхбитовые выходные последовательности на каждом такте определяются входным битом и двумя предыдущими входными битами, т.е. двумя битами, содержащимися в первых двух ячейках регистра сдвига. Бит в последней ячейке регистра сдвига перемещается направо (покидает регистр) и не влияет больше на выход. Так что мы можем сказать, что трёхбитовая последовательность выхода для каждого входного бита определяется входным битом и четырьмя возможными состояниями регистра сдвига, обозначенными a=00, b=01, c=10, d=11. Если пометить узлы дерева этими же метками, найдем, что на третьем такте имеются два узла с пометкой a, два с пометкой b, два с пометкой c и два с пометкой d. Теперь видим, что все ветви, исходящие из двух узлов с одинаковой меткой (одинаковым состоянием) являются идентичными в том смысле, что они генерируют одинаковые выходные последовательности. Это означает, что два узла, имеющие одинаковую метку, можно слить. Если мы это сделаем в дереве, показанном на рис. 8.2.4 , мы получим другую диаграмму, которая более компактна, именно получим *решётку*. Для примера решётчатая диаграмма для свёрточного кодера рис. 8.2.2 показана на рис. 8.2.5.

Чтобы изучить эту диаграмму, договоримся о том, что сплошные линии означают выходы, генеририруемые входом 0, а пунктирные – выходы, генеририруемые входом 1. В примере, который мы рассматриваем, видим, что после начального состояния решётка содержит четыре узла на каждом шаге, соответствующие четырем состояниям регистра сдвига a, b, c и d. После второго шага каждый узел в решётке имеет два входных пути и два выходных. Из выходных путей один соответствует входу 0, а второй – входу 1.

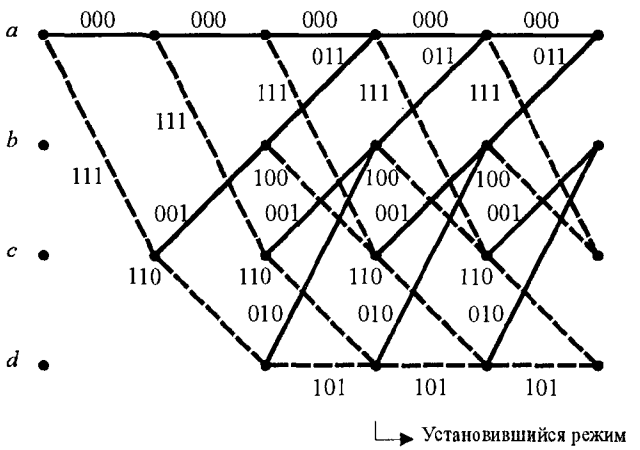


Рис. 8.2.5. Решётчатая диаграмма для свёрточного кода, имеющего скорость 1/3, K=3

Поскольку выход кодера определяется входом и состоянием кодера, ещё более компактной, чем решётка, является диаграмма состояний. Диаграмма состояний – это просто граф возможных состояний кодера и возможных переходов из одного состояния в другое. Для примера на рис. 8.2.6 показана диаграмма состояний для кодера, показанного на рис. 8.2.2.

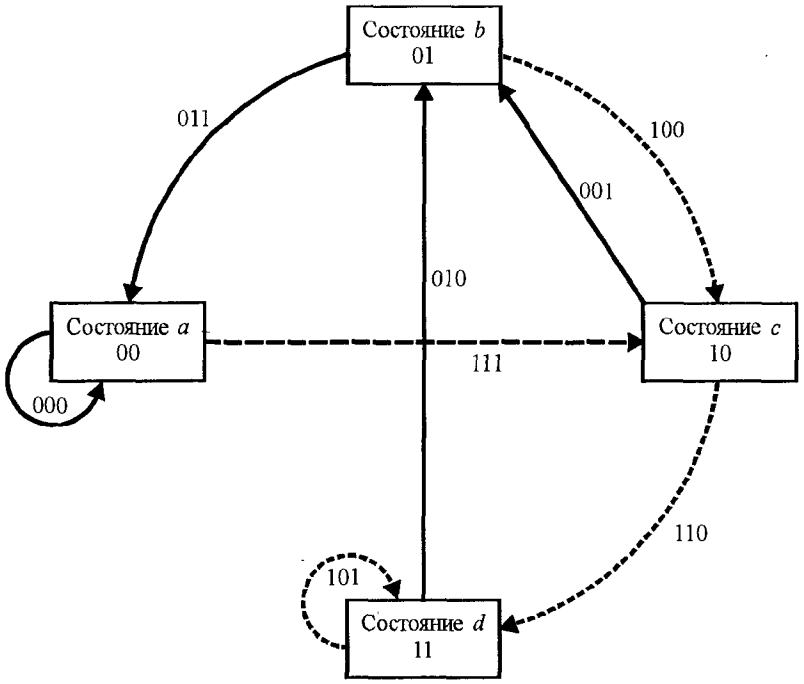


Рис. 8.2.6. Диаграмма состояний для свёрточного кода, имеющего скорость 1/3, K=3

Эта диаграмма показывает, что возможные переходы таковы

$$a \xrightarrow{0} a, a \xrightarrow{1} c, b \xrightarrow{0} a, b \xrightarrow{1} c, c \xrightarrow{0} b, c \xrightarrow{1} d, d \xrightarrow{0} b, d \xrightarrow{1} d,$$

где $\alpha \xrightarrow{1} \beta$ означает переход из состояния α в β, когда входной бит 1. Три бита, показанных далее на каждой ветви диаграммы состояний, представляют выходные биты. Пунктирная линия на графе означает, что входной бит 1, в то время как сплошная линия указывает, что входной бит 0.

Пример 8.2.2. Рассмотрим свёрточный код со скоростью 2/3, k=2, описанный в примере 8.2.1 и показанный на рис. 8.2.3. Первые два входных бита могут быть 00, 01, 10 или 11. Соответствующие выходные биты – 000, 010, 111, 101. Когда следующая пара входных битов входит в кодер, первая пара передвигается в следующую ячейку. Соответствующие выходные биты зависят от пары битов, переместившихся во вторую ячейку и новой пары входных битов. Следовательно, древовидная диаграмма для этого кода, показанная на рис. 8.2.7, имеет четыре ветви на узел, соответствующие четырём возможным парам входных символов. Поскольку кодовое ограничение кодера $K = 2$, дерево начинает повторяться после второго шага. Как показано на рис. 8.2.7, все ветви, исходящие из узла, обозначенного a (состояния a) дают идентичные выходы. Путём слияния узлов, имеющих одинаковое название, мы получаем решётку, показанную на рис. 8.2.8. Итоговая диаграмма состояний для этого кодера показана на рис. 8.2.9.

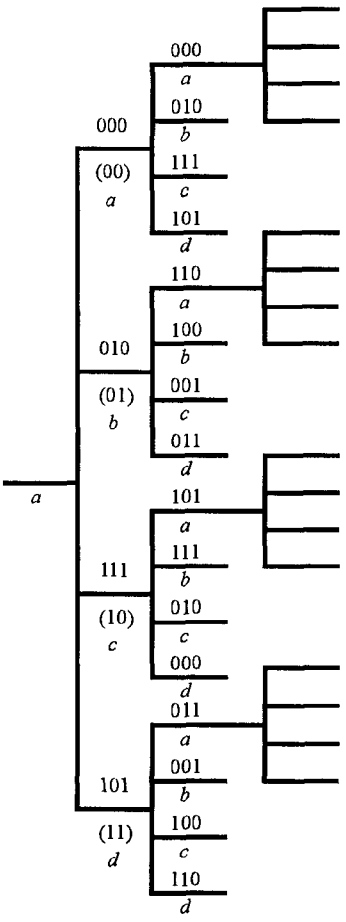


Рис. 8.2.7. Древовидная диаграмма для свёрточного кода, имеющего параметры K=2, k=2, n=3

Для обобщения отметим, что свёрточный код со скоростью k/n и кодовым ограничением K характеризуется 2^k ветвями, уходящими от каждого узла на древовидной диаграмме. Решётка и диаграмма состояний имеют (каждая из них) $2^{k(K-1)}$ возможных состояний. Имеются 2^k ветвей, входящих в каждое состояние, и 2^k ветвей, покидающих каждое состояние (для решётки и дерева это верно после наступления установившегося режима).

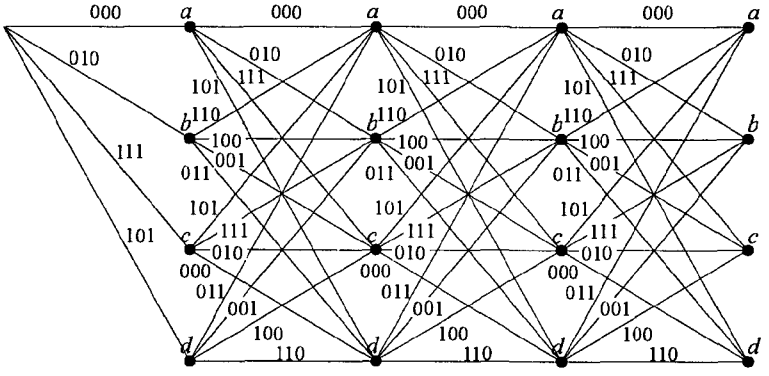


Рис. 8.2.8. Решётчатая диаграмма для свёрточного кода, имеющего параметры K=2, k=2, n=3

Три типа диаграмм, описанных выше, используются также для представления недвоичных свёрточных кодов. Если число символов в алфавите равно $q = 2^k$, $k > 1$, результирующий недвоичный код можно также представить как эквивалентный двоичный код. Следующий пример рассматривает свёрточный код этого типа.

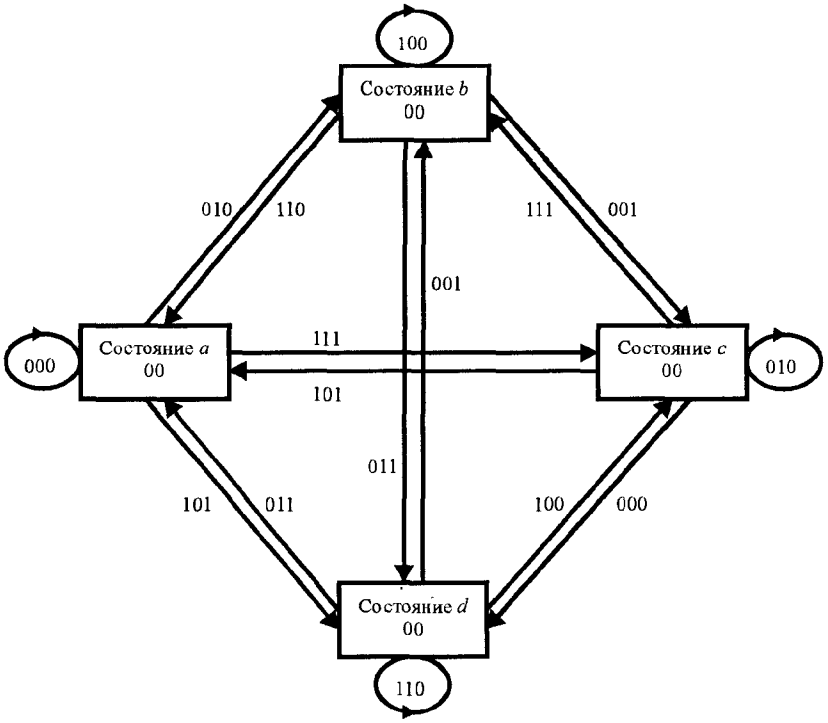


Рис. 8.2.9. Диаграмма состояний для свёрточного кода, имеющего параметры K=2, k=2, n=3

Пример 8.2.3. Рассмотрим свёрточный код, генерируемый кодером, показанном на рис. 8.2.10. Этот код можно описать как двоичный свёрточный код с параметрами $K = 2$, $k = 2$, $n = 4$, $Rc = 1/2$ и с генераторами

$$\mathbf{g}_1 = [1010], \quad \mathbf{g}_2 = [0101], \quad \mathbf{g}_3 = [1110], \quad \mathbf{g}_4 = [1001].$$

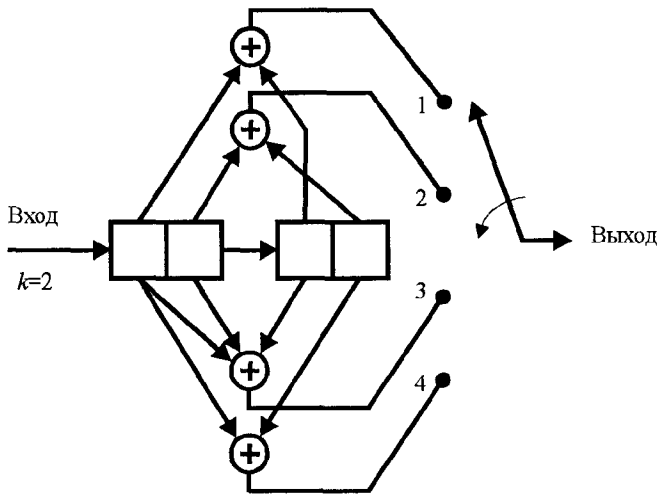


Рис. 8.2.10. Свёрточный кодер с K=2, k=2, n=3

За исключением различия в скорости, этот код похож по форме на свёрточный код со скоростью 2/3, k=2, рассмотренный в примере 8.2.1.

Альтернативно, код, генерируемый кодером рис. 8.2.10, можно описать как не двоичный (q=4) код с одним четырехпозиционным символом на входе и двумя четырехпозиционных символов на выходе. Действительно, если выходные символы кодера трактуются модулятором и демодулятором как q-ичные (q=4), которые передаются по каналу посредством некоторой M-ичный (M=4) модуляции, код можно соответственно рассмотреть как недвоичный.

В любом случае, древовидные, решётчатые диаграммы и диаграммы состояний независимы от того, как мы смотрим на код. Это значит, что этот частный код, характеризуется деревом с четырьмя ветвями, исходящими от каждого узла, или решёткой с четырьмя возможными состояниями и четырьмя ветвями, входящими и покидающими каждое состояние, или, что эквивалентно, посредством диаграммы состояний, имеющей те же параметры, что и решётка.

8.2.1. Передаточная функция свёрточного кода

Дистанционные свойства и характеристики качества (вероятность ошибки) свёрточного кода можно получить из его диаграммы состояний. Поскольку свёрточный код линейный, набор расстояний Хемминга между кодовыми последовательностями, генерируемыми на определённом шаге дерева и последовательностью из одних нулей, такой же, как набор расстояний кодовых последовательностей по отношению к другой кодовой последовательности. Следовательно, мы предположим, без потери общности, что входом кодера является последовательность из одних нулей.

Диаграмма состояний, показанная на рис. 8.2.6, будет использована для демонстрации метода получения дистанционных свойств свёрточного кода. Сначала мы пометим ветви на диаграмме состояний как $D^0 = 1$, D^1, D^2 или D^3, где показатель у D означает расстояние Хемминга между выходной битовой последовательностью, соответствующей ветви, и выходной последовательностью, соответствующей ветви из одних нулей. Собственную петлю у узла a можно исключить, поскольку она ничего не вносит в дистанционные свойства кодовой последовательности относительно кодовой последовательности из одних нулей. Дальше, узел a расщепляется на два узла, один из них представляет вход, а другой выход диаграммы состояний. Рис. 8.2.11 иллюстрирует результирующую диаграмму.

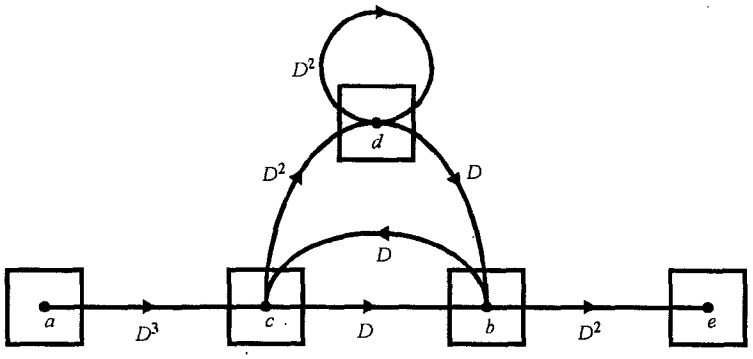


Рис. 8.2.11. Диаграмма состояний для свёрточного кода, имеющего скорость 1/3, K=3

Мы используем эту диаграмму, которая теперь состоит из пяти узлов, поскольку узел a был расщеплен на два узла, для написания четырех уравнений состояния:

$$X_c = D^3 X_a + DX_b, \quad X_b = DX_c + DX_d,$$
$$X_d = D^2 X_c + D^2 X_d, \quad X_e = D^2 X_b. \tag{8.2.1}$$

Передаточная функция кода определяется как $T(D) = X_e / X_a$. Решив уравнения состояния, данные выше, мы получим

$$T(D) = \frac{D^6}{1-2D^2} = D^6 + 2D^8 + 4D^{10} + 8D^{12} + \ldots = \sum_{d=6}^{\infty} a_d D^d, \tag{8.2.2}$$

где, по определению,

$$a_d = \begin{cases} 2^{(d-6)/2} & (\text{четные} \quad d) \\ 0 & (\text{нечетные } d). \end{cases} \tag{8.2.3}$$

Передаточная функция этого кода указывает на то, что имеется единственный путь с расстоянием Хемминга d=6 от пути из одних нулей, который сливается с путём из одних нулей при данном узле. Из диаграммы состояний, показанной на рис. 8.2.6, или решётчатой диаграммы, показанной на рис. 8.2.5, видно, что путь с d=6 это *acbe*. Нет других путей из узла *a* до узла *e*, имеющих расстояние d=6. Второе слагаемое в (8.2.2) указывает на то, что есть два пути от уэла *a* до узла *e*, имеющих расстояние d=8. Снова из диаграммы состояний или решётки мы видим, что этими путями являются *acdbe* и *acbcbe*. Третье слагаемое в (8.1.2) указывает, что есть четыре пути с расстоянием d=10 и так далее. Таким образом, передаточная функция даёт нам дистанционные свойства свёрточного кода. Минимальное расстояние кода *называется минимальным свободным расстоянием* и обозначается $d_{\text{св}}$. В нашем примере $d_{\text{св}} = 6$.

Передаточную функцию можно использовать для получения более детальной информации, чем только расстояния различных путей. Введём множитель N для всех переходов ветвей, вызванных входным битом 1. Тогда, поскольку каждая ветвь пересекается, совокупный показатель N увеличивается на единицу только тогда, когда переход ветви обусловлен входным битом 1. Далее мы вводим множитель J для каждой ветви диаграммы состояний так, что показатель J будет служить счётной величиной, указывающей число ветвей для любого данного пути от узла *a* к узлу *e*. Для свёрточного кода со скоростью 1/3 в нашем примере диаграмма состояний, которая объединяет суммируемые множители J и N, показана на рис. 8.2.12.

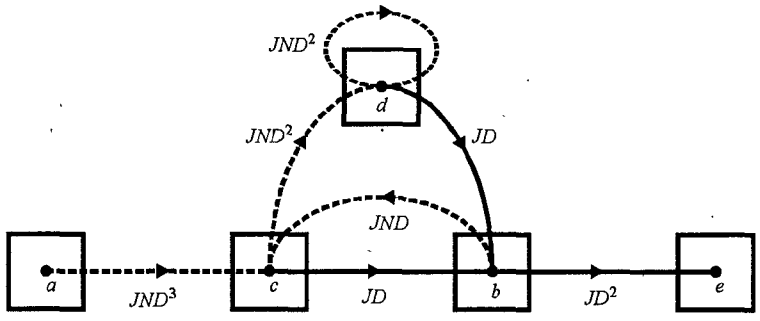


Рис. 8.2.12. Диаграмма состояний для свёрточного кода, имеющего скорость 1/3, K=3

Уравнения для диаграммы состояний, показанной на рис. 8.2.12, таковы:

$$X_c = JND^3 X_a + JNDX_b, \quad X_b = JDX_c + JDX_d,$$
$$X_d = JND^2 X_c + JND^2 X_d, \quad X_e = JD^2 X_b. \tag{8.2.4}$$

Решая эти уравнения, получаем передаточную функцию

$$T(D,N,J)=\frac{J^3ND^6}{1-JND^2(1+J)}=J^3ND^6+J^4N^2D^8+J^5N^2D^8+ \\ +J^5N^3D^{10}+2J^6N^3D^{10}+J^7N^3D^{10}+\ldots \qquad (8.2.5)$$

Эта форма передаточной функции даёт свойства всех путей свёрточного кода. Это значит, что первое слагаемое в выражении $T(D,N,J)$ указывает на то, что путь с расстоянием $d=6$ имеет длину 3 и что из трёх информационных битов один «1». Второе и третье слагаемые в выражении $T(D,N,J)$ указывают на то, что из двух слагаемых с расстоянием $d=8$ одно длиной 4, а второе длиной 5. Два из четырёх информационных бит в пути длиной 4 и два из пяти информационных бит в пути длиной 5 являются «1».

Таким образом, показатель множителя J указывает длину пути, который сливается первый раз с путём из одних нулей, показатель множителя N указывает число «1» в информационной последовательности для этого пути, а показатель D указывает расстояние от последовательности кодированных битов этого пути от последовательности с одними нулями.

Множитель J особенно важен, если мы передаем последовательность конечной длины, скажем, m битов. В этом случае свёрточный код повторяется после m узлов или m ветвей. Это подразумевает, что передаточная функция для усечённого кода получается при усечении $T(D,N,J)$ по слагаемому J^m. С другой стороны, если мы передаем экстремально длинную последовательность, т.е. существенно неограниченную по длине последовательность, мы хотим подавить зависимость $T(D,N,J)$ от параметра J. Это легко выполнить, положив J=1. Так для примера, данного выше, мы имеем

$$T(D,N,1)=T(D,N)=\frac{ND^6}{1-ND^2}=ND^6+2N^2D^8+4N^3D^{10}+\ldots=\sum_{d=6}^{\infty}a_dN^{(d-4)/2}D^d, \qquad (8.2.6)$$

где коэффициенты $\{a_d\}$ определены (8.2.3).

Процедуру, которую мы описали в общих чертах выше для определения передаточной функции двоичного свёрточного кода, легко расширить на недвоичные коды. В следующем примере мы определим передаточную функцию для недвоичного свёрточного кода, ранее введённого в примере 8.2.3.

Пример 8.2.4. Свёрточный код, показанный на рис. 8.2.10, имеет параметры $K=2$, $k=2$, $n=4$.

Допустим, что мы трактуем код как недвоичный. Так, вход кодера и выход кодера можно трактовать как четверичные символы. В частности, если мы трактуем вход и выход как чётверичные символы 00, 01, 10 и 11, то расстояние, измеряемое в символах между последовательностями 0111 и 0000, равно 2. Далее предположим, что входной символ 00 декодируется как символ 11, тогда мы имеем одну ошибку в символе. Это соглашение, применённое к свёрточному коду, показанному на рис. 8.2.10, приводит к диаграмме состояний, иллюстрируемой рис. 8.2.13.

Из неё мы получаем уравнения состояний:

$$\begin{aligned} X_b &= NJD^2X_a+NJDX_b+NJDX_c+NJD^2X_d, \\ X_c &= NJD^2X_a+NJD^2X_b+NJDX_c+NJDX_d, \\ X_d &= NJD^2X_a+NJDX_b+NJD^2X_c+NJDX_d, \\ X_e &= JD^2(X_b+X_c+X_d). \end{aligned} \qquad (8.2.7)$$

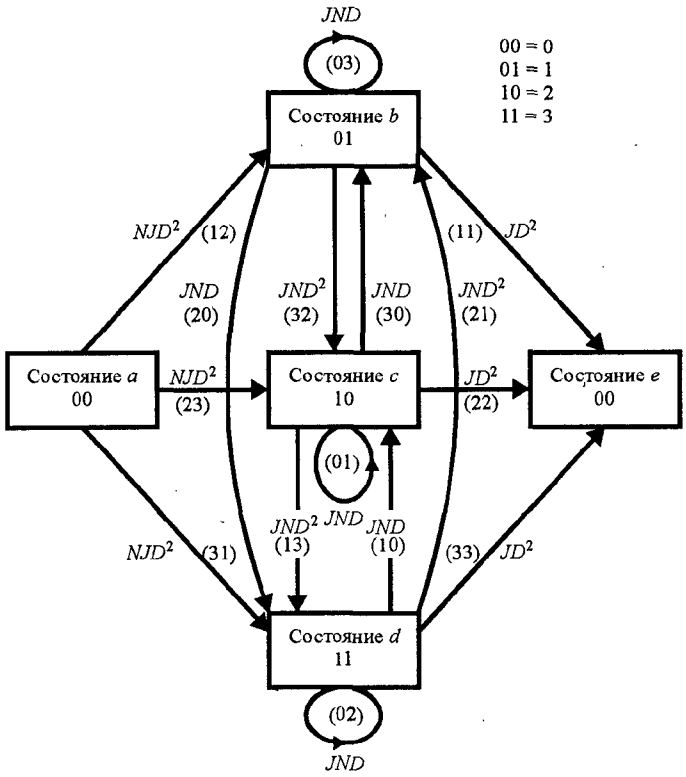


Рис. 8.2.13. Диаграмма состояний для недвоичного свёрточного кода, имеющего параметры K=2, k=2, скорость 1/2

Решение этих уравнений приводит к передаточной функции

$$T(D,N,J)=\frac{3NJ^2D^4}{1-2JD-NJD^2}. \qquad (8.2.8)$$

Это выражение для передаточной функции особенно применимо тогда, когда четверичные символы на выходе кодера отображаются в соответствующий ансамбль четверичных сигналов $s_m(t)$, $m=1, 2, 3, 4$, например, четырьмя ортогональными сигналами. Таким образом здесь имеется взаимно-однозначное соответствие между кодовыми символами и сигналами.

Альтернативно, для примера, выход кодера можно передать как последовательность двоичных символов посредством двоичной ФМ. В таком случае следует измерять расстояние в битах. Если использовать такое соглашение, диаграмма состояний имеет вид рис. 8.2.14.

Решение уравнений состояний, полученное из этой диаграммы состояний, приводит к передаточной функции, которая отличается от той, которая дана (8.2.8).

Некоторое свёрточные коды проявляют характерное поведение, называемое *катастрофическим размножением ошибок.*

Когда код, который имеет такие характеристики, используется в двоичном симметричном канале, возможно, при ограниченном числе ошибок в канале, неограниченное число ошибок декодирования. Такой код можно идентифицировать из его диаграммы состояний. Она может содержать путь с нулевым расстоянием (путь с множителем $D^0=1$) от некоторого ненулевого состояния обратно в то же самое состояние. Это означает, что может образоваться петля вокруг этого пути с нулевым расстоянием неограниченное число раз без увеличения расстояния относительно пути с одними нулями. Но если эта собственная петля соответствует передаче 1, декодер будет

делать неограниченное число ошибок. Поскольку такие коды легко распознать, их легко избежать на практике.

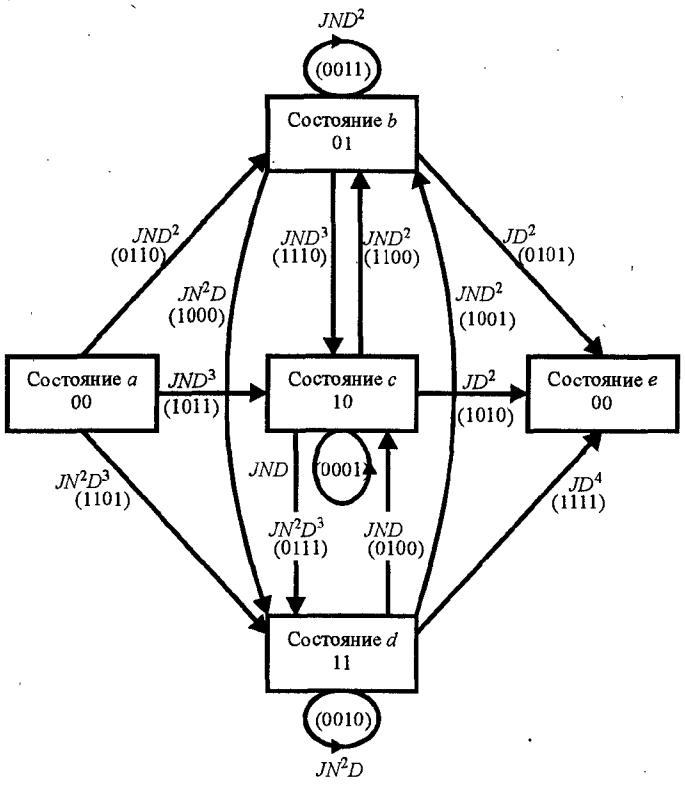


Рис. 8.2.14. Диаграмма состояний для недвоичного свёрточного кодера, имеющего параметры K=2, k=2, скорость 1/2, выходы которого интерпретируются как двоичные последовательности

8.2.2. Оптимальное декодирование для свёрточных кодов – алгоритм Витерби

При декодировании блокового кода в канале без памяти, мы вычисляем расстояние (расстояние Хемминга при декодировании жёстких решений и расстояние Евклида при декодировании мягких решений) между принимаемым кодовыми словами и 2^k возможными к передаче кодовыми словами. Затем мы выбираем кодовое слово, которое наиболее близко по расстоянию к принятому кодовому слову. Это правило решения, которое требует вычисления 2^k метрик, оптимально в том смысле, что оно приводит к минимуму средней вероятности ошибки в двоичном симметричном канале с АБГШ и $p < \frac{1}{2}$.

В отличие от блокового кода, который имеет фиксированную длину n, свёрточный код порождается устройством с ограниченным числом состояний. Как следствие, оптимальный декодер является максимально правдоподобным последовательным оценивателем (МППО) вида, описанного в разделе 5.1.4 для сигналов с памятью, таких как ДБНП и МНФ. Поэтому оптимальное декодирование свёрточных кодов включает поиск по решётке наиболее правдоподобной последовательности. В зависимости от того, формирует ли детектор, за которым следует декодер, жёсткие или мягкие решения, соответствующие метрики при поиске по решётке могут быть или метриками Хемминга или метриками Евклида, соответственно. Более подробно мы это рассмотрим ниже, используя решётку рис. 8.2.5 для свёрточного кода, показанного на рис. 8.2.2.

Рассмотрим два пути на решётке, которые начинаются в начальном состоянии a и сливаются в состоянии a после трёх переходов (трёх ветвей), которые соответствуют двум информационным последовательностям 000 и 100 и передаваемым последовательностям 000 000 000 и 111 001 011 соответственно. Обозначим переданные биты через $\{c_{jm},\ j=1,2,3;\ m=1,2,3\}$, где индекс j указывает на j-ю ветвь, а индекс m указывает на m-й бит в этой ветви. Соответственно определим $\{r_{jm},\ j=1,2,3;\ m=1,2,3\}$ как выход демодулятора. Если детектор формирует жёсткие решения, его выходом для каждого переданного бита является или 0, или 1. С другой стороны, если используется декодирование мягких решений, а кодированная последовательность передаётся двоичной когерентной ФМ, то входные величины для декодера определяются так:

$$r_{jm}=\sqrt{\mathscr{E}_c}\left(2c_{jm}-1\right)+n_{jm}, \tag{8.2.9}$$

где n_{jm} представляет собственный шум, а $\mathscr{E}_c$ – энергия сигнала каждого переданного кодового символа (бита).

Для j-й ветви и i-го пути по решётке метрики определяются как логарифм совместной плотности вероятности последовательности $\{r_{jm},\ m=1,2,3\}$ при условии передачи последовательности $\{c_{jm}^{(i)},\ m=1,2,3\}$ для i-го пути.

То есть

$$\mu_j^{(i)}=\log P\left(\mathbf{Y}_j\middle|\mathbf{C}_j^{(i)}\right),\qquad j=1,2,3,\ldots \tag{8.2.10}$$

Далее, метрика i-го пути решётки, содержащего B ветвей, определяется так:

$$PM^{(i)}=\sum_{j=1}^{B}\mu_j^{(i)}. \tag{8.2.11}$$

Правило для решения между двумя путями по решётке сводится к выбору того, у которого больше метрика. Это правило максимизирует вероятность правильного решения или, что эквивалентно, минимизирует вероятность ошибки в информационной последовательности. Например, предположим, что демодулятором формируются жёсткие решения при принимаемой последовательности {101 000 100}. Пусть i=0 означает путь с тремя ветвями из одних нулей, а i=1 означает второй путь с тремя ветвями, который начинается в начальном состоянии a и сливается с путём из одних нулей в состоянии a после трёх переходов. Метрики для этих двух путей таковы

$$\begin{aligned}PM^{(0)}&=6\log(1-p)+3\log p,\\ PM^{(1)}&=4\log(1-p)+5\log p,\end{aligned} \tag{8.2.12}$$

где p – вероятность ошибочного приёма бита. Предположив, что $p<\frac{1}{2}$, находим, что метрика $PM^{(0)}$ больше, чем метрика $PM^{(1)}$. Этот результат согласуется с наблюдением, что путь из одних нулей (i=0) имеет расстояние Хемминга d=3 от принимаемой последовательности, в то время как путь с i=1 имеет расстояние Хемминга d=5 от принимаемого пути. Таким образом, расстояние Хемминга является эквивалентной метрикой для декодирования с жёстким решением.

Аналогично предположим, что используется декодирование мягких решений, а канал прибавляет к сигналу АБГШ. Тогда выход демодулятора описывается статистически через условную ФПВ

$$p\left(r_{pn}\middle|c_{pn}^{(i)}\right)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{\left[r_{pn}-\sqrt{\mathscr{E}_c}\left(2c_{pn}^{(i)}-1\right)\right]^2}{2\sigma^2}\right\}, \tag{8.2.13}$$

где $\sigma^2 = \frac{1}{2}N_0$ – дисперсия аддитивного гауссовского шума. Если мы пренебрегаем слагаемыми, общими для всех метрик ветвей, метрику j-й ветви i-го пути можно выразить так:

$$\mu_j^{(i)} = \sum_{m=1}^{n} r_{pn}\left(2c_{jm}^{(i)} - 1\right), \tag{8.2.14}$$

где в нашем примере n=3. Таким образом, корреляционные метрики для двух путей при оговоренных условиях

$$\begin{aligned} CM^{(0)} &= \sum_{j=1}^{3}\sum_{m=1}^{3} r_{pn}\left(2c_{jm}^{(0)} - 1\right), \\ CM^{(1)} &= \sum_{j=1}^{3}\sum_{m=1}^{3} r_{pn}\left(2c_{jm}^{(1)} - 1\right). \end{aligned} \tag{8.2.15}$$

Имея метрики ветвей и метрики путей, рассчитанные декодером, мы теперь рассмотрим использование алгоритма Витерби для оптимального декодирования информационной последовательности при свёрточном кодировании. Мы рассмотрим два пути, описанные выше, которые сливаются в состоянии *a* после трёх переходов. Заметим, что какой-либо частный путь по решётке, который ответвляется от этого узла, будет суммировать идентичные слагаемые в метриках путей $CM^{(0)}$ и $CM^{(1)}$. Как следствие, если $CM^{(0)} > CM^{(1)}$, у сливающегося узла *a* после трёх переходов, $CM^{(0)}$ будет продолжать быть больше $CM^{(1)}$ для любого пути, который ответвляется от узла *a*. Это значит, что путь, соответствующий $CM^{(1)}$, можно исключить из дальнейшего решения. Путь, соответствующий метрике $CM^{(0)}$, является *выжившим*. Аналогично, один из двух путей, которые сливаются в состоянии *b*, может быть исключен на основе двух соответствующих метрик. Эта процедура повторяется в состоянии *c* и состоянии *d*. Как результат, после трёх первых переходов имеются четыре выживших пути, один кончающийся на каждом состоянии, и соответствующие метрики для каждого выжившего пути. Эта процедура повторяется на каждом шаге решётки по мере того, как принимаются новые сигналы в последующих временных интервалах.

В общем, когда декодируется двоичный свёрточный код с $k = 1$ и кодовым ограничением K посредством алгоритма Витерби, имеются 2^{k-1} состояний. Следовательно, имеются 2^{k-1} выживших путей на каждом шаге и 2^{k-1} метрик, по одной для каждого выжившего пути. Далее, двоичный свёрточный код с k входными битами, которые сдвигаются по кодеру, содержащему K (k-битовых) ячеек регистров сдвига, генерирует решётку, которая имеет $2^{k(K-1)}$ состояний. Следовательно, декодирование такого кода посредством алгоритма Витерби требует сохранить следы $2^{k(K-1)}$ выживших путей и $2^{k(K-1)}$ метрик. На каждом шаге решётки имеются 2^k путей, которые сливаются в каждый узел. Поскольку каждый путь, который сходится в общей узел, требует вычисления метрик, имеются 2^k метрик, вычисленных на каждом узле. Из 2^k путей, которые сливаются в каждом узле, только один выживет, и это наиболее вероятный (с минимальным расстоянием) путь. Таким образом, число вычислений при декодировании, выполняемых на каждом шаге, возрастает экспоненциально с k и K. Экспоненциальный рост вычислительных затрат ограничивает использование алгоритма Витерби только для относительно малых значений K и k.

Задержка при декодировании при декодировании длинных информационных последовательностей, которые были кодированы свёрточным кодом, обычно слишком велика для большинства практических применений. Более того, память, требуемая для

хранения всей длины выживших последовательностей, велика и дорогая. Как указано в разделе 5.1.4, решение этой проблемы заключается в модификации алгоритма Витерби таким образом, чтобы установить фиксированную задержку декодирования без существенной потери качества оптимальности алгоритма. Напомним, что модификация сводится к тому, чтобы удержать в заданное время t только самые последние δ декодированных информационных бит (символов) в каждой выжившей последовательности. По мере приёма новых информационных битов (символов) оптимальное решение принимается относительно бита (символа), принятого на δ ветвей раньше по решётке, путём сравнения метрик выживших последовательностей и принятия решение о бите в последовательности, имеющей наибольшую метрику. Если δ выбран достаточно большим, все выжившие пути будут содержать одинаковый декодируемый бит (символ), принятый на δ ветвей раньше. Это значит, что с большой вероятностью, все выжившие последовательности к моменту t выходят из одного и того же узла в момент времени $t-\delta$. Экспериментально показано (компьютерное моделирование), что задержка $\delta \geq 5K$ обеспечивает пренебрежимо малое ухудшение качества относительно оптимального алгоритма Витерби.[1]

8.2.3. Вероятность ошибки при декодирования мягких решений

Тема этого подраздела – качество алгоритма Витерби в канале с АБГШ при декодировании мягких решений.

При расчёте вероятности ошибки для свёрточных кодов для упрощения расчётов используются линейные свойства этого класса кодов. Это значит, что мы предполагаем, что передаётся последовательность из одних нулей и мы определяем вероятность ошибки при решении в пользу другой последовательности. Считается, что кодированные двоичные символы в j-й ветви свёрточного кода, обозначенные как $\{c_{jm},\ m = 1, 2, \ldots, n\}$ и определённые в разделе 8.2.2, передаются по каналу двоичной ФМ (или четырёхпозиционной ФМ) и детектируется когерентно в демодуляторе. На выходе демодулятора, который соединён со входом декодера Витерби, образуется последовательность $\{r_{jm},\ m = 1, 2, \ldots, n;\ j = 1, 2, \ldots\}$, где r_{jm} определено (8.2.9).

Декодер Витерби мягких решений формирует метрики ветвей, определённые (8.2.14), и по ним рассчитываются метрики путей

$$CM^{(i)} = \sum_{j=1}^{B} \mu_j^{(i)} = \sum_{j=1}^{B} \sum_{m=1}^{n} r_{jm}\left(2c_{jm}^{(i)} - 1\right), \qquad (8.2.16)$$

где i означает один из конкурирующих путей в каждом узле, а B – это число ветвей (информационных символов) в одном пути. Например, путь из одних нулей, обозначенный i=0, имеет метрику пути

$$CM^{(0)} = \sum_{j=1}^{B} \sum_{m=1}^{n} \left(-\sqrt{\mathscr{E}_c} + n_{jm}\right)(-1) = \sqrt{\mathscr{E}_c} Bn + \sum_{j=1}^{B} \sum_{m=1}^{n} n_{jm}. \qquad (8.2.17)$$

Поскольку свёрточный код не обязательно имеет фиксированную длину, мы определим его качество через вероятность ошибки для последовательности, которая первая сливается с последовательностью из одних нулей в узле решётки. В частности, мы определим вероятность первого пересечения другого пути в узле B с путём из одних нулей, как

---

[1] При фиксированной задержке модификации алгоритма Витерби используют правило приёма в целом с поэлементным решением (ПЦПР), которое совместно с обратной связью по решению составляют основу алгоритма Кловского–Николаева (АКН) (прп).

вероятность того, что этот путь имеет метрику, которая превосходит метрику пути из одних нулей первый раз. Предположим, что неправильный путь, отметим его i=1, который сливается с путём из одних нулей, отличается от пути из одних нулей на d бита, то есть в пути i=1 имеются d единиц, а остальные элементы – нули.

Вероятность ошибки при сравнении матриц $CM^{(0)}$ и $CM^{(1)}$ равна

$$P_2(d) = P\left(CM^{(1)} \geq CM^{(0)}\right) = P\left(CM^{(1)} - CM^{(0)} \geq 0\right),$$

$$P_2(d) = P\left[2\sum_{j=1}^{B}\sum_{m=1}^{n} r_{pn}\left(c_{pn}^{(1)} - c_{pn}^{(0)}\right) \geq 0\right]. \qquad (8.2.18)$$

Поскольку кодированные символы в двух путях одинаковы, за исключением d позиций, можно (8.2.18) записать в более простой форме

$$P_2(d) = P\left(\sum_{l=1}^{d} r_l' \geq 0\right), \qquad (8.2.19)$$

где индекс l пробегает по набору из d позиций символов, в которых два пути отличаются, а набор $\{r_l'\}$ представляет вход декодера по этим символам.

Слагаемые $\{r_l'\}$ представляют собой независимые и одинаково распределённые случайные гауссовские величины со средним $-\sqrt{\mathscr{E}_c}$ и дисперсией $\frac{1}{2}N_0$. Следовательно, вероятность ошибки при попарном сравнении этих двух путей, которые отличаются в d позициях, равна

$$P_2(d) = Q\left(\sqrt{\frac{2\mathscr{E}_c}{N_0}d}\right) = Q\left(\sqrt{2\gamma_b R_c d}\right), \qquad (8.2.20)$$

где $\gamma_b = \mathscr{E}_b/N_0$ – ОСШ на бит принимаемого сигнала, а R_c – скорость кода.

Хотя мы определили вероятность первого пересечения пути, который в d позициях отличается от пути из одних нулей, имеется много возможных путей с различными расстояниями, которые сливаются с путём из одних нулей в узле B. Действительно, передаточная функция $T(D)$ дает полное описание всех возможных путей, которые пересекаются(сливаются) с путём из одних нулей в узле В, а также их расстояний. Поэтому мы можем суммировать вероятность ошибки (8.2.20) по всем возможным расстояниям путей. Выполнив такое суммирование, мы получаем верхнюю границу для вероятности ошибки первого пересечения в виде

$$P_c \leq \sum_{d=d_{\text{св}}}^{\infty} a_d P_2(d) \leq \sum_{d=d_{\text{св}}}^{\infty} a_d Q\left(\sqrt{2\gamma_b R_c d}\right), \qquad (8.2.21)$$

где a_d означает число путей с расстоянием d от пути из одних нулей, которые дает первое пересечение с путём из одних нулей.

Имеются два соображения, почему (8.2.21) является верхней границей вероятности ошибки первого пересечения. Первое – это то, что события, которые определяют вероятности ошибок $\{P_2(D)\}$, совместные. Это можно видеть из рассмотрения решётки. Второе – при суммировании по всем возможным $d \geq d_{\text{св}}$ мы безоговорочно предполагаем, что свёрточный код имеет неограниченную длину. Если код повторяется периодически после B узлов, верхнюю границу в (8.2.21) можно улучшить путём суммирования ошибочных событий по $d_{\text{св}} \leq d \leq B$. Эта тонкость имеет некоторое достоинство при определении качества коротких свёрточных кодов, но её влияние на качество пренебрежимо, когда B велико.

Если $k>1$, эквивалентная вероятность ошибки на бит получается путём деления (8.2.26) и (8.2.27) на k.

Выражения для вероятности ошибки, данные выше, базируются на предположении, что кодовые символы передаются двоичной ФМ при когерентном приеме. Результаты для P_b также верны для четырёхфазной когерентной ФМ, поскольку эта техника модуляции/демодуляции эквивалентна двум независимым (квадратурным по фазе) двоичным системам ФМ. Для другой техники модуляции и демодуляции, такая как когерентная и некогерентная ЧМ, результаты можно приспособить путём пересчета парной вероятности ошибки $P_2(d)$. Это значит, что выбор техники модуляции и демодуляции, использованной для передачи кодированной информационной последовательности, влияет только на расчет $P_2(d)$. Расчёт P_b остаётся тем же.

Хотя приведенные выше расчёты вероятности ошибки при декодировании по Витерби свёрточного кода применимы для двоичных свёрточных кодов, относительно легко обобщить их на недвоичные свёрточные коды, в котором каждый недвоичный символ отображается различным сигналом. В частности, коэффициенты $\{\beta_d\}$ в выражении производной $T(D,N)$, даваемые (8.2.25), представляют число ошибок в символах в двух путях, разделенных расстоянием(числом символов) в d символов. Снова мы обозначим вероятность ошибки при парном сравнении двух путей, которые разделены расстоянием d, через $P_2(d)$. Тогда вероятность ошибки символа, для k-битового символа, ограничена сверху

$$P_M \le \sum_{d=d_{св}}^{\infty} \beta_d P_2(d).$$

Вероятность ошибки символа можно превратить в эквивалентную вероятность ошибки на бит. Для примера, если используются 2^k ортогональных сигналов для передачи k-битовых символов, эквивалентная вероятность ошибки на бит равна P_M, умноженной на множитель $2^{k-1}/(2^k-1)$, как указано в главе 5.

8.2.4. Вероятность ошибки при декодировании жёстких решений

Теперь рассмотрим качество, достижимое алгоритмом декодирования Витерби в двоичном симметричном канале. При декодировании жёстких решений свёрточного кода метрики для алгоритма Витерби являются расстояния Хемминга между принимаемой последовательностью и $2^{k(K-1)}$ выжившими последовательностями в каждом узле решётки. Как при нашей трактовки декодирования мягких решений, мы начнем с расчета вероятности первого ошибочного события. Считается, что передается путь из одних нулей. Предположим, что путь, который сравнивается с путём из одних нулей в некотором узле B, имеет расстояние d относительно пути из одних нулей. Если d нечётно, путь из одних нулей будет выбран без ошибок, если число ошибок в принимаемой последовательности меньше, чем $\frac{1}{2}(d+1)$; в противном случае будет выбран неправильный путь. Следовательно, вероятность выбора неправильного пути равна

$$P_2(d) = \sum_{k=(d+1)/2}^{d} \binom{d}{k} p^k (1-p)^{d-k}, \qquad (8.2.28)$$

где p – вероятность ошибочного приема символа в двоичном симметричном канале. Если d чётно, то неправильный путь выбирается, когда число ошибок превышает $\frac{1}{2}d$. Если число ошибок равно $\frac{1}{2}d$, то имеется связь между метриками двух путей, которую можно

Верхнюю границу в (8.2.21) можно выразить в несколько другой форме, если учесть верхнюю экспоненциальную границу для Q-функции:

$$Q\left(\sqrt{2\gamma_b R_c d}\right) \le e^{-\gamma_b R_c d} = D^d\Big|_{D=e^{-\gamma_b R_c}} . \qquad (8.2.22)$$

Если использовать (8.2.22) в (8.2.21), верхняя граница для вероятности ошибки первого пересечения выражается так:

$$P_e < T(D)\Big|_{D=e^{-\gamma_b R_c}} . \qquad (8.2.23)$$

Хотя вероятность ошибки при первом пересечении дает меру качества свёрточного кода, более часто используемой мерой качества является вероятность ошибки на бит. Для этой вероятности можно получить верхнюю границу посредством процедуры, использованной для получения верхней границы вероятности ошибки при первом пересечении. Подробнее, мы знаем, что, когда выбирается неправильный путь, информационные символы, по которым отобранный путь отличается от правильного пути, будут декодированы неправильно. Мы также знаем, что показатель в множителе N, содержащийся в передаточной функции $T(D,N)$, указывает на число ошибок по информационным символам (число «l») при выборе неправильного пути, который сливается с путём из одних нулей в одном и том же узле B. Если мы умножим вероятность ошибки двоичного перехода $P_2(d)$ на число неправильно декодированных информационных символов в неправильном пути у узла, где он пересекается с правильным, мы получим вероятность ошибки на бит для этого пути. Средняя вероятность ошибки на бит ограничена сверху путём умножения каждой парной вероятности ошибки $P_2(d)$ на соответствующее число неправильно декодированных информационных символов, для каждого возможного неправильного пути, который сливается с правильным путём у B-го узла, и суммированием по всем d.

Подходящие множители для умножения, соответствующие числу ошибок по информационным символам для каждого неправильно выбранного пути можно получить дифференцирования $T(D,N)$ по N. В общем $T(D,N)$ можно выразить так

$$T(D,N) = \sum_{d=d_{cв}}^{\infty} a_d D^d N^{f(d)} , \qquad (8.2.24)$$

где $f(d)$ означает показатель N как функция от d. Взяв производную от $T(D,N)$ по N и положив затем $N=1$, мы получим

$$\frac{dT(D,N)}{dN}\Big|_{N=1} = \sum_{d=d_{cв}}^{\infty} a_d f(d) D^d = \sum_{d=d_{cв}}^{\infty} \beta_d D^d , \qquad (8.2.25)$$

где $\beta_d = a_d f(d)$.

Таким образом, вероятность ошибки на бит при $k=1$ ограничена сверху:

$$P_b < \sum_{d=d_{cв}}^{\infty} \beta_d P_2(d) < \sum_{d=d_{cв}}^{\infty} \beta_d Q\left(\sqrt{2\gamma_b R_c d}\right). \qquad (8.2.26)$$

Если Q-функцию ограничить сверху экспонентой, как указано в (8.2.22), тогда (8.2.26) можно выразить в простой форме

$$P_b < \sum_{d=d_{cв}}^{\infty} \beta_d D^d\Big|_{D=e^{-\gamma_b R_c}} < \frac{dT(D,N)}{dN}\Big|_{N=1,D=e^{-\gamma_b R_c}} . \qquad (8.2.27)$$

разрешить случайным выбором одного из путей; так что ошибка возникнет в половине случаев. Следовательно, вероятность выбора неправильного пути равна

$$P_2(d) = \sum_{k=d/2+1}^{d} \binom{d}{k} p^k (1-p)^{d-k} + \frac{1}{2}\binom{d}{\frac{1}{2}d} p^{d/2}(1-p)^{d/2}. \quad (8.2.29)$$

Как указано в разделе (8.2.3), имеется много возможных путей с различными расстояниями, которые сливаются в данном узле с путём из одних нулей. Следовательно, вероятность первого ошибочного события не являются точным выражением. Однако мы можем определить верхнюю границу посредством суммирования парных ошибок $P_2(d)$ по всем возможным путям, которые сливаются при данном узле с путём из одних нулей. Таким образом мы получим объединенную границу

$$P_c < \sum_{d=d_{св}}^{\infty} a_d P_2(d), \quad (8.2.30)$$

где коэффициенты $\{a_d\}$ представляют число путей, соответствующих набору расстояний $\{d\}$. Эти коэффициенты являются коэффициентами в выражении передаточной функции $T(D)$ или $T(D,N)$.

Вместо использования выражения для $P_2(d)$, даваемое (8.2.29) и (8.2.30), мы можем использовать верхнюю границу

$$P_2(d) < \left[4p(1-p)\right]^{d/2}, \quad (8.2.31)$$

которая была дана в разделе 8.1.5. Использование этой границы в (8.2.30) дает свободную верхнюю границу для вероятности первого ошибочного события в виде

$$P_e < \sum_{d=d_{св}}^{\infty} a_d \left[4p(1-p)\right]^{d/2} < T(D)\Big|_{D=\sqrt{4p(1-p)}}. \quad (8.2.32)$$

Теперь определим вероятность ошибки на бит. Как в случае декодирования мягких решений, мы используем тот факт, что показатели в множителях N, которые имеются в передаточной функции $T(D,N)$ указывают на число ненулевых информационных символов, которые ошибочно приняты, когда неправильный путь выбран вместо пути из одних нулей. Путём дифференцирования $T(D,N)$ по N и приравнивания $N=1$ в результате получим значения для соответствующих вероятностей первых ошибочных событий $P_2(d)$. Так мы получим выражение для верхней границы вероятности ошибки на бит в виде

$$P_b < \sum_{d=d_{св}}^{\infty} \beta_d P_2(d), \quad (8.2.33)$$

где $\{\beta_d\}$ – коэффициенты в выражении для производной $T(D,N)$, рассчитанной при $N=1$. Для $P_2(d)$, мы можем использовать или выражения, данные (8.2.28) и (8.2.29), или верхнюю границу (8.2.31). Если используется последняя, верхнюю границу для P_b можно выразить так

$$P_b < \frac{dT(D,N)}{dN}\Big|_{N=1, D=\sqrt{4p(1-p)}}. \quad (8.2.34)$$

Если $k>1$, результат (8.2.33) и (8.2.34) для P_b надо разделить на k.

На рис. 8.2.15 дано сравнение вероятности ошибки для свёрточного кода со скоростью 1/3, $K=3$ при декодировании мягких и жёстких решений. Заметим, что верхняя граница

Чернова (8.2.24) хуже на 1 дБ относительно плотной верхней границы (8.2.33) в соединении с (8.2.28) и (8.2.29). Преимущество границы Чернова – простота вычислений. Сравнивая качество декодирования мягких и жёстких решений, отметим, что разница, получаемая от верхних границ примерно 2, 5 дБ для $10^{-6} \le P_b \le 10^{-2}$.

В заключение мы хотим напомнить, что средняя по ансамблю вероятность ошибки для свёрточного кода в дискретном канале без памяти, так же как в случае блокового кода, можно выразить через предельную скорость R_0 (доказательство см. Витерби и Омура, 1979)

$$\overline{P_b} < \frac{(q-1)q^{-KR_0/R_c}}{\left[1-q^{-(R_0-R_c)/R_c}\right]^2}, \qquad R_c \le R_0,$$

где q – число входных символов канала, K – кодовое ограничение, R_c – скорость кода, R_0 – предельная скорость, определённая в разделе 7.2 и 8.1. Заключения, полученные путём вычисления R_0 для различных условий в канале, применимы как к блоковым, так и свёрточным кодам.

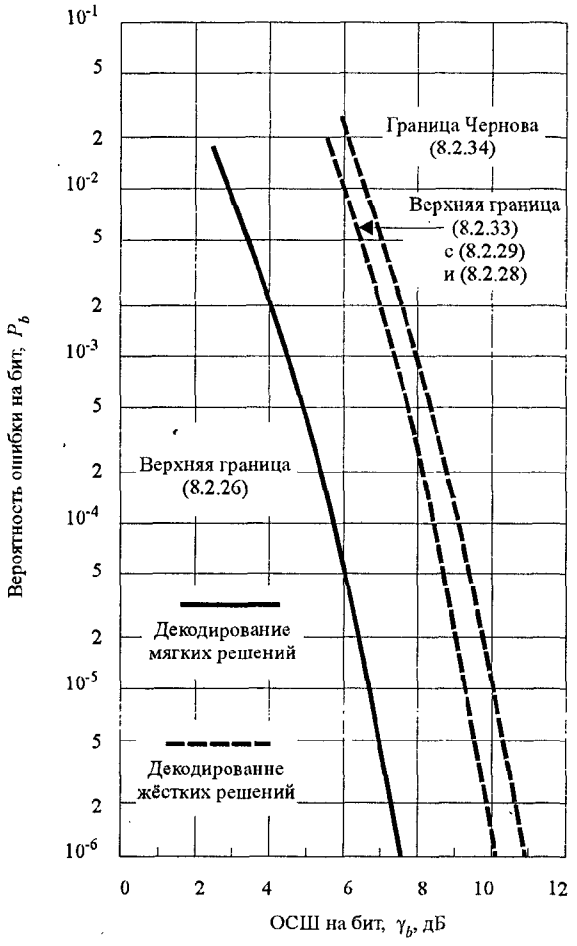


Рис. 8.2.15. Сравнение декодирования мягких и жёстких решений для свёрточного кода с K=3, k=1, n=3

8.2.5. Дистанционные характеристики двоичных свёрточных кодов

В этом подразделе мы хотим свести в таблицу минимальные свободные расстояния и генераторы для нескольких свёрточных кодов с малыми кодовыми ограничениями и для

нескольких скоростей кода. Эти двоичные коды оптимальны в том смысле, что при заданным, скорости кода и кодовому ограничению, они имеют наибольше возможное $d_{св}$. Генераторы и соответствующие значения $d_{св}$, табулированные ниже, были получены Оденвальдером (1970), Ларсеном (1973), Пааске (1974) и Даутом и др. (1982) посредством компьютерных методов исследования.

Хеллер (1968) нашел относительно простую верхнюю границу для минимального свободного расстояния для свёрточного кода со скоростью $1/n$. Она определяется как

$$d_{св} \le \min_{l>1} \left\lfloor \frac{2^{l-1}}{2^{l-1}-1}(K+l-1)n \right\rfloor, \qquad (8.2.35)$$

где $\lfloor x \rfloor$ означает наибольшее целое, содержащееся в x. С целью сравнения эта верхняя граница также дана в таблицах для скорости кода $1/n$. Для свёрточных кодов со скоростью k/n Даут и др. (1982) дали модификацию границы Хеллера. Значения, полученные посредством этой верхней границы для кодов со скоростью k/n, также табулированы.

В таблицах 8.2.1–8.2.7 даны параметры свёрточных кодов, имеющих скорость $1/n$ при $n = 2, 3, \ldots, 8$. В таблицах 8.2.8–8.2.11 даны параметры свёрточных кодов, имеющих скорость k/n для $k \le 4$, $n \le 8$.

Табл. 8.2.1. Максимальное свободное расстояние кодов со скоростью 1/2

Кодовое ограничение K	Порождающие полиномы (в восьмеричной записи)		$d_{св}$	Верхняя граница для $d_{св}$
3	5	7	5	5
4	15	17	6	6
5	23	35	7	8
6	53	75	8	8
7	133	171	10	10
8	247	371	10	11
9	561	753	12	12
10	1.167	1.545	12	13
11	2.335	3.661	14	14
12	4.335	5.723	15	15
13	10.533	17.661	16	16
14	21.675	27.123	16	17

Источники: *Odenwalder* (1970) и *Larsen* (1973)

Табл. 8.2.2. Максимальное свободное расстояние кодов со скоростью 1/3

Кодовое ограничение K	Порождающие полиномы (в восьмеричной записи)			$d_{св}$	Верхняя граница для $d_{св}$
3	5	7	7	8	8
4	13	15	17	10	10
5	25	33	37	12	12
6	47	53	75	13	13
7	133	145	175	15	15
8	225	331	367	16	16
9	557	663	711	18	18
10	1.117	1.365	1.633	20	20
11	2.353	2.671	3.175	22	22
12	4.767	5.723	6.265	24	24
13	10.533	10.675	17.661	24	24
14	21.645	35.661	37.133	26	26

Источники: *Odenwalder* (1970) и *Larsen* (1973)

Табл. 8.2.3. Максимальное свободное расстояние кодов со скоростью 1/4 [*Larsen* (1973)]

Кодовое ограничение K	Порождающие полиномы (в восьмеричной записи)				$d_{св}$	Верхняя граница для $d_{св}$
3	5	7	7	7	10	10
4	13	15	15	15	13	15
5	25	33	33	33	16	16
6	53	53	53	53	18	18
7	135	145	145	145	20	20
8	235	331	331	331	22	22
9	463	663	663	663	24	24
10	1.117	1.365	1.365	1.365	27	27
11	2.387	2.671	2.671	2.671	29	29
12	4.767	5.723	5.723	5.723	32	32
13	11.145	17.661	17.661	17.661	33	33
14	21.113	37.133	37.133	37.133	36	36

Табл.8.2.4 Максимальное свободное расстояние кодов со скоростью 1/5 [*Daut* и др. (1982)]

Кодовое ограничение K	Порождающие полиномы (в восьмеричной записи)					$d_{св}$	Верхняя граиица для $d_{св}$
3	7	7	7	5	5	13	13
4	17	17	13	15	15	16	16
5	37	27	33	25	35	20	20
6	75	71	73	65	57	22	22
7	175	131	135	135	147	25	25
8	257	233	323	271	357	28	28

Табл. 8.2.5. Максимальное свободное расстояние кодов со скоростью 1/6 [*Daut* и др. (1982)]

Кодовое ограничение K	Порождающие полиномы (в восьмеричной записи)			$d_{св}$	Верхняя граница для $d_{св}$
3	7	7	7	16	16
	7	5	5		
	17	17	13	20	20
4	13	15	15		
5	37	35	27	24	24
	33	25	35		
6	73	75	55	27	27
	65	47	57		
7	173	151	135	30	30
	135	163	137		
8	253	375	331	34	34
	235	313	357		

Табл. 8.2.6. Максимальное свободное расстояние кодов со скоростью 1/7 [*Daut* и др. (1982)]

Кодовое ограничение K	Порождающие полиномы (в восьмеричной записи)				$d_{св}$	Верхняя граница для $d_{св}$
3	7	7	7	7	18	18
	5	5	5			
	17	17	13	13	23	23
4	13	15	15			
5	35	27	25	27	28	28
	33	35	37			
6	53	75	65	75	32	32
	47	67	57			
7	165	145	173	135	36	36
	135	147	137			
8	275	253	375	331	40	40
	235	313	357			

Табл. 8.2.7. Максимальное свободное расстояние кодов со скоростью 1/8

Кодовое ограничение K	Порождающие полиномы (в восьмеричной записи)				$d_{св}$	Верхняя граница для $d_{св}$
3	7	7	5	5	21	21
	5	7	7	7		
	17	17	13	13	26	26
4	13	15	15	17		
5	37	33	25	25	32	32
	35	33	27	37		
6	57	73	51	65	36	36
	75	47	67	57		
7	153	111	165	173	40	40
	135	135	147	137		
8	275	275	253	371	45	45
	331	235	313	357		

Табл. 8.2.8. Максимальное свободное расстояние кодов со скоростью 2/3

Кодовое ограничение K	Порождающие полиномы (в восьмеричной записи)			$d_{св}$	Верхняя граница для $d_{св}$
2	17	06	15	3	4
3	27	75	72	5	6
4	236	155	337	7	7

Табл. 8.2.9. Максимальное свободное расстояние кодов со скоростью *k*/5

Скорость	Кодовое ограничение K	Порождающие полиномы (в восьмеричной записи)					$d_{св}$	Верхняя граница для $d_{св}$
2/5	2	17	07	11	12	04	6	6
	3	27	71	52	65	57	10	10
	4	247	366	171	266	373	12	12
3/5	2	35	23	75	61	47	5	5
4/5	2	237	274	156	255	337	3	4

Табл. 8.2.10. Максимальное свободное расстояние кодов со скоростью *k*/7

Скорость	Кодовое ограничение	Порождающие полиномы (в восьмеричной записи)				$d_{св}$	Верхняя граница для $d_{св}$
2/7	2	05	06	12	15	9	9
		15	13	17			
	3	33	55	72	47	14	14
		25	53	75			
	4	312	125	247	366	18	18
		171	266	373			
3/7	2	45	21	36	62	8	8
		57	43	71			
4/7	2	130	067	237	274	6	7
		156	255	337			

Табл. 8.2.11. Максимальное свободное расстояние кодов со скоростью 3/4 и 3/8

Скорость	Кодовое ограничение K	Порождающие полиномы (в восьмеричной записи)				$d_{св}$	Верхняя граница для $d_{св}$
3/4	2	13	25	61	47	4	4
3/8	2	15	42	23	61	8	8
		51	36	75	47		

Источник таблиц 8.2.7–8.2.11: *Daut* и др. (1982)

8.2.6. Недвоичные k-дуальные коды и каскадные коды

Наше обсуждение свёрточных кодов до сих пор было сконцентрировано прежде всего на двоичные коды. Двоичные коды особенно пригодны для каналов в которых возможно использовать двоичную или четверичную ФМ с когерентной демодуляцией.

Однако имеется много приложений для которых ФМ и когерентная демодуляция не подходит или невозможна. В этих случаях используется другая техника модуляции, например M-ичная ЧМ, в соединении с некогерентной демодуляцией. Недвоичные коды особенно хорошо согласуются с M-ичными сигналами, которые демодулируются некогерент-но.

В этом разделе мы опишем класс недвоичных свёрточных кодов, называемых k-дуальными кодами, которые легко декодируются посредством алгоритма Витерби, используя декодирование мягких или жёстких решений Они также подходят как внешний код или как внутренний код в каскадном коде, который также будет описан ниже.

Дуальный k свёрточный кодер со скоростью 1/2 можно представить так, как показано на рисунке (8.2.16). Он состоит из двух $(K=2)$ k-битовых регистров сдвига и $n=2k$ генераторов функций. У него два k-битовых выхода. Заметим что код, рассмотренный в примере (8.2.3), является 2-дуальным свёрточным кодом.

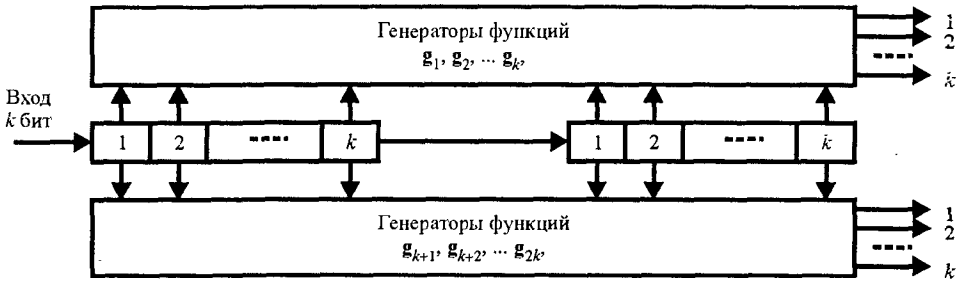


Рис. 8.2.16. Кодер для k-дуального кода со скоростью 1/2

$2k$ генераторов функций для k-дуальных кодов были даны Витерби и Джекобсом (1975). Их можно выразить в форме

$$\begin{bmatrix} \leftarrow g_1 \rightarrow \\ \leftarrow g_2 \rightarrow \\ \vdots \\ \leftarrow g_k \rightarrow \end{bmatrix} = \left[\begin{array}{ccccccc:ccc} 1 & 0 & 0 & \dots & 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 & 0 & \dots & 0 & 1 \end{array}\right] = \left[\mathbf{I}_k \quad \mathbf{I}_k\right]$$

$$\begin{bmatrix} \leftarrow g_{k+1} \rightarrow \\ \leftarrow g_{k+2} \rightarrow \\ \vdots \\ \leftarrow g_{2k} \rightarrow \end{bmatrix} = \left[\begin{array}{ccccccc:ccccc} 1 & 1 & 0 & 0 & & \dots & 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & & \dots & 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & 1 & 0 & \dots & 0 & 0 & 0 & 1 & 0 \dots & 0 \\ 0 & 0 & 0 & & & \dots & 0 & 1 & & & & \\ 1 & 0 & 0 & & & \dots & 0 & 0 & 0 & & \dots & 0 \quad 1 \end{array}\right]$$

$$= \left[\begin{array}{ccccccc:ccc} 1 & 1 & 0 & 0 & & \dots & 0 & & & \\ 0 & 0 & 1 & 0 & & \dots & 0 & & & \\ 0 & 0 & 0 & 1 & 0 & \dots & 0 & & \mathbf{I}_k & \\ \hdashline 0 & 0 & 0 & & & \dots & 0 & 1 & & \\ 1 & 0 & 0 & & & \dots & 0 & 0 & & \end{array}\right], \qquad (8.2.36)$$

где $\mathbf{I}_k$ означает $k \times k$ единичную матрицу.

Общая формула для передаточной функции k-дуального кода со скоростью 1/2 была найдена Оденвальдером (1976).Она выражается так:

$$T(D,N,J)=\frac{\left(2^k-1\right)D^4J^2N}{1-NJ\left[2D+\left(2^k-3\right)D^2\right]}=\sum_{i=4}^{\infty}a_iD^iN^{f(i)}J^{h(i)}, \quad (8.2.37)$$

где D представляет расстояние Хемминга для q-ичных $\left(q=2^k\right)$ символов, показатель $f(i)$ для N представляет число ошибок в информационных символах, которые имеют место при выборе ветви на дереве или в решётке, отличающейся от соответствующих ветвей пути из одних нулей, показатель $h(i)$ для J равен числу ветвей для данного пути. Заметим, что минимальное свободное расстояние равно $d_{\text{св}}=4$ символам ($4k$ битам)

Низкоскоростные k-дуальные свёрточные коды можно генерировать различными путями, простейший сводится к повторению каждого символа, генерированного кодом со скоростью 1/2, r раз, где $r=1, 2, \ldots, m$ ($r=1$ соответствует появлению каждого символа один раз). Если каждый символ в частной ветви дерева или решётки или диаграммы состояний повторяется r раз, то увеличивается параметр расстояний от D до D^r. Следовательно, передаточные функции для k-дуального кода со скоростью $1/2r$ равна

$$T(D,N,J)=\frac{\left(2^k-1\right)D^{4r}J^2N}{1-NJ\left[2D^r+\left(2^k-3\right)D^{2r}\right]}. \quad (8.2.38)$$

При передаче длинных информационных последовательностей параметр длина пути J в передаточной функции можно подавить, положив $J=1$. Результирующую передаточную функцию $T(D,N)$ можно дифференцировать по N и положить затем N=1.Это даёт

$$\left.\frac{dT(D,N,J)}{dN}\right|_{N=1}=\frac{\left(2^k-1\right)D^{4r}}{\left[1-2D^r-\left(2^k-3\right)D^{2r}\right]^2}=\sum_{t=4r}^{\infty}\beta_tD^t, \quad (8.2.39)$$

где β_t представляет число ошибок в символах, связанных с путём имеющим расстояние D^t от пути с одними нулями, как было описано ранее в разделе (8.2.3).Выражение (8.2.3.) можно использовать для расчёта вероятности ошибки для k-дуальных кодов при различных условиях в канале.

Качество k-дуальных кодов с M-ичной модуляцией. Предположим, что k-дуальный код используется в соединении с M-ичными ортогональными сигналами в модуляторе, где $M=2^k$. Каждый символ кодера отображается в один из M возможных ортогональных сигналов. Считается, что в канале действует АБГШ. Демодулятор состоит из M согласованных фильтров.

Если декодер осуществляет декодирование жёстких решений, качество кода определяется вероятностью ошибки символа P_M. Эта вероятность ошибки была рассчитана в главе 5 для когерентного и некогерентного детектирования. По P_M мы можем определить $P_2(d)$ согласно (8.2.28) или (8.2.29), что является вероятностью ошибки при парном сравнении пути из одних нулей с путём, который отличается в d символах. Вероятность ошибки на бит имеет верхнюю границу

$$P_b<\frac{2^{k-1}}{2^k-1}\sum_{d=4r}^{\infty}\beta_dP_2(d). \quad (8.2.40)$$

Множитель $2^{k-1}/\left(2^k-1\right)$ используется для превращения вероятности ошибки символа в вероятность ошибки на бит.

Вместо декодирования жёстких решений предположим, что декодер осуществляет декодирование мягких решений, используя выход демодулятора, который использует квадратичный детектор. Выражение для вероятности ошибки на бит (8.2.40) всё ещё применимо, но теперь $P_2(d)$ определяется так (смотри раздел 12.1.1):

$$P_2(d) = \frac{1}{2^{2d-1}} \exp\left(-\tfrac{1}{2}\gamma_b R_c d\right) \sum_{i=0}^{d-1} K_i \left(\tfrac{1}{2}\gamma_b R_c d\right), \qquad (8.2.41)$$

где

$$K_i = \frac{1}{i!} \sum_{i=0}^{d-1-i} \binom{2d-1}{l},$$

а $R_c = 1/2r$ скорость кода. Это выражение следует из результата (8.1.63.).

Каскадные коды. В разделе 8.1.8 мы рассмотрели каскадное объединение двух блоковых кодов для того, чтобы сформировать длинный блоковый код. Теперь, поскольку мы описываем свёрточные коды, мы расширим нашу точку зрения и рассмотрим каскадное объединение блокового кода со свёрточным кодом или каскадное объединение двух свёрточных кодов.

Как было описано раньше, внешний код обычно выбирается недвоичным с выбором кодового символа из алфавита $q = 2^k$ символов. Этот код может быть блоковым кодом, таким как код Рида-Соломона или свёрточным кодом, таким как k-дуальный код. Внутренний код может быть как двоичным, так и недвоичным и блоковым или свёрточным. Для примера, в качестве внешнего кода можно выбрать код Рида-Соломона, а k-дуальный код можно выбрать в качестве внутреннего. В такой каскадной схеме число символов внешнего кода $q = 2^k$, так что каждый символ внешнего кода отображается k-битовым символом внутреннего k-дуального кода. Для передачи символов можно использовать M-ичные ортогональные сигналы.

Декодирование такого кода также может принимать различные формы. Если внутренний код свёрточный и имеет короткое кодовое ограничение, алгоритм Витерби обеспечивает эффективный способ декодирования, используя декодирование либо мягких, либо жёстких решений.

Если внутренний код блоковый, и декодер для этого кода обеспечивает декодирование мягких решений, внешний декодер также может выполняться с мягким решением, используя в качестве входов метрики, соответствующие каждому кодовому слову внутреннего кода. С другой стороны, внутренний декодер может выполнять декодирование жёстких решений после получения кодовых слов и отправлять жёсткое решение внешнему декодеру. Тогда внешний декодер должен формировать декодирование жёстких решений.

Следующий пример описывает каскадный код, в котором внешний код свёрточный, а внутренний – блоковый.

Пример 8.2.5. Предположим, что мы сконструировали каскадный код, выбрав k-дуальный код в качестве внешнего кода и блоковый код Адамара в качестве внутреннего. Для конкретности, выберем 5-дуальный код со скоростью 1/2 и код Адамара (16, 5) в качестве внутреннего кода. Дуальный код со скоростью 1/2 имеет минимальное свободное расстояние $D_{\text{св}} = 4$, а код Адамара имеет минимальное расстояние $d_{\min} = 8$. Следовательно, каскадный код имеет эффективное минимальное расстояние 32. Поскольку имеется 32 кодовых слова в коде Адамара и 32 возможных символа во внешнем коде, в

итоге каждый символ внешнего кода отображается в одно из 32 кодовых слов кода Адамара.

Вероятность ошибочного декодирования символа внутренним кодом можно определить из результатов качества блоковых кодов, данных в разделах 8.1.4 и 8.1.5, соответственно, для декодирования мягких и жёстких решений.

Сначала предположим, что во внутреннем декодере осуществляется жёсткое решение с вероятностью ошибки декодирования кодовых слов (символов внешнего кода), обозначенную P_{32}, так как $M = 32$. Тогда качество внешнего кода и, следовательно, качество каскадного кода можно получить, используя эту вероятность ошибки в соединении с передаточной функцией для 5-дуального кода, определенной в (8.2.32).

С другой стороны, если декодирование мягких решений используется во внешнем и внутреннем декодерах, метрики для мягкого решения о каждом принимаемом кодовом слове кода Адамара передаются на алгоритм Витерби, который вычисляет результирующие метрики для конкурирующих путей решётки. Мы дадим численные результаты качества каскадных кодов этого типа в нашей дискуссии о кодировании в каналах с релеевскими замираниями.

8.2.7. Другие алгоритмы декодирования свёрточных кодов

Алгоритм Витерби, описанный в разделе 8.2.2, это оптимальный алгоритм декодирования для свёрточных кодов. Однако он требует вычисления 2^{kK} метрик каждого узла решётки и хранения $2^{k(K-1)}$ метрик и $2^{k(K-1)}$ выживших последовательностей, каждая из них имеет длину около $5kK$ бит. Вычислительное время и память для хранения, требуемого для реализации алгоритма Витерби, делают его практически неприемлемым для свёрточных кодов с большим кодовым ограничением.

Еще до открытия оптимального алгоритма Витерби, были придуманы другие алгоритмы для декодирования свёрточных кодов. Самым ранним был алгоритм последовательного декодирования, впервые предложенный Возенкрафтом и впоследствии модифицированный Фано (1963).

Последовательный алгоритм декодирования Фано ищет наиболее правдоподобный путь по дереву, или решётке путём проверки каждый раз одного пути. Приращение, добавляемое к метрике вдоль ветви пропорционально вероятности принимаемого сигнала для этой ветви, как в алгоритме Витерби, за исключением того, что к метрике каждой ветви добавляется отрицательная константа прибавляется. Величина этой константы выбирается так, что метрика правильного пути будет в среднем увеличиваться, в то время как метрика для любого неправильного пути в среднем будет уменьшаться. Путём сравнения метрики претендующего пути с меняющимся (увеличивающимся) порогом алгоритм Фано обнаруживает и отвергает неправильные пути.

Для большей конкретности рассмотрим канал без памяти.

Метрика для i-го пути по дереву или решётке от первой ветви до B-й ветви можно выразить так

$$CM^{(i)} = \sum_{j=1}^{B} \sum_{m=1}^{n} \mu_{jm}^{(i)}, \tag{8.2.42}$$

где

$$\mu_{jm}^{(i)} = \log_2 \frac{p\left(r_{jm} \middle| c_{jm}^{(i)}\right)}{p\left(r_{jm}\right)} - \mathcal{K}. \tag{8.2.43}$$

В (8.2.43) r_{jm} – выходная последовательность демодулятора, $p(r_{jm}|c_{jm}^{(i)})$ означает условную ФПВ r_{jm} при условии кодового бита $c_{jm}^{(i)}$ (m-й бит в j-й ветви i-го пути), а $\mathcal{K}$ – положительная константа. $\mathcal{K}$ выбирается, как сказано выше, так, что неправильные пути будут иметь уменьшающиеся метрики, в то время как правильный путь в среднем увеличивает метрику. Заметим, что член $p(r_{jm})$ в знаменателе не зависит от кодовой последовательности и, следовательно, может быть включен в постоянную.

Метрика, определённая (8.2.43), в общем применима при декодировании как жёстких, так и мягких решений. Однако она может быть особенно упрощена, когда используется декодирование жёстких решений. В частном случае, если мы имеем ДСК с переходной вероятностью ошибки p, метрика для каждого принятого символа, согласующаяся с формулой (8.2.43), определяется так

$$\mu_{jm}^{(i)} = \begin{cases} \log_2\left[2(1-p)\right] - R_c & \text{при} \quad \tilde{r}_{jm} = c_{jm}^{(i)} \\ \log_2 2p - R_c & \text{при} \quad \tilde{r}_{jm} \neq c_{jm}^{(i)} \end{cases} \qquad (8.2.44)$$

где $\tilde{r}_{jm}$ – выходы жёстких решений демодулятора,

$c_{jm}^{(i)}$ – m-й кодовый символ в j-й ветви i-го пути дерева,

R_c – скорость кода.

Заметим, что эти метрики требуют хотя бы приближённого знания вероятности ошибки p.

Пример 8.2.6. Предположим, что для передачи информации по ДСК с вероятностью ошибки $p = 0{,}1$ используется двоичный свёрточный код со скоростью $R_c = 1/3$. Вычисление согласно (8.2.44) дает

$$\mu_{jm}^{(r)} = \begin{cases} 0.52 & \text{ïðè} \quad \tilde{r}_{jm} = c_{jm}^{(i)}, \\ -2.65 & \text{ïðè} \quad \tilde{r}_{jm} \neq c_{jm}^{(i)}. \end{cases} \qquad (8.2.45)$$

Для упрощения расчётов метрики (8.2.45) можно нормировать. Они хорошо аппроксимируются так:

$$\mu_{jm}^{(r)} = \begin{cases} 1 & \tilde{r}_j = c_{jm}^{(i)}, \\ -5 & \tilde{r}_{jm} \neq c_{jm}^{(i)}. \end{cases} \qquad (8.2.46)$$

Поскольку скорость кода равна 1/3, то кодер имеет три выходных символа на каждый входной символ. Тогда метрики ветвей, согласующиеся с (8.2.46), равны

$$\mu_j^{(i)} = 3 - 6d$$

или, что эквивалентно,

$$\mu_j^{(i)} = 1 - 2d, \qquad (8.2.47)$$

где d – хеммингово расстояние трёх принятых бит от трёх битов на ветви.

Таким образом, метрики $\mu_j^{(i)}$ просто связаны с расстоянием Хемминга между принимаемыми символами и кодовыми символами j-й ветви i-го пути.

Первоначально декодер можно заставить выбирать правильную траекторию путём передачи известной цепочки данных. Тогда он продвигается вперед от узла к узлу, выбирая наиболее вероятную (с большей метрикой) ветвь в каждом узле и увеличивая порог так, что его изменение никогда не больше, чем некоторая заранее выбранная величина, скажем, τ. Теперь предположим, что аддитивный шум (при декодировании мягких решений) или ошибки демодуляции, возникающие из-за шума в канале (при декодировании жёстких

решений) заставят декодер принять неверный путь, поскольку он кажется более правдоподобным, чем правильный путь. Это иллюстрирует рис. 8.2.17.

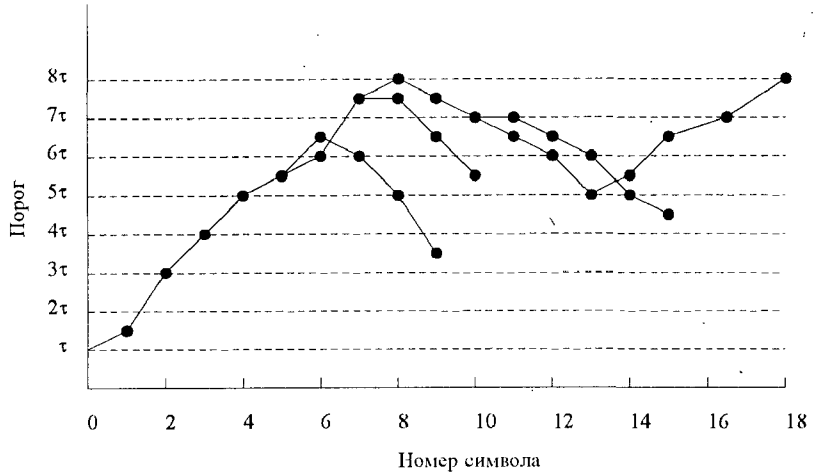


Рис. 8.2.17. Пример поиска пути при последовательном декодировании [*Jordan* (1966), © 1966 *IEEE*]

Поскольку метрики неверного пути в среднем уменьшаются, метрика упадет ниже текущего порога, скажем τ_0. Если это случится, декодер возвращается обратно и берёт альтернативный путь по дереву (решётке) с меньшей метрикой ветви, в попытке найти другой путь с большей метрикой, которая превосходит порог τ_0. Если приходит удача в альтернативном пути, это продолжается вдоль пути все время, и выбирается наиболее правдоподобная ветвь в каждом узле. С другой стороны, если не существует пути, который превышает порог τ_0, порог уменьшается на величину τ и декодер возвращается к первоначальному пути. Если первоначальный путь не находится выше нового порога, декодер начинает обратный поиск других путей. Эта процедура повторяется с уменьшением порога на τ при каждом повторении до тех пор, пока декодер не найдет путь, который остаётся выше установленного порога. Упрощённая структурная схема алгоритма Фано показана на рис. 8.2.18.

Алгоритм последовательного декодирования требует буферную память в декодере для хранения поступающих данных декодирования в течение периодов, когда декодер ищет альтернативные пути. Когда поиск заканчивается, декодер должен быть в состоянии обрабатывать демодулированные символы достаточно быстро, чтобы освободить буфер для начала нового поиска. Иногда в течение экстремально длинных поисков буфер может переполниться. Это вызывает потерю данных, которые можно восстановить повторением потерянной информации. В этой связи мы хотим напомнить, что предельная скорость R_0 имеет особый смысл при последовательном декодировании. Это скорость, выше которой среднее число операций декодирования на декодированный символ становится неограниченным и оно выражает *предельную вычислительную скорость* $R_{\text{пв}}$. На практике последовательные декодеры обычно работают при скоростях, близких к R_0.

Алгоритм последовательного декодирования Фано успешно применяется в различных системах связи. Его качество по вероятности ошибки сравнимо с декодером Витерби. Однако по сравнению с декодером Витерби последовательное декодирование имеет значительно большую задержку декодирования. С другой стороны, последовательное декодирование требует меньше памяти, чем декодирование по Витерби, и, следовательно, оно является привлекательным для свёрточных кодов с большим кодовым ограничением.

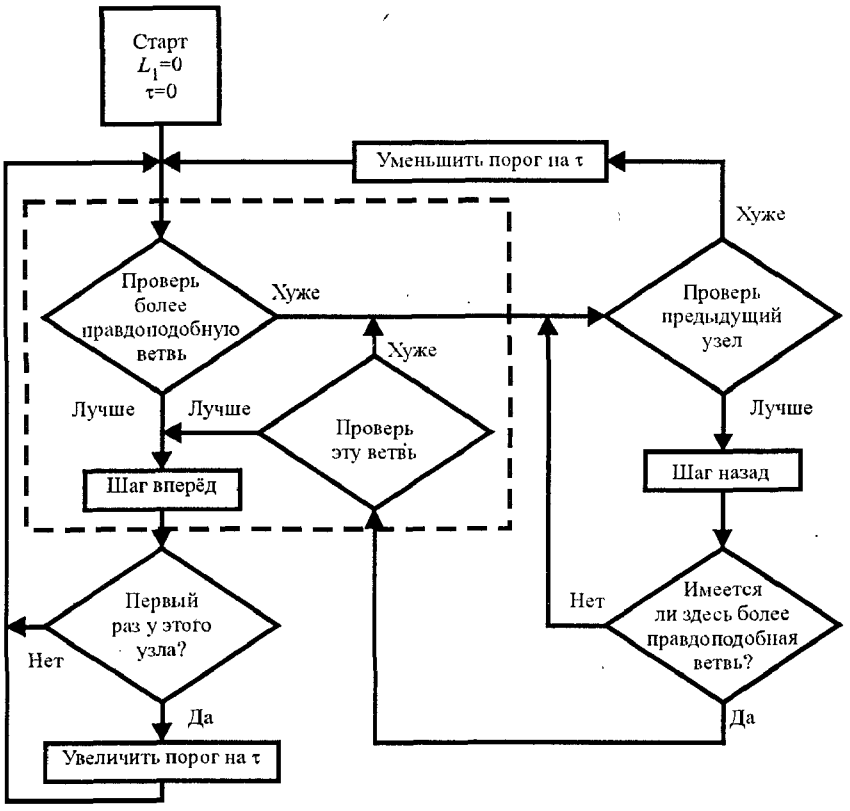
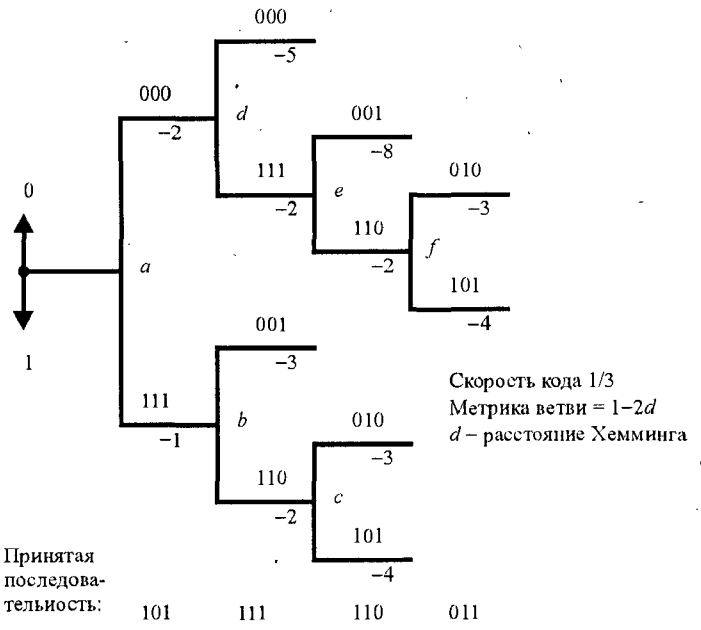


Рис. 8.2.18. Упрощённая структура алгоритма Фано
[*Jordan* (1966), © 1966 *IEEE*]

Исследования вычислительной сложности и требований к памяти для последовательного декодирования вызывают интерес, и они все еще продолжаются. Для анализа этих вопросов и других характеристик алгоритмов Фано интересующемуся читателю рекомендуются книги Галлагера (1986), Возенкрафта и Джекобса (1965), Сэвейдж (1966) и Форни (1974). Другой тип алгоритма последовательного декодирования, названный *стек-алгоритмом*, был предложен независимо Зигангировым (1966) и Елинеком (1969). В противовес алгоритму Витерби, который сохраняет траектории $2^{(K-1)k}$ путей и соответствующих метрик, стек-алгоритм последовательного декодирования работает с несколькими путями и их соответствующими метриками. В стек-алгоритме более вероятные пути упорядочиваются согласно их метрик, причём в верхней части (в голове стека) располагается путь, имеющий наибольшую метрику. На каждом шаге алгоритма только путь в голове стека проверяется по разветвлению. Это приносит 2^k продолжений и их соответствующих метрик. Эти 2^k продолжения вместе с другими путями затем упорядочиваются согласно величинам их метрик, и все пути с метриками, которые располагаются ниже некоторой выбранной величины от метрики главного пути, отбрасываются. Затем процесс продолжения путей с наибольшими метриками повторяется. Рис. 8.2.19 иллюстрирует несколько первых шагов стек-алгоритма.

Очевидно, что если ни одно из 2^k продолжений пути с наибольшей метрикой не остаётся в голове стека, следующий шаг поиска выполняет продолжение другого пути, который приближается к голове стека. Отсюда следует, что алгоритм не всегда даёт быстрое продвижение по начатой траектории при любой итерации. Следовательно, некоторую величину памяти нужно зарезервировать для вновь принятых символов и ранее принятых символов для того, чтобы позволить алгоритму расширить поиск по одному из более коротких путей, когда такой путь достигает головы стека.



Стек с накопленными метриками путей

Шаг *a*	Шаг *b*	Шаг *c*	Шаг *d*	Шаг *e*	Шаг *f*
−1	−2	−2	−2	−2	−3
−2	−2	−3	−3	−3	−3
	−3	−3	−3	−3	−3
		−4	−3	−4	−4
			−5	−5	−4
				−8	−5
					−8

Рис. 8.2.19. Пример работы стек-алгоритма для декодирования свёрточного кода со скоростью 1/3

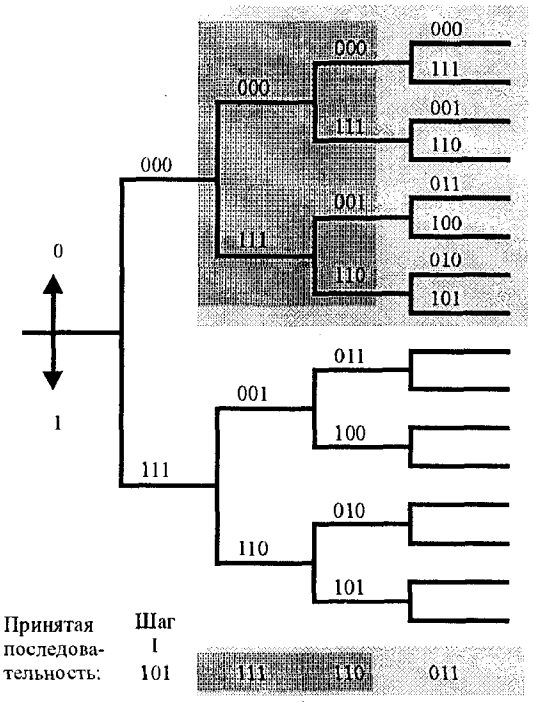
Из сравнении стек-алгоритма с алгоритмом Витерби следует, что стек-алгоритм требует малого числа сравнений метрик, но его вычислительная экономия в большой степени снижается за счёт вычислений, требуемых для упорядочивания стека после каждой итерации. По сравнению с алгоритмом Фано, стек-алгоритм в вычислительном отношении проще, поскольку здесь нет возвращения по тому же пути, как это делает алгоритм Фано. С другой стороны, стек-алгоритм требует больше памяти, чем алгоритм Фано.

Третья альтернатива оптимального декодера Витерби – это метод, названный декодированием с обратной связью (Хеллер 1975), который был разработан для декодирования в ДСК (декодирование жёстких решений). При декодировании с обратной связью декодер делает жёсткое решение об информационном символе на j-м шаге, основываясь на метриках, вычисленных от j-го до $(j+m)$-го шага, где m – выбранное целое число. Таким образом, решение об информационном символе в пользу 0 или 1 зависит от того, каково минимальное расстояние по Хеммингу для пути, который начинается на j-м шаге и кончается на $(j+m)$-м шаге и сколько содержится «0» или «1» в ветви, исходящих от шага j. Решение выносится один раз об информационном символе на j шаге, и только часть дерева, которая связана с этим символом, сохраняется (половина путей, исходящих из узла j), а остальные пути исключаются. Так осуществляется обратная связь в декодере. Следующий шаг заключается в расширении части дерева, которая вышла до шага $j+1+m$ и

рассмотрении путей от шага (j+1)-го до (j+1+m)-го для принятия решения о символе на шаге j+1. Так процедура повторяется на каждом шаге. Параметр m – это просто число шагов по дереву, которое декодер учитывает при вынесении жёсткого решения. Поскольку большая величина m ведет к большой величине памяти, желательно m выбрать как можно меньше. С другой стороны, m должно быть достаточно большим, чтобы избежать существенного ухудшения качества. Чтобы сбалансировать эти два противоречивых требования, m обычно выбирается в области $K \le m \le 2K$, где K – кодовое ограничение. Заметим, что эта задержка при декодировании значительно меньше, чем задержка при использовании алгоритма Витерби, которая обычно около $5K$.

Пример 8.2.7. Рассмотрим использование декодера с обратной связью для свёрточного кода со скоростью 1/3, показанного на рис. 8.2.2. Рис. 8.2.20 иллюстрирует древовидную диаграмму и операции декодера с обратной связью при m=2. Это значит, что при декодировании символа ветви j, декодер рассматривает пути на ветвях j, j+1 и j+2. Начиная с первой ветви, декодер рассчитывает восемь метрик (расстояние Хемминга) и решает, что символ в первой ветви является 0, если путь с минимальным расстоянием находится в верхней части дерева и, что символ 1, если путь с минимальным расстоянием находится в нижней части дерева. В этом примере принимаемая последовательность для первых трёх ветвей полагается такой 101 111 110, поэтому путь с минимальным расстоянием находится в верхней части дерева, следовательно, первый выход символа декодера 0.

Следующие шаги сводятся к расширению верхней части дерева (той части дерева, которая выжила) на одну ветвь и к вычислению восьми метрик в ветвях 2, 3 и 4. Для предположения входной последовательности 111 110 011 путь с минимальным расстоянием находится в нижней части секции дерева, вычислившей после первого шага. Итак, второй выходной символ декодера 1. Третий шаг заключается в расширении этой нижней части дерева и повторении процедуры, описанной для двух первых шагов.



Шаг 1: Метрики верхней части дерева: 7, 6, 5, 2 } →0
метрики нижней части дерева: 5, 4, 3, 4 }
Шаг 2: Метрики верхней части дерева: 7, 6, 5, 6 } →1
метрики нижней части дерева: 3, 6, 1, 2. }

Рис. 8.2.20. Пример декодирования с обратной связью для свёрточного кода со скоростью 1/3

Вместо вычисления метрик описанным выше способом, декодер с обратной связью в ДСК можно эффективно реализовать путём вычисления синдрома по принимаемой

последовательности и используя табличный метод коррекции ошибок. Этот метод похож на тот, который был описан выше для декодирования блоковых кодов. Для некоторых свёрточных кодов декодер с обратной связью упрощается к виду, называемому *логический декодер по большинству* или *пороговый декодер* (Месси 1963; Хеллер 1975).

8.2.8. Практические соображения по применению свёрточных кодов

Свёрточные коды широко используются во многих практических приложениях при синтезе систем связи. Декодирование по Витерби предпочтительно используется при малых кодовых ограничениях $(K \le 10)$, в то время как последовательное декодирование используется при больших кодовых ограничениях, когда сложность декодирования по Витерби становиться чрезмерной. Выбор кодового ограничения диктуется требуемым выигрышем кода.

Из результатов для вероятности ошибки при декодировании мягких решений, данных (8.2.26), очевидно, что выигрыш от кодирования, достигаемый свёрточным кодом, относительно системы без кодирования с двоичной ФМ или КФМ равен

$$\text{выигрыш от кодирования} \le 10\lg\left(R_c d_{\text{св}}\right).$$

Мы также знаем, что минимальное свободное расстояние можно увеличить или за счёт уменьшения скорости кода или за счёт увеличения кодового ограничения или тем или другим одновременно. Табл. 8.2.12 дает данные о верхних границах для выигрыша от кодирования для некоторых свёрточных кодов. С целью сравнения табл. 8.2.13 даёт реальные выигрыши от кодирования и верхние границы для некоторых свёрточных кодов с малым кодовым ограничением при использовании декодирования по Витерби. Следует заметить, что выигрыш от кодирования увеличивается по отношению к асимптотическому пределу по мере увеличения ОСШ.

Таблица 8.2.12. Верхние границы выигрыша кода для декодирования мягких решений для некоторых свёрточных кодов

Скорость кода 1/2			Скорость кода 1/3		
Кодовое ограничение K	$d_{\text{св}}$	Верхняя граница дБ	Кодовое ограничение K	$d_{\text{св}}$	Верхняя граница дБ
3	5	3,98	3	8	4,26
4	6	4,77	4	10	5,23
5	7	5,44	5	12	6,02
6	8	6,02	6	13	6,37
7	10	6,99	7	15	6,99
8	10	6,99	8	16	7,27
9	12	7,78	9	18	7,78
10	12	7,78	10	20	8,24

Таблица 8.2.13. Выигрыша кода (дБ) для декодирования мягких решений по Витерби

P_b	$\mathscr{E}_b/N_0$ некодир.	$R_c = 1/3$		$R_c = 1/2$			$R_c = 2/3$		$R_c = 3/4$	
	(дБ)	$K=7$	$K=8$	$K=5$	$K=6$	$K=7$	$K=6$	$K=8$	$K=6$	$K=9$
10^{-3}	6,8	4,2	4,4	3,3	3,5	3,8	2,9	3,1	2,6	2,6
10^{-5}	9,6	5,7	5,9	4,3	4,6	5,1	4,2	4,6	3,6	4,2
10^{-7}	11,3	6,2	6,5	4,9	5,3	5,8	4,7	5,2	3,9	4,8

Источник: *Jacobs* (1974); © *IEEE*

Рассмотренные результаты базируются на декодировании по Витерби мягких решений. Если используется декодирование жёстких решений, выигрыш от кодирования уменьшается примерно на 2 дБ в канале с АБГШ.

Бóльшие значения выигрыша декодирования, чем те, которые указаны выше в таблицах, достигнуты при использовании свёрточных кодов с бóльшими кодовыми ограничениями, например $K = 50$ и использовании последовательного декодирования. В любом случае, последовательное декодирование применяется при декодировании жёстких решений для уменьшения сложности устройств. Рис. 8.2.21 иллюстрирует характеристики качества (вероятность ошибки) для некоторых свёрточных кодов с кодовым ограничением $K = 7$ и скоростей кода 1/2 и 1/3 при использовании последовательного декодирования (жёстких решений) и с кодовым ограничением $K = 41$ при тех же скоростях кода.

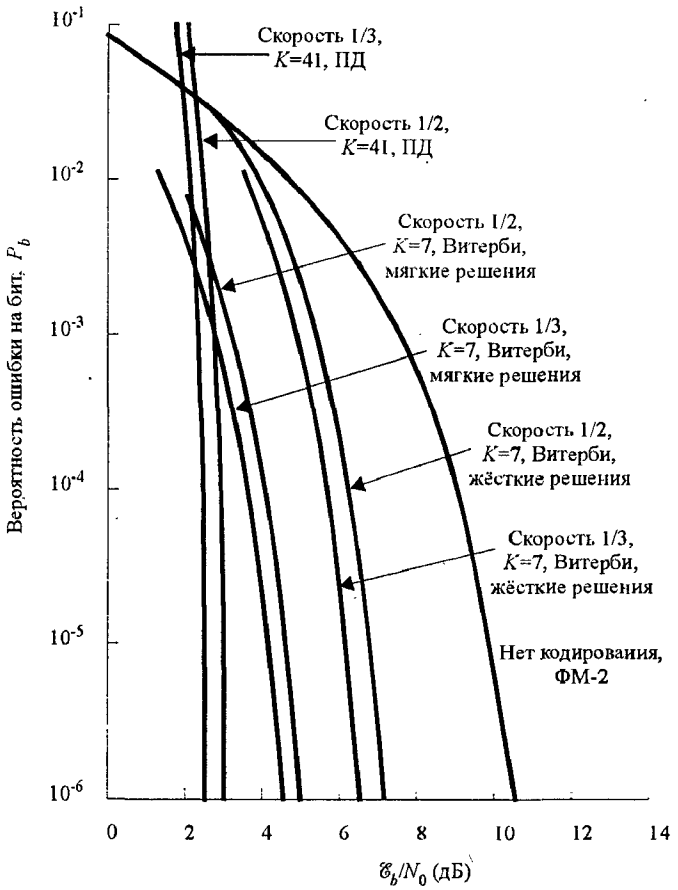


Рис. 8.2.21. Качество при декодировании по Витерби и последовательном декодировании (ПД) при скорости кода 1/2 и 1/3 [*Omura* и *Levitt* (1982). © 1982 *IEEE*]

Заметим, что при $K = 41$ код обеспечивает вероятность ошибки 10^{-4} при ОСШ 2,5...3 дБ, что на 4...4,5 дБ отличается от предела пропускной способности канала, то есть вблизи предельной скорости. Однако коды с $K = 7$ и скоростями 1/2 и 1/3 с декодированием мягких решений по Витерби рассчитанные для вероятности ошибки 10^{-4}, работает при ОСШ соответственно 5 и 4,4 дБ. Эти коды с малым кодовым ограничением достигают выигрыша кодирования примерно 6 дБ при вероятности ошибки 10^{-6}, в то время как коды с большим кодовым ограничением дают выигрыш около 7,5...8 дБ.

Имеются две важных предпосылки к внедрению декодера Витерби:

1) учёт влияния ограничения интервала обработки, которое будет особенно актуальным при использовании мощных кодов и обеспечит фиксированную задержку декодера;

2) допустимая степень квантования входного сигнала для декодера Витерби.

Исходя из опыта, можно утверждать, что усечение памяти кода на величину пяти кодовых ограничений ведёт к пренебрежимо малым потерям в качестве. Рис. 8.2.22 иллюстрирует качество, полученное при моделировании кода со скоростью 1/2, с кодовым ограничением $K = 3$, 5 и 7 при длине памяти пути 32 бита. В дополнение к усеченной памяти пути, вычисления были сделаны при квантовании входных сигналов, поступающих от модулятора на восемь уровень (3 бита). Штриховые кривые дают результат качества, полученный от верхней границы для вероятности ошибки на бит, даваемой (8.2.26).

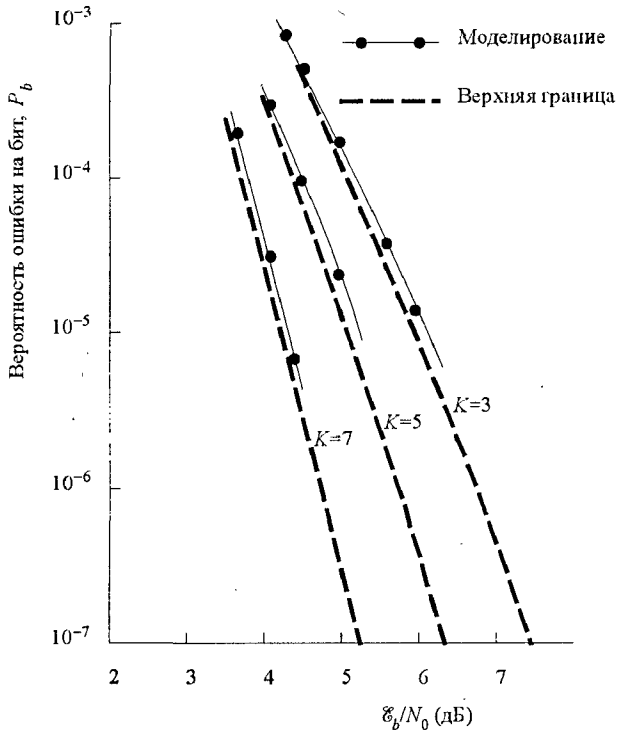


Рис. 8.2.22. Вероятность ошибки на бит для кода со скоростью 1/2 при декодировании по Витерби с 8-уровневым квантованием входных сигналов декодера и памятью пути на 32 бита [*Heller* и *Jacobs* (1971).© 1971 *IEEE*]

Заметим, что результаты моделирования близки к теоретической верхней границе, что указывает на то, что ухудшение качества, обусловленное усечением памяти и квантованием входных сигналов декодера, несущественно (0,2...0,3 дБ).

Рис. 8.2.23 иллюстрирует кривые вероятности ошибки на бит для свёрточного кода с $K = 3...8$, полученные моделированием при декодировании жёстких решений. Заметим, что для кода с $K = 8$ выигрыш от кодирования близок к 4 дБ относительно некодированной КФМ.

Влияние квантования входного сигнала декодера далее иллюстрируются на рис 8.2.24 для кода со скоростью 1/2, $K = 5$. Заметим, что трехбитовое квантование (восемь уровней) примерно на 2 дБ лучше, чем декодирование жёстких решений, что близко к безусловному пределу, получаемому при декодировании мягких решений в канале с АБГШ.

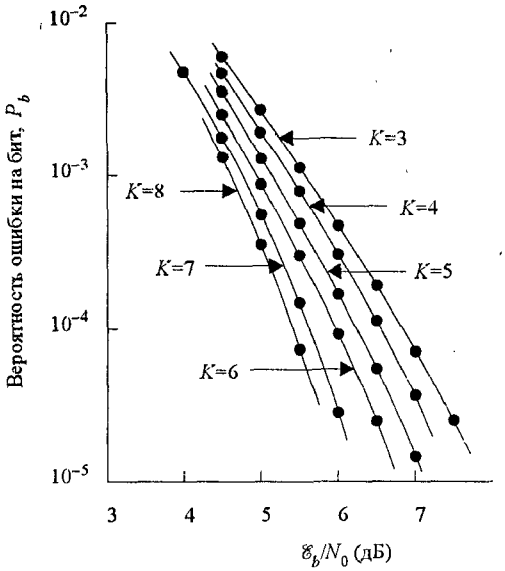


Рис. 8.2.23. Качество для кода со скоростью 1/2 при декодировании жёстких решений по Витерби и усечении памяти пути 32 битами [*Heller* и *Jacobs* (1971).© 1971 *IEEE*]

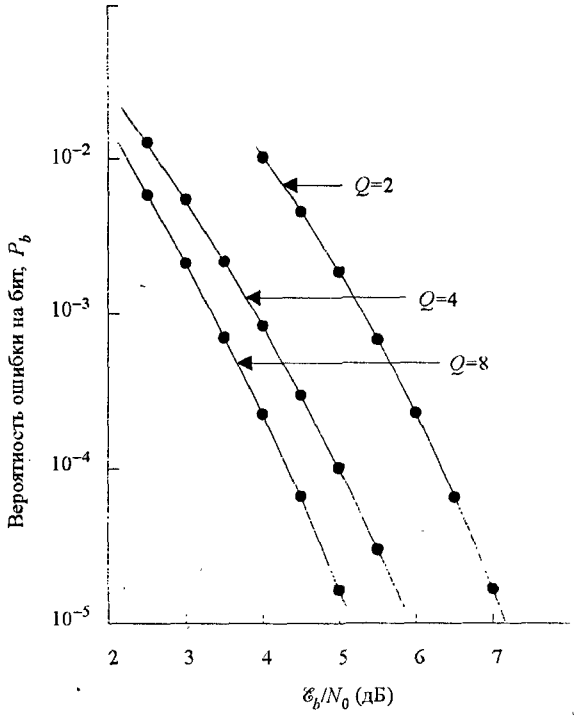


Рис. 8.2.24. Качество для кода со скоростью 1/2, K=5, при 8-, 4- и 2-уровневом квантовании на входе декодера Витерби. Усечение памяти пути = 32 битам. [*Heller* и *Jacobs* (1971).© 1971 *IEEE*]

Совместное влияние квантования сигнала и усечения памяти путей для кода со скоростью 1/2, $K = 5$ при памяти путей на 8, 16 и 32 бита и использовании 2- или 8-уровневого квантования иллюстрирует рис. 8.2.25. Из этих результатов очевидно, что ограничение памяти путей тремя кодовыми ограничениями не ведёт к серьёзному ухудшению качества.

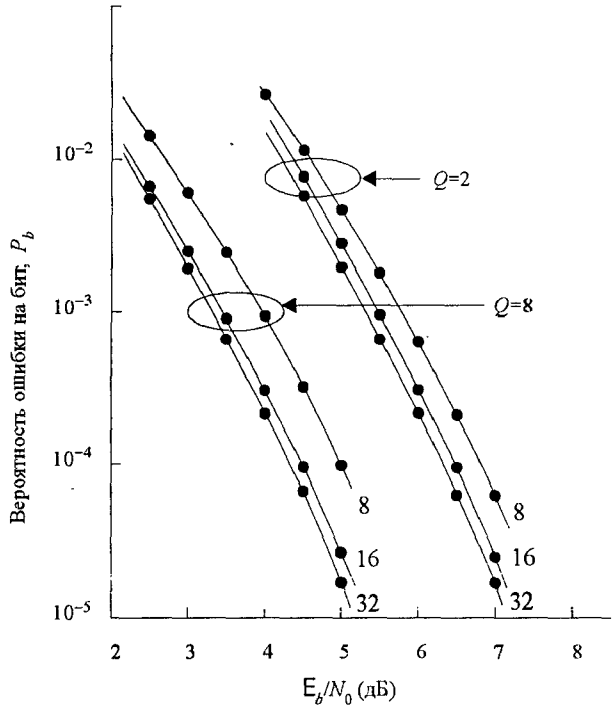


Рис. 8.2.25. Качество для кода со скоростью 1/2, K=5 при 8-, 4- и 2-уровневом квантовании на входе декодера Витерби. Усечение памяти пути = 32 битами. [*Heller* и *Jacobs* (1971). © *1971 IEEE*]

Если сигнал от демодулятора квантован более чем на два уровня, то следует рассмотреть другую проблему – расстояние между уровнями квантования. Рис. 8.2.26 иллюстрирует результаты моделирования для восьмиуровневого квантования с равномерным распределением уровней, как функцию от величины расстояния между порогами. Видим, что имеется оптимальное расстояние между порогами (примерно равное 0,5). Однако, область оптимальности достаточно широкая (0,4...0,7) так что, после начальной установки порогов, имеется малое ухудшение качества при вариации уровня в АРУ порядка ±20%.

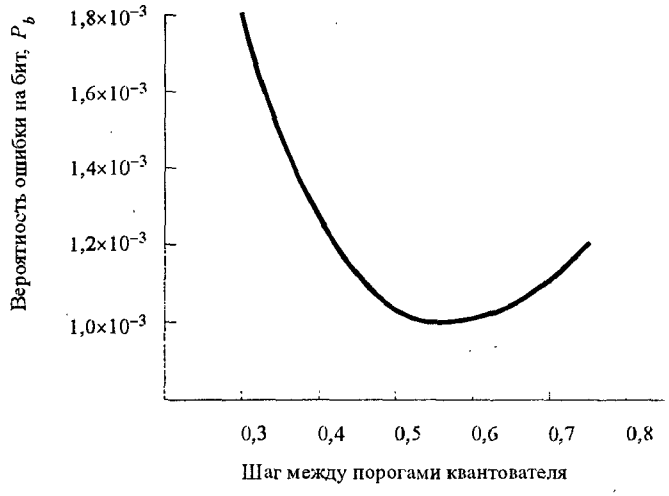


Рис. 8.2.26. Вероятность ошибок для кода со скоростью 1/2, K=5, использовании декодера Витерби при $\mathcal{E}_b/N_0$ = 3,5 дБ и 8-уровневом квантовании как функция от шага равномерного квантования на входе декодера. [*Heller* и *Jacobs* (*1971*). © 1971 *IEEE*]

Наконец, мы можем наблюдать некоторые интересные результаты в изменении качества, обусловленном изменением фазы несущей. Рис.8.2.27 показывает качество кода со скоростью 1/2, $K=7$ с восьмиуровневым квантованием и при отслеживании фазы несущей петлёй, как функцию от ОСШ γ_L.

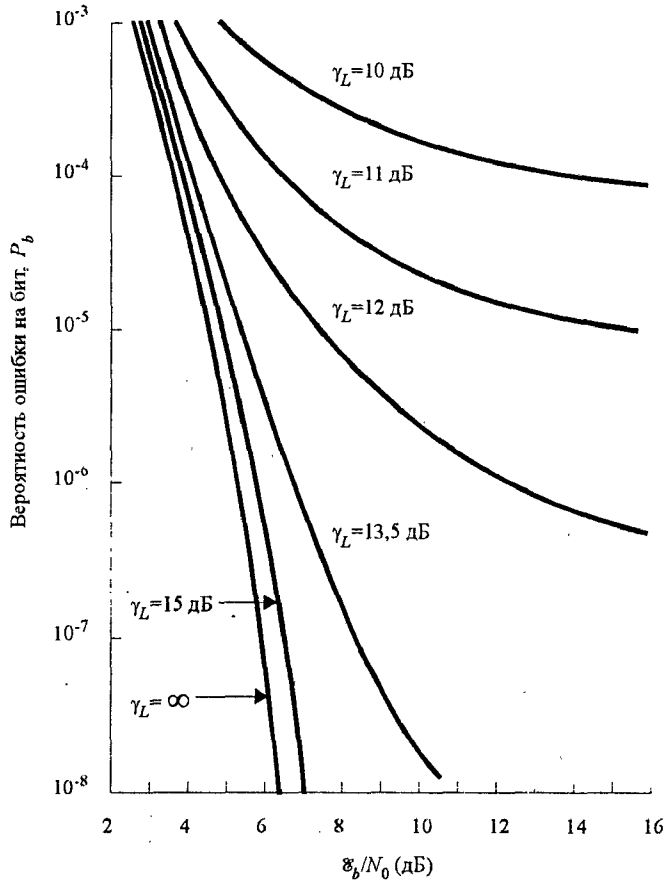


Рис. 8.2.27. Качество для кода со скоростью 1/2, K=7 с декодером Витерби при 8-уровневом квантовании как функция от γ_L – ОСШ флуктуаций фазы несущей [*Heller* и *Jacobs* (*1971*). © 1971 *IEEE*]

Напомним, что в ФАП ошибка фазы имеет дисперсию, которая обратно пропорциональна γ_L. Результаты рис.8.2.27 показывают, что ухудшение велико, если ОСШ мала ($\gamma_L < 12$ дБ)и вызывает насыщение кривой при относительно высоких значениях вероятности ошибки.

8.3. КОДИРОВАННАЯ МОДУЛЯЦИЯ ДЛЯ ЧАСТОТНО-ОГРАНИЧЕННЫХ КАНАЛОВ

При трактовке блоковых и свёрточных кодов в разделе 8.1 и 8.2, соответственно, было отмечено улучшение качества за счёт расширения полосы частот передаваемого сигнала на величину, обратно пропорциональную скорости кода. Напомним, для примера, что улучшение качества, достигаемое двоичным блоковым кодом (n,k) с декодированием мягких решений, примерно равно $10\lg\left(R_c d_{\min} - k\ln 2/\gamma_b\right)$ по сравнению с некодированной двоичной или четверичной ФМ. Для примера, если $\gamma_b = 10$, расширенный код Голея (24,12) дает выигрыш от кодирования 5 дБ. Этот выигрыш от кодирования достигается ценой удвоения полосы передаваемого сигнала и, конечно, дополнительной

оптимальностью, связанной с усложнением реализации приемника. Таким образом, кодирование является эффективным методом для размена полосы и сложность реализации на мощность передатчика. Эта ситуация применима к цифровым системам связи, которые рассчитаны для работы в области ограниченной мощности мощность, когда $R/W < 1$.

В этом разделе мы рассмотрим использование кодированных сигналов для каналов с ограничением полосы. Для таких каналов, цифровые системы связи синтезируются так, чтобы использовать эффективные по полосе системы многоуровневой фазовой модуляции, такие как АИМ, ФМ, ДФМ, КАМ и работать в области, где $R/W > 1$. Если кодирование применяется к ограниченным по полосе каналам, то желателен выигрыш качества без расширения полосы сигналов. Эту цель можно достичь увеличением числа сигналов по сравнению с соответствующими системами без кодирования, чтобы компенсировать избыточность, вносимую кодом.

Для примера, предположим, что система без кодирования, которая использует четырехфазовую ФМ, достигает $R/W = 2(\text{бит}/\text{с}\cdot\text{Гц})$ при вероятности ошибки 10^{-6}. Для этой вероятности ошибки ОСШ на бит γ_b=10,5 дБ. Мы можем попытаться уменьшить ОСШ на бит путём использования кодированных сигналов, но это надо сделать без расширения полосы. Если мы выберем скорость кода $R_c = 2/3$ это должно сопровождаться увеличением числа сигнальных точек от четырех (два бита на символ) до восьми (три бита на символ). Таким образом, код со скоростью 2/3, используемый совместно с восьми фазовой ФМ, проносит например, такое же количество данных (в единицу времени) через канал как некодированная четырехфазовая ФМ. Однако, мы напомним, что увеличение числа фаз сигнала от четырех до восьми требует дополнительно примерно 4 дБ мощности сигнала для поддержания той же вероятности ошибок. Следовательно, если кодирование приносит выгоду, выигрыш качества кода со скоростью 2/3 должен преодолеть этот 4 децибеловый штраф.

Если модуляция трактуется, как отдельная операция, независимая от кодирования, то требуется использование очень мощных кодов (большие кодовые ограничения свёрточных кодов или большая длина блокового кода) для возмещения потерь и обеспечение некоторого достаточного выигрыша от кодирования. С другой стороны, если модуляция является частью единого процесса кодирования и она рассчитывается совместно с кодом, для увеличения минимального евклидового расстояния между парами кодированных сигналов, потери от расширения ансамбля сигналов легко преодолеть и достигается достаточный выигрыш кодирования при относительно простых кодах. Ключ для подхода к этой совместной модуляции и кодирования заключается в изобретении эффективного метода, отображения кодовых битов (символов) в сигнальные точки так, чтобы максимизировать минимальное евклидово расстояние (между парами символов!). Такой метод был разработан Унгербоеком (1982) и основан на принципе отображения рядом сочленений .

Мы опишем эти принципы посредством двух примеров.

Пример 8.3.1: **сигнальное созвездие для восьмеричной ФМ**. Расчленим восьмифазовое сигнальное созвездие, показанное на рис. 8.3.1, на подобразы с возрастающими минимальными евклидовыми расстояниями. В восьмифазовом сигнальном ансамбле сигнальные точки располагаются на окружности радиуса $\sqrt{\mathscr{E}}$ и имеет минимальное взаимное расстояние

$$d_0 = 2\sqrt{\mathscr{E}}\sin\tfrac{1}{8}\pi = \sqrt{\left(2-\sqrt{2}\right)\mathscr{E}} = 0.765\sqrt{\mathscr{E}}\,.$$

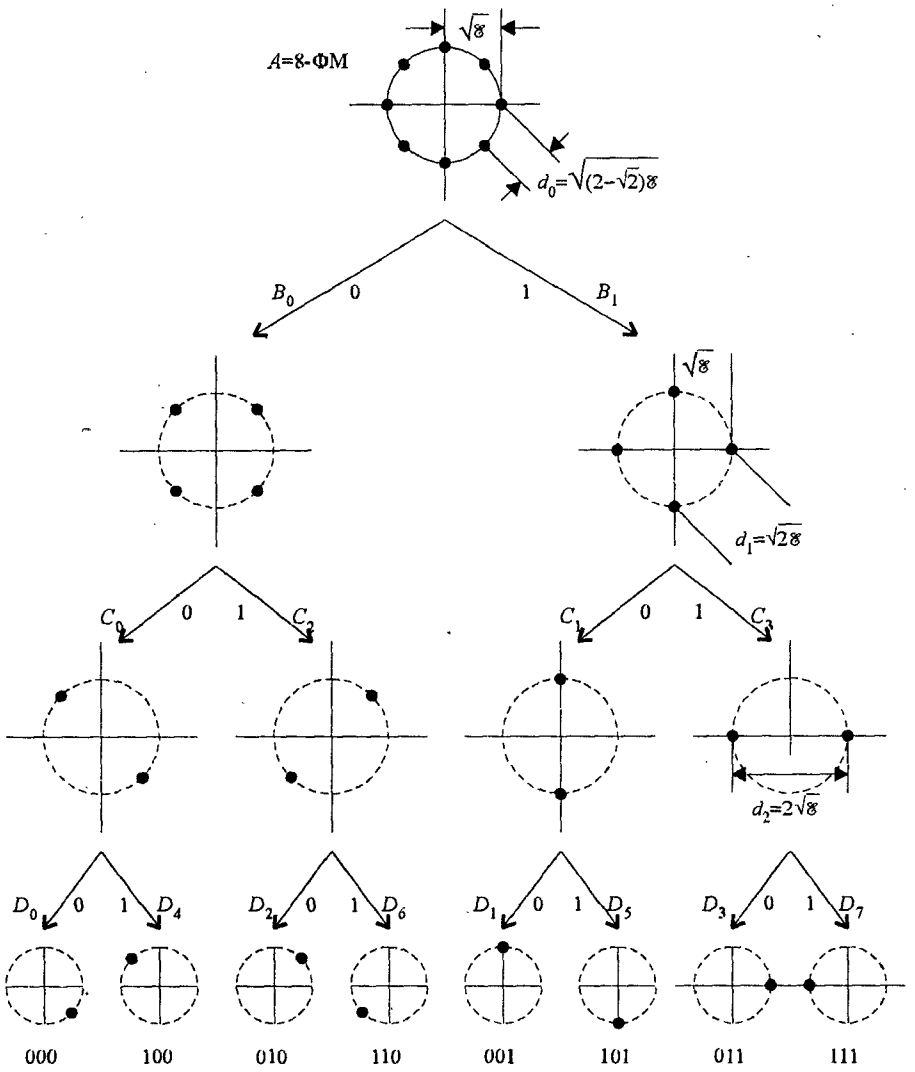


Рис. 8.3.1. Ряд расчленений для восьмеричного ансамбля сигналов ФМ

При первом расчленении восемь точек подразделяются на два подобраза из четырех точек в каждом так, что минимальное расстояние между точками увеличивается до $d_1 = \sqrt{2\mathscr{E}}$. На втором уровне расчленения каждый из двух подобразов разделяются на два подобраза из двух точек так, что минимальное расстояние увеличивается до $d_2 = 2\sqrt{\mathscr{E}}$. В результате получилось четыре подобраза с двумя точками в каждом. Наконец, последняя ступень расчленения ведет к восьми подобразам, где каждый подобраз состоит из единственной точки. Заметим, что каждый уровень расчленения увеличивает минимальное евклидовое расстояние между сигнальными точками. Результат этих трех ступеней расчленения иллюстрирует рис. 8.3.1. Путь, по которому кодовые символы отображаются в расчлененные сигнальные точки, описан ниже.

Пример 8.3.2. Сигнальные созвездия для 16-КАМ

Шестнадцатиточечное прямоугольное сигнальное созвездие показанное на рис. 8.3.2, сначала делится на два подобраза путём назначения альтернативных точек в каждом подобразе, как показано на рисунке. Таким образом, минимальное расстояние между точками увеличивается при первом расчленении с $2\sqrt{\mathscr{E}}$ до $2\sqrt{2\mathscr{E}}$. Дальнейшее расчленение двух подобразов ведет к большому увеличению евклидового расстояния между

сигнальными точками, как показано на рис. 8.3.2. Интересно заметить, что для прямоугольного сигнального созвездия, каждый уровень расчленения увеличивает минимальное евклидово расстояние на $\sqrt{2}$, то есть $d_{i+1}/d_i = \sqrt{2}$ для всех i.

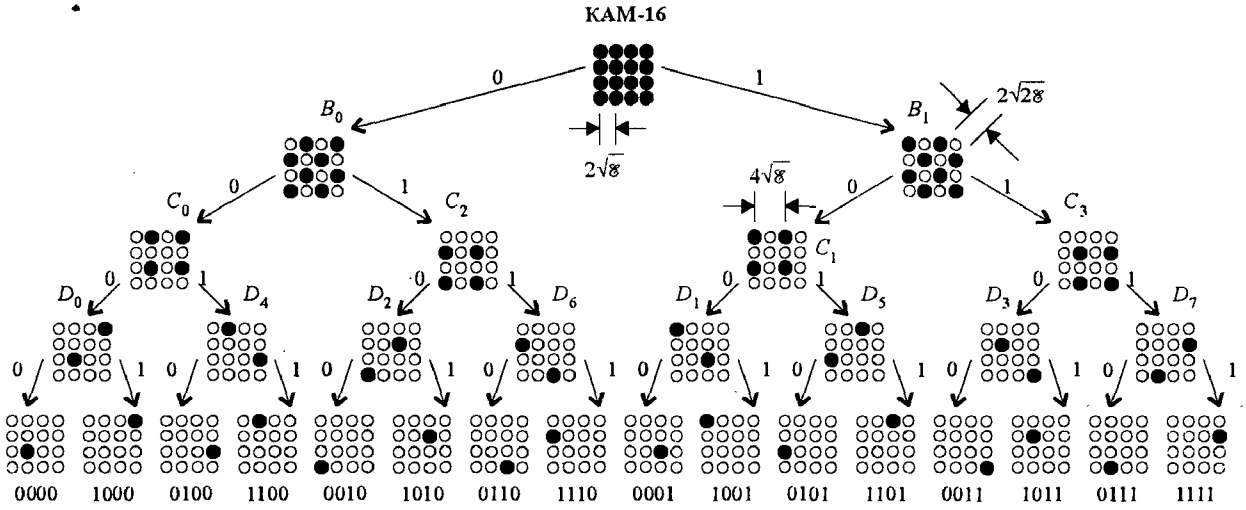


Рис. 8.3.2. Ряд расчленений для ансамбля сигналов КАМ-16

В этих двух примерах расчленения ведется до тех пор, пока каждый подобраз содержит только единственную точку. В общем, это не необходимо. Например, сигнальное созвездие 16 точечной КАМ можно расчленить только дважды, чтобы получить четыре подобраза с четырьмя точками в каждом. Аналогично сигнальное созвездие восьми фазовой ФМ можно расчленить дважды, чтобы получить четыре подобраза с двумя точками в каждом.

Степень до какой сигнал расчленяется зависит от характеристик кода. В общем процесс кодирования выполняется так, как показано на рис. 8.3.3.

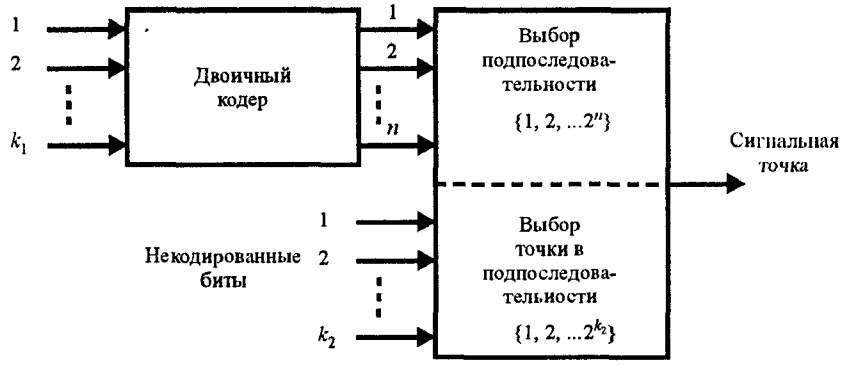


Рис. 8.3.3. Общая структура комбинированного кодера/модулятора

Блок из m информационных символов делится на две группы длиной k_1 и k_2. k_1 символов кодируются в n символов, в то время как k_2 символа остаются не кодированными. Затем n символа кодера используются для выбора одного из 2^n возможных подобразов в расчлененном ансамбле сигналов, в то время как k_2 символа используются для выбора одной из 2^{k_2} сигнальных точек в каждом подобразе. Если $k_2 = 0$, все m информационных символа кодируются.

Пример 8.3.3. Рассмотрим использование свёрточного кода со скоростью 1/2, показанного на рис. 8.3.4 (а) для кодирования некоторого информационного символа, в то время как второй информационный символ остается не кодированным.

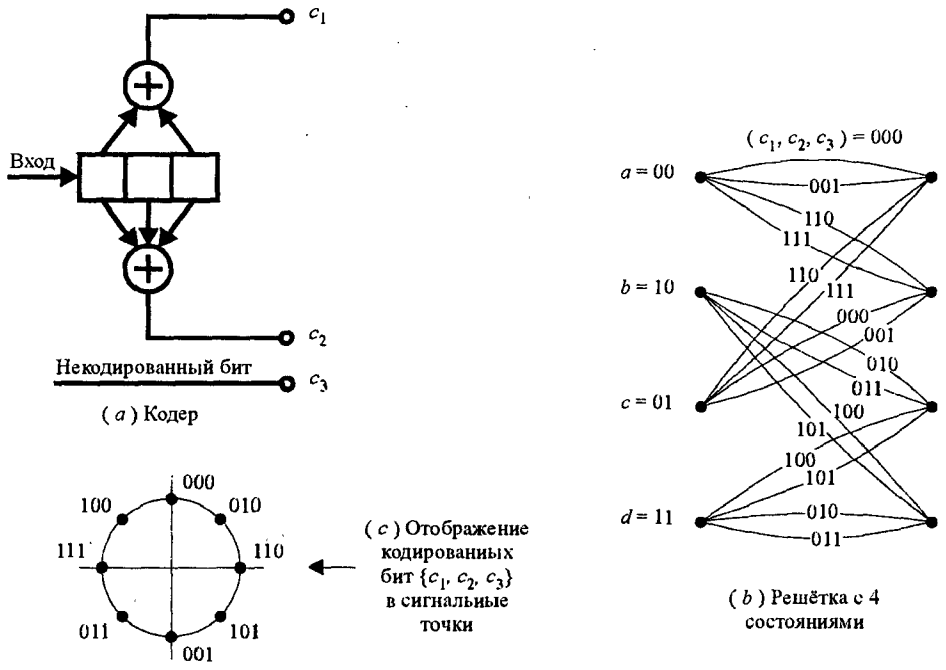


Рис. 8.3.4. Решётка с 4 состояниями для кодированной модуляции восьмеричной ФМ

Используем их в объединении с восьми точечным сигнальным созвездием, например, восьми фазовой ФМ или восьми точечной КАМ. Два кодированных символа используются для выбора одного из четырех подобразов сигнального созвездия, в то время как оставшийся информационный символ используются для выбора одной из двух точек внутри каждого подобраза. В этом случае $k_1 = 1$ и $k_2 = 1$. Решётка с четырьмя состояниями показанная на рис. 8.3.4(b), является в основе своей решёткой для свёрточного кода со скоростью 1/2 с добавлением параллельных путей в каждой позиции для размещения некодированных символов c_3. Таким образом, кодированные символы (c_1, c_2) используются для выбора одного из четырех подобразов, каждый из которых содержит две сигнальные точки, а не кодированные символы используются для выбора одной из двух сигнальных точек внутри каждого подобраза. Заметим, что сигнальные точки внутри подобраза удаленны на расстояние $d_2 = 2\sqrt{\varepsilon}$. Таким образом, расстояние Евклида между параллельными путями равно d_2. Отображение кодовых символов (c_1, c_2, c_3) сигнальными точками иллюстрируется на рис. 8.3.4(c). В качестве альтернативной схемы кодирования мы можем использовать свёрточный кодер со скоростью 2/3 и кодировать 2 информационных символа так, как показано на рис 8.3.5. Это кодирование ведет к решётке с восьмью состояниями и обеспечивает лучшее качество, но требует более сложную реализацию декодера, как описано ниже.

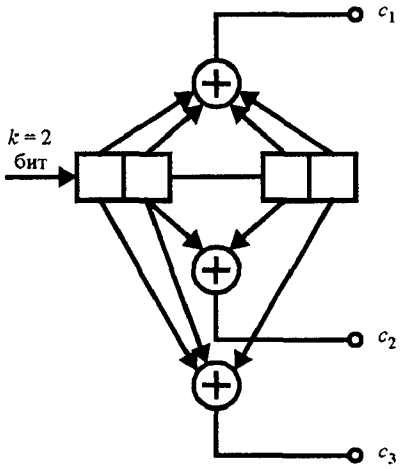
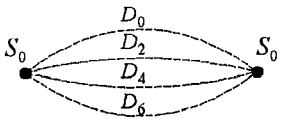


Рис. 8.3.5. Свёрточный кодер со скоростью 2/3 для кодирования обоих информационных символов

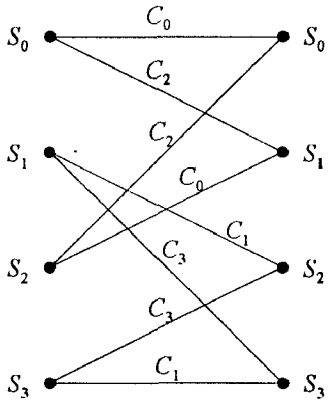
Как блоковый, так и свёрточный коды можно использовать в объединении с расчлененным сигнальным созвездием. В общем, свёрточные коды обеспечивают сравнимый выигрыш кодирования относительно блоковых кодов, а имеющийся в распоряжении алгоритм Витерби приводит к несложной реализации декодирования мягких решений. Из этих соображений мы ограничим наше обсуждение свёрточными кодами (линейными решётчатыми кодами), а в более общем случае (нелинейными) решётчатыми кодами.

Решётчатая кодированная модуляция. Рассмотрим использование сигнального созвездия 8-и уровневой ФМ в объединении с решётчатым кодом. В качестве образца при измерении выигрыша кодирования используем не кодированную четырехфазовую ФМ (4 ФМ). Не кодированная 4 ФМ использует сигнальные точки либо подобраза B_0 либо B_1 на рис 8.3.1, из которого видно, что минимальное расстояние сигнальных точек равно $\sqrt{2\mathscr{E}}$. Заметим, что этот сигнал соответствует тривиальной решётке с одним состоянием и четырьмя параллельными переходами состояний, как показано на рис. 8.3.6(а).

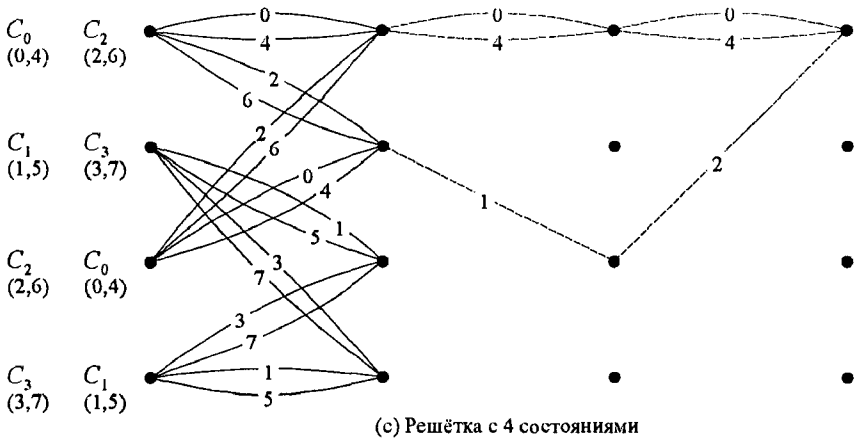
Подобразы D_0, D_2, D_4 и D_6 использованы как сигнальные точки с целью иллюстрации. Для кодированной 8 ФМ модуляции мы можем использовать решётку с четырьмя состояниями, как показано на рис. 8.3.6(б). Заметим, что каждая ветвь решётки соответствует одному из четырех подобразов C_0, C_1, C_2 или C_3. Для восьми точечного созвездия каждый из подобразов C_0, C_1, C_2 и C_3 содержит две сигнальные точки. Следовательно, пподобраз C_0 содержит две сигнальные точки, соответствующие битам (000, 100) или (0, 4) в восьмеричной записи. Аналогично, C_2 содержит две сигнальные точки, соответствующие битам (010, 110) или (2, 6) в восьмеричной записи. C_1 содержит две сигнальные точки, соответствующие битам (001, 101) или (1, 5) в восьмеричной записи и C_3 содержит две сигнальные точки, соответствующие битам (011, 111) или (3, 7) в восьмеричной записи. Таким образом, каждый переход в решётке из четырех состояний содержит два параллельных пути, как показано более подробно на рис. 8.3.6 (*c*). Заметим, что любые два сигнальных пути, которые расходятся из одного состояния и сливаются в то же состояние после более одного перехода имеем квадрат расстояний Евклида $d_0^2 + 2d_1^2 = d_0^2 + d_2^2$ между ними.



(*a*) Решётка с 1 состоянием



(*b*) Решётка с 4 состояниями



(c) Решётка с 4 состояниями

Рис. 8.3.6. Решётка для некодированной 4ФМ и решётчатый код для сигиалов 8ФМ

Для примера, сигнальные пути 0, 0, 0 и 2, 1, 2 разделены на $d_0^2 + d_2^2 = \left[(0.765)^2 + 4\right]\mathscr{E} = 4.585\mathscr{E}$. С другой стороны, квадрат расстояния Евклида между параллельными переходами равно $d_2^2 = 4\mathscr{E}$. Следовательно, минимальное расстояние Евклида между путями, которые расходятся из какого либо состояния и сходятся в то же состояние, равно для решётки с четырьмя состояниями $d_2 = 2\sqrt{\mathscr{E}}$. Это минимальное расстояние в решётчатом коде названо *свободным евклидовым расстоянием* и обозначаются $D_{\text{св}}$.

В решётке из четырех состояний на рис. 8.3.6 (б) $D_{\text{св}} = 2\sqrt{\mathscr{E}}$. Если сравнить с расстоянием Евклида $d_0 = \sqrt{2\mathscr{E}}$ для не кодированной 4 ФМ модуляции, мы видим, что решётчатый код с четырьмя состояниями дает выигрыш кодирования 3 дБ. Мы хотим подчеркнуть, что решётчатый код с четырьмя состояниями, иллюстрированный на рис. 8.3.6(б), оптимален в том смысле, что он обеспечивает наибольшое свободное Евклидово расстояние. Ясно, можно конструировать много других решётчатых кодов с четырьмя состояниями, включая и тот, который показан на рис. 8.3.7, которые четыре отсчётных перехода из одного состояния до всех остальных состояний. Однако, ни этот код, ни любой из других возможных решётчатых кодов с четырьмя состояниями не даст большее значение $D_{\text{св}}$.

Для решётки с четырьмя состояниями рис. 8-3-6 (б) $D_{\text{св}} = 2\sqrt{q}$. Если сравнить с расстоянием Евклида $d_0 = \sqrt{2q}$ для некодированной ЧФМ модуляции мы видим, что решётчатый код с четырьмя состояниями даёт выигрыш кодирования 3 дБ.

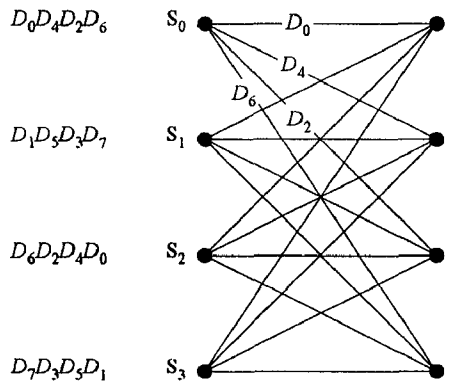


Рис. 8.3.7. Альтернативный решётчатый код с 4 состояниями

Конструирование оптимального решётчатого кода с четырьмя состояниями для восьми точечного созвездия было выполнено на основе следующих эвристических правил:

a) параллельные переходы (когда они возникают) назначаются сигнальным точкам, разделенным максимальным расстоянием Евклида, например $d_2 = 2\sqrt{\mathscr{E}}$ для 8 ФМ в четырех подобразах C_0, C_1, C_2, C_3.

b) переход, начинающийся и сходящийся в некоторое состояние назначается подобразом (C_0, C_2) или (C_1, C_3), которые имеют максимальное расстояние $d_1 = \sqrt{2\mathscr{E}}$.

c) сигнальные точки возникают с одинаковой частотой.

Заметим, что правила (*a*) и (*b*) гарантируют, что Евклидово расстояние связанное с одним и многими путями, которые выходят из определенного состояния и сходятся в это состояние, превосходят Евклидово расстояние для не кодированной 4 ФМ. Правило (*c*) гарантирует, что решётчатый код будет иметь регулярную структуру.

Хотим указать на то, что специальное отображение кодовых символов в сигнальные точки, как иллюстрировано на рис. 8.3.1, где восемь сигнальных точек представлено в эквивалентную двоичную форму, существенно. Можно разработать иные отображения путём перестановки подобразов таким путём, чтобы сохранить основное свойство – увеличение минимального расстояния среди подобразов.

В решётчатом коде с четырьмя состояниями параллельные переходы разделяются евклидовым расстоянием $2\sqrt{\mathscr{E}}$, что тоже равно $D_{\text{св}}$. Таким образом, выигрыш кодирования в 3 дБ ограничивается расстоянием параллельных переходов. Больший выигрыш в качестве относительно не кодированной 4 ФМ можно достичь путём использования решётчатых кодов с большим числом состояний, что позволяет исключить параллельные переходы. Решётчатые коды с восемью или большим числом состояний следует использовать различимые переходы для получения больших значений $D_{\text{св}}$. Например, на рис. 8.3.8 мы иллюстрируем решётчатый код с восемью состояниями, разработанный Унгербоеком (1982 год) для 8 ФМ сигнального созвездия.

Для максимизации свободного Евклидова расстояния переходы состояний были определены с учетом трех базовых правил, данных выше. В этом случае можно заметить, что минимум квадрата Евклидова расстояния равно

$$D_{\text{св}}^2 = d_0^2 + 2d_1^2 = 4.585\mathscr{E},$$

что, по сравнению с расстоянием $d_0^2 = 2\mathscr{E}$ для некодированной 4 ФМ представляет выигрыш 3,6 дБ. Унгербоек (1982...1987) также нашел решётчатые коды со скоростью 2/3 и с 16, 32, 64, 128 и 256 состояниями, которые достигают выигрыш кодирования от 4 до 5,75 дБ для 8 ФМ модуляции.

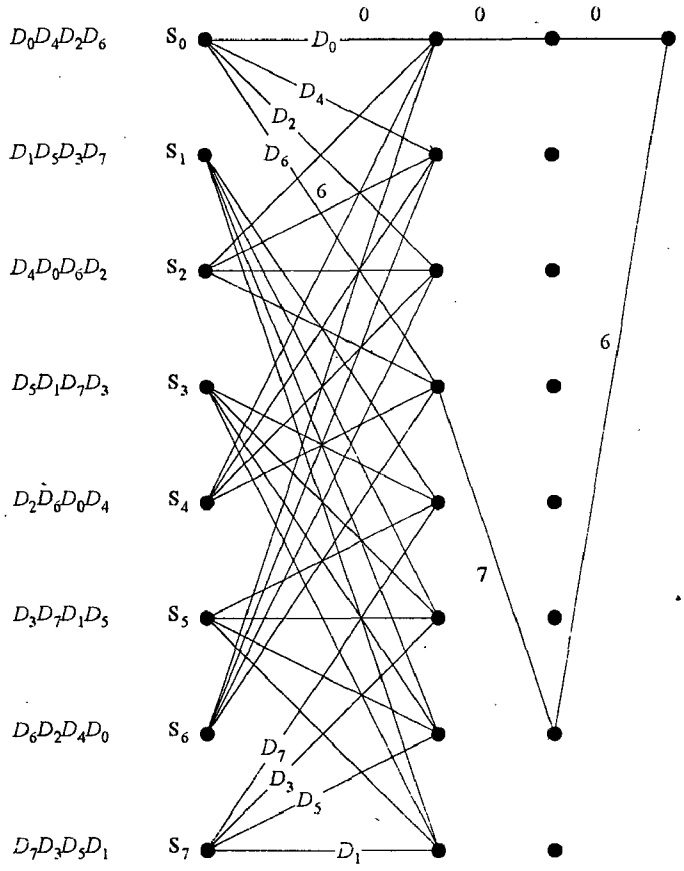


Рис. 8.3.8. Решётчатый код с 8 состояниями для кодированной модуляции 8ФМ

Базовый принцип расчленения ансамбля легко расширить на большие сигнальные созвездия с ФМ, которые дают большую частотную эффективность. Например, 3 (бит/с·Гц) можно достичь или не кодированным 8 ФМ или кодированной модуляцией 16 ФМ с решётчатым кодом. Унгербоек (1987) предложил решётчатые коды и рассчитал выигрыш кодирования, достигнутый посредством простых свёрточных кодов со скоростью 1/2 и 2/3 для 16 ФМ сигнального созвездия. Эти результаты суммированы ниже.

Декодирование Витерби мягких решений для решётчато-кодированной модуляции выполняется двумя ступенями. Поскольку каждая ветвь решётки соответствует сигнальному подобразу, то – первая ступень декодирования сводится к определению наилучшей точки сигнала внутри каждого подобраза, то есть точку в каждом подобразе, которая ближе по расстоянию в к принятой точке. Мы можем это назвать *декодированием под образов*. На второй ступени сигнальные точки, выбранные в каждом подобразе и их метрики квадратов расстояний используются для соответствующей ветви в алгоритме Витерби для определения пути сигнала в кодовой решётке, который имеет минимальную сумму квадратов расстояний от принимаемых сигнальных последовательностей (зашумленный выход канала).

Вероятность ошибки решётчато-кодированных сигналов в присутствии аддитивного гауссовского шума можно рассчитать, следуя процедуре, описанной в разделе 8.2 для свёрточных кодов. Напомним, что эта процедура включает расчет вероятности ошибки для всех различных событий, приводящих к ошибке и к суммированию вероятностей этих событий для получения объединенной верхней границы для вероятности ошибки

декодирования при первом пересечении. Заметим, однако, что при больших ОСШ, вероятность ошибки при первом пересечении в основном определяется лидирующим слагаемым, который имеет минимальное расстояние $D_{св}$. Следовательно, при больших ОСШ вероятность ошибки при первом пересечении хорошо аппроксимируется так:

$$P_e \approx N_{св} Q\left(\sqrt{\frac{D_{св}^2}{2N_0}}\right), \qquad (8.3.1)$$

где $N_{св}$ означает число сигнальных последовательностей с расстоянием $D_{св}$, которые выходят из определённого состояния и возвращаются в то же состояние после одного или больше переходов.

При вычислении выигрыша кодирования, достигаемого посредством решётчето-кодированной модуляции мы обычно сосредотачиваемся на выигрыш, достигаемый путём увеличения $D_{св}$ и пренебрегаем влиянием $N_{св}$. Однако, решёточные коды с большим числом состояний могут привести к большим значениям $N_{св}$, что нельзя игнорировать при оценивании всего выигрыша кодирования.

В дополнении к решётчето-кодированной ФМ модуляции, описанной выше, мощные решёточатые коды были также разработаны для сигнальных созвездий АМ и КАМ. Особенно важен для практики класс решётчато-кодированных двухмерных прямоугольных сигнальных созвездий. Рис. 8.3.9 иллюстрирует эти сигнальные созвездия для М-КАМ, где М=16, 32, 64 и 128.

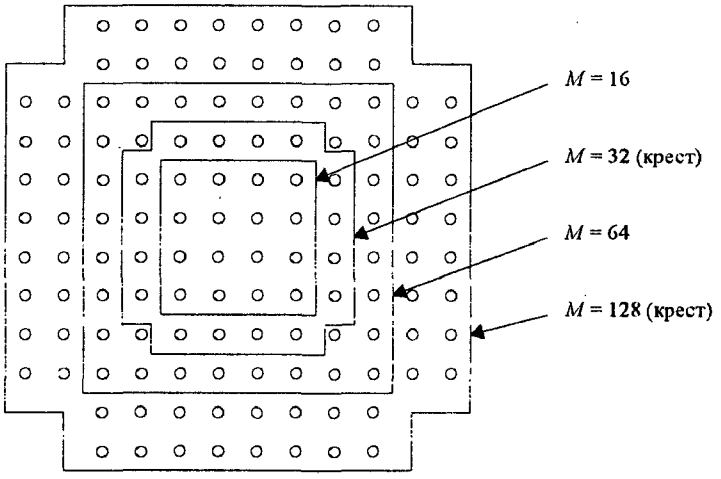


Рис. 8.3.9. Прямоугольное двухмерное (КАМ) сигнальное созвездие

Созвездия с М=2 и 128 имеют крестообразные формы и их иногда называют *крест созвездиями*. Ниже расположенные прямоугольные сетки, содержащие сигнальные точки М-КАМ, названы *решётками типа* Z_2 (индекс указывает на размерность пространства). Если расчленение ансамбля применить к этому классу созвездий, минимальное евклидово расстояние между последовательными расчленениями равно $d_{i+1}/d_i = \sqrt{2}$ для всех i, как мы видели прежде в примере 8.3.2.

Рис. 8.3.10 иллюстрирует решётчатый код с восьмью состояниями, который можно использовать в прямоугольном сигнальном созвездии М-КАМ, $M = 2^k$, где k=4, 5, 6 ... и так далее.

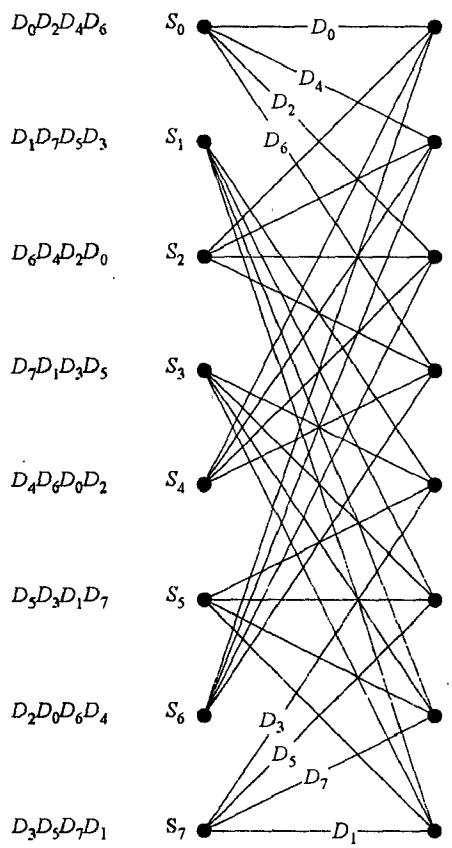


Рис. 8.3.10. Решётка с 8 состояниями для прямоугольных сигнальных созвездий с КАМ

С решёткой из восьми состояний мы связываем восемь сигнальных подобразов так, что подходит любой из M-КАМ сигнальных ансамблей при $M \geq 16$. Для $M = 2^{m+1}$ два входных символа $(k_1 = 2)$ кодируются в $n = 3$ $(n = k_1 + 1)$ символа, которые используются для выбора одного из восьми состояний, соответствующих восьми подобразам. Дополнительные $k_2 = m - k_1$ входных символа используются для выбора сигнальных точек внутри подобраза и они приводят к параллельным переходам в решётке из восьми состояний. Таким образом, 16 КАМ включает два параллельных перехода в каждой ветви решётки. В более общем виде, выбор $M = 2^{m+1}$-точечного сигнального созвездия КАМ подразумевает, что решётки с восьмью состояниями содержат 2^{m-2} параллельных переходов в каждой ветви.

Задание сигнальных подобразов для переходов основывается на том же наборе базовых (эвристических) правил, описанных выше 8 ФМ сигнального созвездия. Так четыре (ветвевых) переходов, начинающихся от или входящих в то же состояние, задаются подобразами D_0, D_2, D_4, D_6 или D_1, D_3, D_5, D_7. Параллельные переходы задают сигнальные точки, содержащиеся внутри соответствующих подобразов. Этот решётчатый код с восьмью состояниями обеспечивает выигрыш от кодирования 4 дБ. Евклидово расстояние параллельных переходов превышает свободное евклидово расстояние и, следовательно, качество кода не ограниченно параллельными переходами.

Решётчатые коды с большими размерами для М-КАМ обеспечивают даже большие выигрыши от кодирования. Для примера, решётчатые коды с 2^v состояниями для

$M = 2^{m+1}$ сигнального созвездия КАМ можно сконструировать путём свёрточного кодирования k_1 входных символов в $k_1 + 1$ выходных символов. Так, для этой цели используется свёрточный код со скоростью $R_c = k_1 / (k_1 + 1)$. Обычно выбор $k_1 = 2$ обеспечивает достаточную долю общего достигаемого выигрыша от кодирования. Дополнительные $k_2 = m - k_1$ входных символов не кодируются, и они передаются в каждом сигнальном интервале, путём выбора символьных точек внутри подобраза.

Таблицы 8.3.1...8.3.3, взятые из статей Унгербоека (1987) дают суммарный выигрыш от кодирования, достигаемый при помощи решётчато – кодированной модуляции. Табл. 8.3.1 суммирует выигрыши от кодирования, достигаемые при решётчато – кодированной (одномерной) модуляцией с АМ со скоростями решётчатых кодов 1/2. Заметим, что выигрыш от кодирования посредством решётчатого кода с 128 состояниями, равен 5,8 дБ для восьми уровневой АМ, что близко к предельной скорости канала R_0 и теряет 4 дБ относительно предела пропускной способности канала для частости ошибки в области $10^{-6}...10^{-8}$. Мы можем также видеть, что число путей $N_{св}$ с евклидовым расстоянием $D_{св}$ становится больше с увеличением числа состояний.

Табл. 8.3.2 показывает выигрыш кодирования для решётчато-кодированной 16 ФМ. Снова видим, что выигрыш кодирования для восьми и более ступеней решётки превышает 4 дБ относительно некодированной 8 ФМ. Простой код со скоростью 1/2 и 128 ступенями решётки дает выигрыш 5,33 дБ.

Табл. 8.3.3 содержит выигрыш кодирования, получаемый решётчато – кодированными сигналами КАМ. Относительно простой решётчатый код со скоростью 2/3 и 128 ступенями решётки дает выигрыш около 6 дБ при m=3 и 4.

Таблица 8.3.1. Выигрыши кодирования для решётчато-кодированных АМ сигналов

Число состояний	k_1	Скорость кода $k_1/(k_1+1)$	m=1 Выигрыш кода (дБ) 4АМ относительно некодированной 2АМ	m=2 Выигрыш кода (дБ) 8АМ относительно некодированной 4АМ	$m \to \infty$ $m \to \infty$ Асимптотический выигрыш кода (дБ)	$m \to \infty$ $N_{св}$
4	1	1/2	2,55	3,31	3,52	4
8	1	1/2	3,01	3,77	3,97	4
16	1	1/2	3,42	4,18	4,39	8
32	1	1/2	4,15	4,91	5,11	12
64	1	1/2	4,47	5,23	5,44	36
128	1	1/2	5,05	5,81	6,02	66

Источник: *Ungerboek* (1987)

Таблица 8.3.2. Выигрыши кодирования для решётчато-кодированных 16-ФМ сигналов

Число состояний	k_1	Скорость кода $k_1/(k_1+1)$	m=3 Выигрыш кода (дБ) 16ФМ относительно некодированной 8ФМ	$m \to \infty$ $N_{св}$
4	1	1/2	3,54	4
8	1	1/2	4,01	4
16	1	1/2	4,44	8
32	1	1/2	5,13	8
64	1	1/2	5,33	2
128	1	1/2	5,33	2
256	2	2/3	5,51	8

Источник: *Ungerboek* (1987)

Таблица 8.3.3. Выигрыши кодирования для решётчато-кодированных КАМ сигналов

Число состояний	k_1	Скорость кода $k_1/(k_1+1)$	m=3 Выигрыш кода (дБ) 16КАМ относительно некодированной 8КАМ	m=4 Выигрыш кода (дБ) 32КАМ относительно некодированной 16КФМ	m=5 Выигрыш кода (дБ) 64КАМ относительно некодированной 32КАМ	$m \to \infty$ Асимптотический выигрыш кода (дБ)	$m \to \infty$ $N_{св}$
4	1	1/2	3,01	3,01	2,80	3,01	4
8	2	2/3	3,98	3,98	3,77	3,98	16
16	2	2/3	4,77	4,77	4,56	4,77	56
32	2	2/3	4,77	4,77	4,56	4,77	16
64	2	2/3	5,44	5,44	4,23	5,44	56
128	2	2/3	6,02	6,02	5,81	6,02	344
256	2	2/3	6,02	6,02	5,81	6,02	44

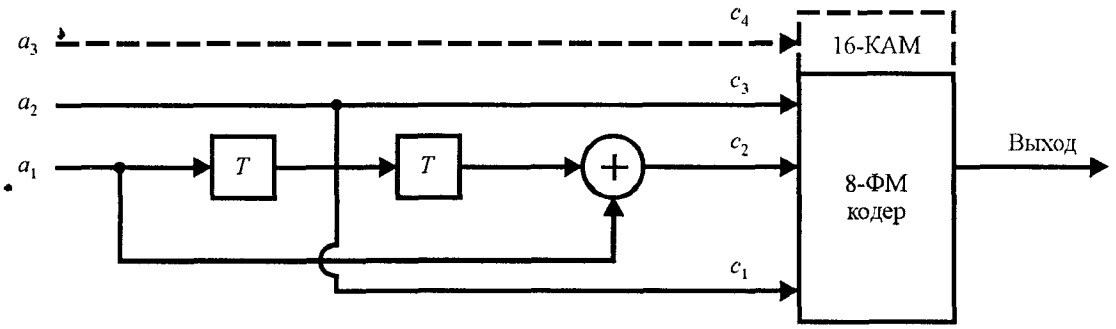
Источник: *Ungerboek* (1987)

Результаты в этих таблицах ясно иллюстрируют существенные выигрыши кодирования, достигаемые относительно простыми решётчатыми кодами. Выигрыш от кодирования в 6 дБ близок к результату для предельной скорости R_0 для рассматриваемого сигнального ансамбля. Дополнительный выигрыш, который может вести к передаче вблизи границы пропускающей способности канала трудно получить без достаточного увеличения сложности техники кодирования/декодирования.

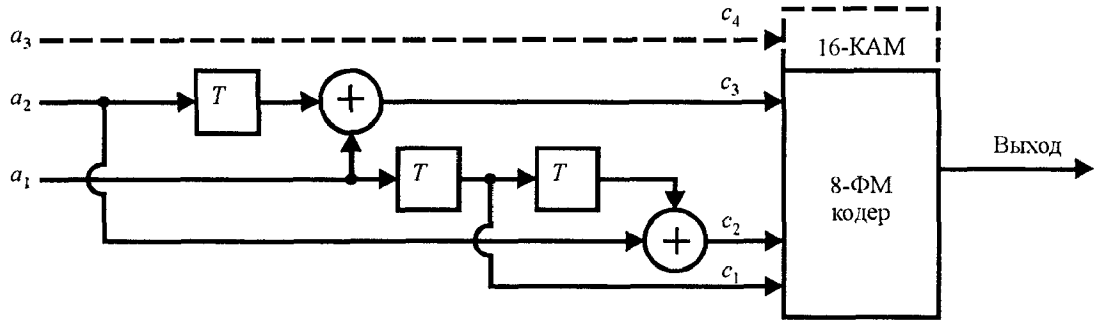
Поскольку пропускная способность канала определяет окончательный предел качества кодирования, мы можем подчеркнуть, что непрерывное расчленение больших ансамблей сигналов быстро ведет к разделению сигнальных точек внутри некоторого подобраза, что расширяет свободное евклидово расстояние кода. В таких случаях параллельные переходы больше не являются ограничивающим фактором для $D_{св}$. Обычно расчленение на восемь подобразов достаточно для получения выигрыша кодирования 5-6 дБ при помощи решётчатых кодов со скоростью 1/2 или со скоростью 2/3 с 64 или 128 ступенями решётки, как указанно в табл. 8.3.1...8.3.3. Свёрточные кодеры для линейных решётчатых кодов, иллюстрируемые в табл. 8.3.1...8.3.3 для сигнальных созвездий M-АМ, M-ФМ и M-КАМ даны в статьях Унгербоека (1982...1987). Кодеры могут быть реализованы с обратной связью или без неё. Например, рис. 8.3.11 иллюстрирует три свёрточных кодера без обратной связи, соответствующие решётчатым кодам с 4, 8, 16 состояниями для сигнальных созвездий 8 ФМ и 16 КАМ.

Эквивалентная реализация этих решётчатых кодов, основанная на систематических свёрточных кодерах с обратной связью даны на рис. 8.3.12. Обычно систематические свёрточные кодеры предпочтительны в практических разработках.

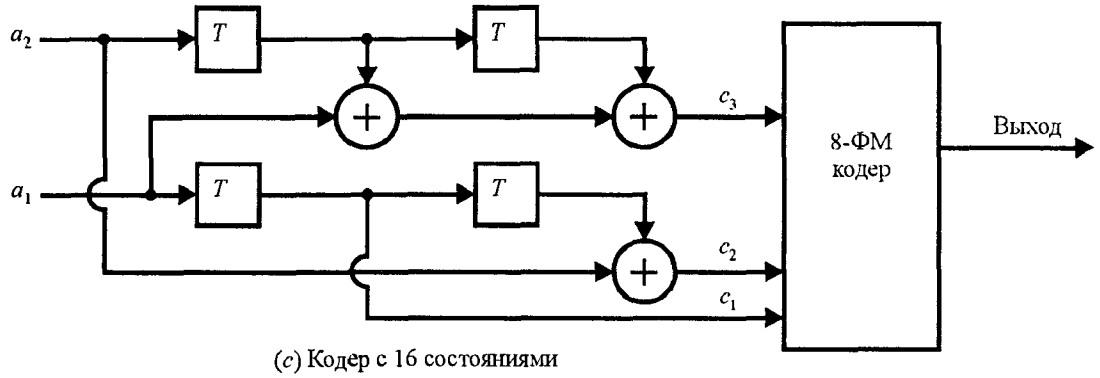
Важнейшая проблема для линейных свёрточных кодов заключается в том, что ансамбль модулированных сигналов обычно не инвариантен к изменению фазы. Это ставит проблему в практических разработках, где обычно применяется дифференциальное кодирование, так чтобы преодолеть фазовые флуктуации в случаях, когда приемник должен восстановить фазу несущей после временных ослаблений сигналов. Проблема постоянства фазы и дифференциального кодирования/декодирования была решена Вайем (1984 a,b), который изобрел линейные и нелинейные решётчатые коды, которые инварианты к вращению фазы соответственно на 180^0 или 90^0. Для примера, рис.8.3.13 иллюстрирует нелинейный свёрточный кодер с восемью состояниями для прямоугольного сигнального созвездия 32-КАМ, который инвариантен к вращению фазы на 90^0. Эти решётчатые коды были приняты как международный стандарт для линейных модемов телефонной связи со скоростями 9600 и 14000 бит/с (высокоскоростные).



(*a*) Кодер с 4 состояниями



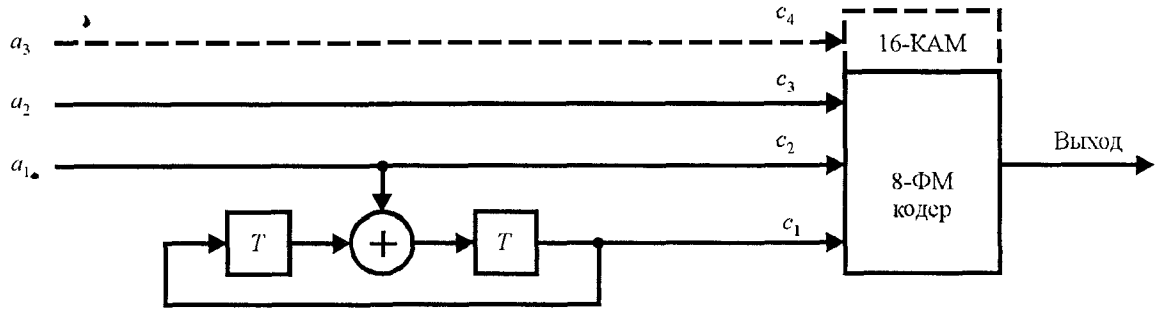
(*b*) Кодер с 8 состояниями



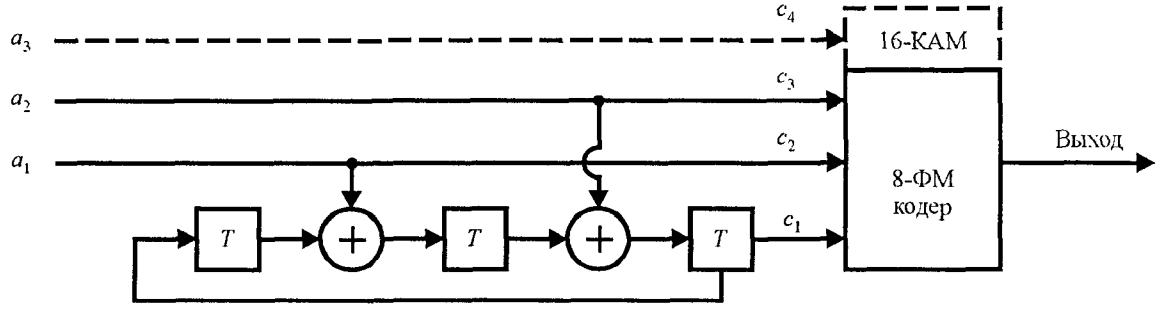
(*c*) Кодер с 16 состояниями

Рис. 8.3.11. Минимальный кодер без обратной связи для сигналов 8-ФМ и 16-КАМ [*Ungerboek* (1982), © 1982 *IEEE*]

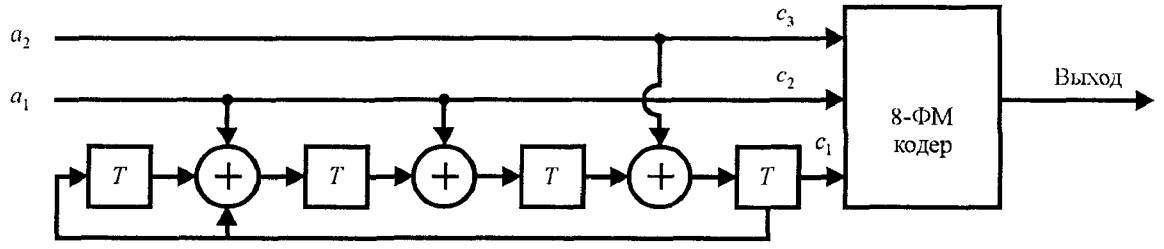
Схемы решётчато-кодированной модуляции были также разработаны для многомерных сигналов. В практических системах многомерные сигналы передаются как последовательность или одномерных (АМ) или двухмерных (КАМ) сигналов. Решётчатые коды, основанные на 4, 8, и 16 сигнальных созвездий были сконструированы и некоторые из этих кодов были внедрены в имеющиеся в распоряжении коммерческих модемов. Потенциальное преимущество решётчато – кодированных многомерных сигналов заключается в том, что мы можем использовать узко избирательные двухмерные сигнальные созвездия, что позволяет осуществлять выбор между выигрышем кодирования и сложностью реализации. Статьи Вая (1987), Унгербоека (1987), Гершо и Лоуренса (1984) и Форни и др. (1984) трактуют многомерные сигнальные созвездия для решётчато-кодовой модуляции.



(*a*) Кодер с 4 состояниями



(*b*) Кодер с 8 состояниями



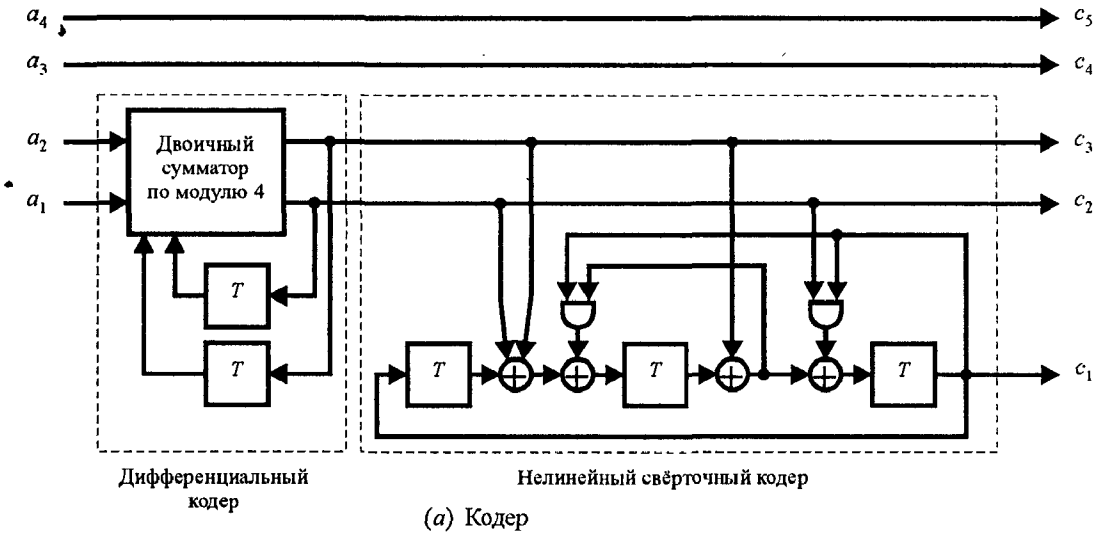
(*c*) Кодер с 16 состояниями

Рис. 8.3.12. Эквивалентная реализация систематических свёрточных кодов с обратной связью для 8ФМ и 16КАМ
[*Ungerboek* (1982). © 1982 *IEEE*]

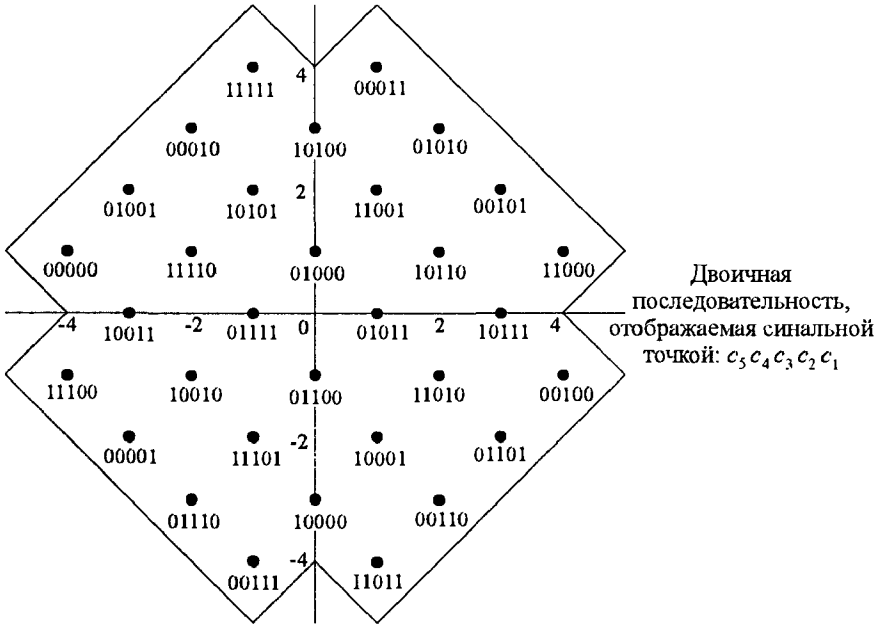
Наконец мы хотим напомнить, что новая техника синтеза решётчато-кодированной модуляции, основанная на решётках и объединении решёток были описаны Кальдербанком и Слоэном (1987) и Форни (1988). Это метод конструирования решётчатых кодов обеспечивает альтернативу методу расчленения ансамблей, описанному выше. Однако, эти два метода тесно связаны. Этот новый метод привёл к открытию новых мощных свёрточных кодов, включающих большие сигнальные созвездия. Многие из них приведены в статье Кальдербанка и Слоэна (1987).

8.4. БИБЛИОГРАФИЧЕСКИЕ ЗАМЕЧАНИЯ И ССЫЛКИ

Пионерские работы по кодированию и кодированным сигналам для цифровой связи были сделаны Шенноном (1948a, b), Хеммингом (1950) и Голеем (1949). За этими работами скоро последовали статьи по качеству кодирования Гильберта (1952), по новым кодам Маллера (1954) и Рида (1954) и по технике кодирования для каналов с шумами Элиаса (1954...1955) и Слепяна (1956).



(*a*) Кодер



(*b*) 32-точечный КАМ сигнал (крест)

Рис. 8.3.13. Нелинейный свёрточный кодер с 8 состояниями для сигнального ансамбля 32-КАМ, который проявляет инвариантность к повороту фазы на 90°

В течение периода 1960...1970 появился целый ряд существенных вкладов в развитие теории кодирования и алгоритмов декодирования. В частности, мы цитировали статьи Рида и Соломона (1960) и коды Рида-Соломона, статьи Хоквингема (1959) и Боуза и Роя-Чоудхурн (1960 a, b) по БЧХ-кодам и диссертация на степень доктора философии Форни (1966 a) по каскадным кодам. За этими работами последовали статьи Гоппа (1970, 1971) по конструированию нового класса линейных циклических кодов, теперь называемые кодами Гоппа (смотри также Берлекэмп, 1973) и статьи Джастисена (1972) по конструктивной технике асимптотически хороших кодов. В течение этого периода работы по асимптотике декодирования были первоначально сфокусированы на БЧХ коды. Первый алгоритм декодирования для двоичных БЧХ кодов был разработан Питерсоном (1960). Большое число утончённых разработок и обобщений Чайна (1964), Форни (1965), Мэсси (1965) и Берлекемма (1968) вели к разработке эффективных в вычислительном отношении

алгоритмов декодирования БЧХ кодов, которые детально описаны Лином и Костелло (1983).

Параллельно с этими разработками по блоковым кодам шли разработки свёрточных кодов, которые были открыты Эллиасом (1955). Важнейшая проблема свёрточных кодов – их декодирование. Возенкрафт и Райфен (1961) описали алгоритм последовательного декодирования для свёрточных кодов. Этот алгоритм был позже модифицирован и уточнен Фано (1963) и теперь он называется *алгоритмом Фано*. Впоследствии были изобретены стек-алгоритмы Зигангировым (1966) и Елинеком (1969), а алгоритм Витерби был открыт Витерби (1967). Оптимальность и умеренная сложность при малых кодовых ограничениях способствовали тому, что алгоритм Витерби стал наиболее популярен для декодирования свёрточных кодов с кодовым ограничением $K \leq 10$.

Одним из наиболее важных вкладов в кодирование в течение 70-х годов были работы Унгербоека и Чайка (1976) по кодированию в ограниченных по полосе каналах. В этих статьях было показано, что можно достичь достаточный выигрыш кодирования путём введения избыточности в сигнале при ограниченном по полосе канале и были предложены решётчатые коды, достигающие выигрыш кодирования порядка 3-4 дБ. Эта работа вызвала большой интерес среди исследователей и привела к большому числу публикаций за последние 10 лет. Много ссылок можно найти в статьях Унгербоека (1982, 1987) и Форни и других (1984). Дополнительные статьи по кодированной модуляции для ограниченных по полосе каналов можно найти в «Special Issue on Voiceband Telephone Data Transmission, IEEE Journal on Selected Areas in Communication» (September 1984). Всесторонняя трактовка решётчато-кодированной модуляции дана в книге Биглиери и др. (1991).

Дополнительно к ссылкам, данным выше по кодированию, декодированию и синтезу кодированных сигналов, мы хотим вспомнить коллекцию статей, опубликованных в IEEE Press и озаглавленных «Key Papers in the Development of Coding Theory», изданных Берлекэмпом (1974). Эта книга содержит важные статьи, которые были опубликованы в первые 25 лет развития теории кодирования. Мы хотим также процитировать Special Issue on Error – Correcting Codes, IEEE Transactions on Communication (октябрь 1971).

ЗАДАЧИ

8.1. Дана порождающая матрица линейного блокового двоичного кода.

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

a) выразите $\mathbf{G}$ в систематической форме $[\mathbf{I}|\mathbf{P}]$.

b) определите проверочную матрицу $\mathbf{H}$ для кода.

c) сконструируйте таблицу синдромов для кода.

d) определите минимальное расстояние кода.

e) покажите, что кодовое слово, соответствующее информационной последовательности 101, ортогонально к $\mathbf{H}$.

8.2. Напишите кодовые слова, генерируемые матрицами, данными в (8.1.35) и (8.1.37) и затем покажите, что эти матрицы генерирует одинаковый ансамбль кодовых слов.

8.3. Распределение весов кодов Хемминга известно. Выраженное в виде полинома степеней x распределение весов двоичного кода Хемминга с длиной блока n имеет вид

$$A(x) = \sum_{i=0}^{n} A_i x^i = \frac{1}{n+1}\left[(1+x)^n + n(1+x)^{(n-1)/2}(1-x)^{(n+1)/2}\right],$$

где A_i – число кодовых слов с весом i. Используете эту формулу для определения распределения весов кода Хемминга (7.4) и сравните ваш результат со списком кодовых слов, данном в табл. 8.1.2.

8.4. Полином

$$g(p) = p^4 + p + 1$$

является порождающим для двоичного кода Хемминга (15, 11).

a) определите порождающею матрицу $\mathbf{G}$ для этого кода в систематической форме.

b) определите порождающею матрицу для дуального кода.

8.5. Для циклического кода Хемминга (7, 4) с порождающим полиномом

$$g(p) = p^3 + p^2 + 1$$

сконструируйте расширенный код Хемминга (8, 4) и приведите список кодовых слов. Каково d_{min} для расширенного кода(?).

8.6. Линейный блоковый код (8, 4) сконструирован укорочением кода Хемминга (15, 11), генерированного порождающим полиномом

$$g(p) = p^4 + p + 1.$$

a) Сконструируйте кодовые слова для кода (8, 4) и дайте их список.

b) Каково минимальное расстояние кода (8, 4)?

8.7. Факторизация полинома $p^{15} + 1$ даёт

$$p^{15} + 1 = (p^4 + p^3 + 1)(p^4 + p^3 + p^2 + p + 1)(p^4 + p + 1)(p^2 + p + 1)(p + 1).$$

a) сконструируйте систематический код (15, 5), использующий порождающий полином

$$g(p) = (p^4 + p^3 + p^2 + p + 1)(p^4 + p + 1)(p^2 + p + 1).$$

b) Каково минимальное расстояние кода?

c) Сколько случайных ошибок в кодовом слове можно исправить?

d) Сколько ошибок можно обнаружить кодом?

e) Напишите кодовые слова кода (15, 2), сконструированного порождающим полиномом

$$g(p) = (p^{15} + 1)/(p^2 + p + 1),$$

и определите его минимальное расстояние.

8.8. Сконструируйте проверочные матрицы $\mathbf{H}_1$ и $\mathbf{H}_2$ соответствующие порождающим матрицам $\mathbf{G}_1$ и $\mathbf{G}_2$, данные, соответственно, (8.1.34) и (8.1.35).

8.9. Сконструируйте расширенный код (8,4) из кода Хемминга (7.4) путём видоизменения порождающей и проверочной матриц.

8.10. Систематический код (6, 3) имеет проверочную матрицу

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

Сконструируйте стандартное расположение и определите корректируемые образцы ошибок и соответствующие им синдромы.

8.11. Сконструируйте стандартное расположение для кода (7, 3) с порождающей матрицей

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{bmatrix}$$

н определите корректируемые образцы ошибок и соответствующие им синдромы.

8.12. Определите корректируемые образцы ошибок (с наименьшим весом) и их синдромы для циклического кода Хемминга (7, 4).

8.13. Докажите, что если сумма двух образцов ошибки $\mathbf{e}_1$ и $\mathbf{e}_2$ является действительным кодовым словом $\mathbf{C}_j$, то каждый образец ошибки имеет одинаковый синдром.

8.14. Пусть

$$g(p) = p^8 + p^6 + p^4 + p^2 + 1$$

является полиномом относительно двоичного поля.

a) найдите циклический код с наименьшей скоростью, для которого порождающим полиномом является $g(p)$. Какова скорость этого кода?

b) найдите минимальное расстояние кода, найденного в (a).

c) каков выигрыш кодирования для кода найденного в (a).

8.15. Рассматривается полином $g(p) = p + 1$ относительно двоичного поля.

a) покажите, что этот полином может генерировать циклический код при любом выборе n. Найдите соответствующее значение k.

b) найдите систематическую форму **G** и **H** для кода, генерируемого посредством $g(p)$.

c) можете ли сказать, какого типа код создается этим порождающим полиномом?

8.16. Синтезируйте циклический код (6, 2) выбором возможно короткого порождающего полинома.

a) определите порождающую матрицу **G** (в систематической форме) для этого кода и найдите все возможные кодовые слова.

b) сколько ошибок может корректировать этот код?

8.17. Докажите, что суммирование любых двух комбинаций в одной строке стандартного расположения даёт возможное кодовое слово.

8.18. Исходя из БЧХ кода (15, 7), сконструируйте укороченный код (12, 4). Дайте порождающую матрицу для укороченного кода.

8.19. В разделе 8.1.2 было указано, что когда код Адамара (n,k) отображается в сигналы посредством двоичной ФМ соответствующие $M = 2^k$ сигналов ортогональны. Определите показатель расширения полосы для M ортогональных сигналов и сравните с требованиями по полосе для ортогональных сигналов ФМ при когерентном детектировании.

8.20. Покажите, что сигналы, генерируемые кодом регистра сдвига максимальной длины при отображении каждого кодового символа в кодовом слове двоичным ФМ сигналом одинаково коррелированны с коэффициентом корреляции $\rho_r = -1/(M-1)$, то есть M сигналов образует симплексный ансамбль.

8.21. Вычислите вероятность ошибки декодирования, получаемой в канале с АБГШ при использовании кода Хемминга (7, 4) при декодировании жёстких и мягких решений. Используйте (8.1.50), (8.1.52), (8.1.82), (8.1.90) и (8.1.91).

8.22. Используйте результаты раздела 2.1.6 полученные для границы Чернова при декодировании жёстких решений (формулы (8.1.89) и (8.1.90)). Предположив, что передано кодовое слово из одних нулей, определите верхнюю границу для вероятности того, что выбрана декодером кодовое слово C_m, имеющий вес w_m. Это имеет место, если $\frac{1}{2}w_m$ или больше символов принято с ошибкой. Для определения границы Чернова определите последовательность из w_m случайных величин, как

$$X_i = \begin{cases} 1 \text{ с вероятностью } p, \\ -1 \text{ с вероятностью } 1-p, \end{cases}$$

где $i = 1,2,...,w_m$, а p – вероятность ошибки в канале. Для ДСК величины $\{X_i\}$ статистически независимы.

8.23. Свёрточный код описывается генераторами

$$\mathbf{g}_1 = [100], \quad \mathbf{g}_2 = [101], \quad \mathbf{g}_3 = [111]$$

a) получите кодер, соответствующий этому коду,

b) получите диаграмму состояний для этого кода,

c) получите решётчатую диаграмму для этого кода,

d) найдите передаточную функцию и свободное расстояние этого кода,

e) проверьте, является ли этот код катастрофическим.

8.24. Свёрточный код из задачи 8.23 используется для передачи по каналу с АБГШ при декодировании жёстких решений. Выходом детектора (демодулятора) является (101001011110111...). Используя алгоритм Витерби, найдите переданную последовательность.

8.25. Повторите задачу 8.23 для кода с генераторами

$\mathbf{g}_1 = [110], \mathbf{g}_2 = [101], \mathbf{g}_3 = [111]$

8.26. Блок-схема двоичного свёрточного кодера показана на рисунке Р8.26.

a) Получите диаграммы состояний кода.

b) Найдите передаточную функцию кода T(D).

c) Каково минимальное свободное расстояние кода $d_{св}$?

d) Предположите, что сообщение кодируется этим кодом и передается через двоичный симметричный канал с вероятностью ошибки $p = 10^{-5}$. Используя алгоритм Витерби найдите переданную последовательность символов, если принята последовательность

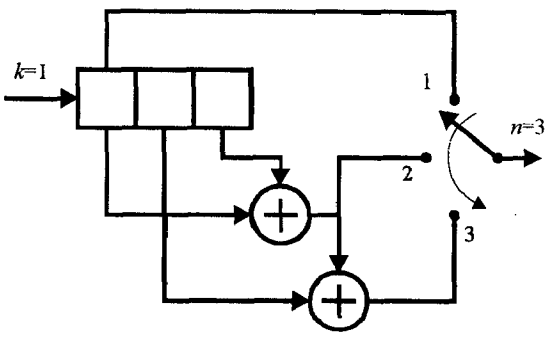


Рис. Р.8.26

$\mathbf{r} = (110, 110, 110, 111, 010, 101, 101)$.

e) Найдите верхнюю границу для вероятности ошибочного декодирования, если используется указанный выше двоичный симметричный канал. Сделайте приемлимую аппроксимацию.

8.27. Блок схема свёрточного кода (3, 1) показана на рисунке Р8.27.

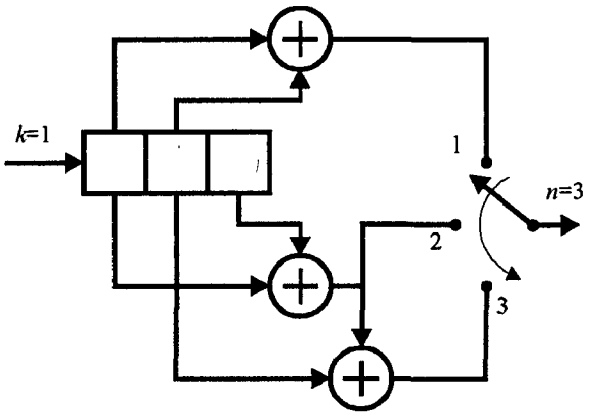


Рис. Р.8.27

a) Получите диаграммы состояний кода.

b) Найдите передаточную функцию кода $T(D)$.

c) Найдите минимальное свободное расстояни кода $(d_{св})$ и покажите соответствующий путь (и расстоянии $d_{св}$ относительно кодового слова из одни нулей) на решётке.

d) Предположите, что четыре двоичны информационных символа (x_1, x_2, x_3, x_4), з которыми следуют два нуля закодированы. Эт информация передается по двоичному симметричном каналу с переходной вероятностью ошибки 0, 1 Принимаемая последовательность такова (111, 111, 111 111, 111, 111). Используйте алгоритм декодировани Витерби, чтобы найти переданную последовательност данных.

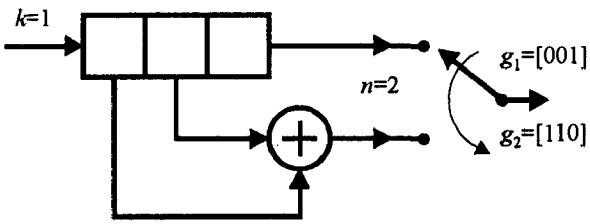


Рис. Р.8.28

8.28. Для свёрточного кода, генерируемого кодером показанным на рисунке Р8.28:

а) Найдите передаточную функцию кода в форм $T(N, D)$

b) Найдите свободное расстояние кода $d_{св}$.

c) Найдите верхнюю границу для средне вероятностной ошибки на бит, используя границу дл вероятности ошибочного декодирования жёстки решений, если код используется в канале пр условии, что переходная вероятность ошибки канале $p = 10^{-6}$.

8.29. На рис. Р8.29 изображён свёрточный код со скоростью 1/2 и кодовым ограничением K=2

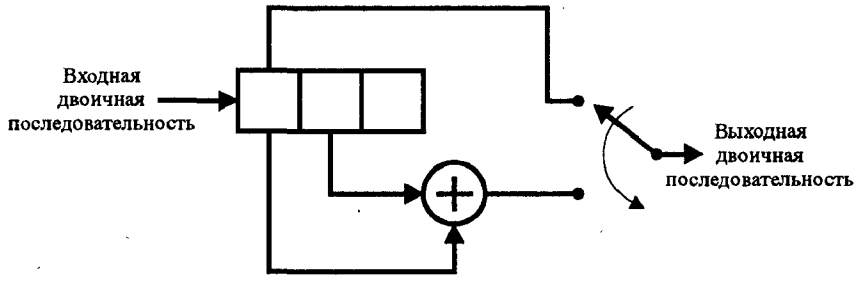


Рис. Р.8.28

a) Нарисуйте древовидную диаграмму кода, решётчатую диаграмму и диаграмму состояний.

b) Найдите передаточною функцию $T(N, D, J)$ и по ней определите минимальное свободно расстояние.

8.30. На рисунке Р.8.30 показан свёрточный код со скоростью 1/2, K=3:

a) Получите дерево кода, диаграмму решётки и диаграммы состояний.

b) Определите передаточную функцию $T(N, D, J)$ и с ее помощью найдите минимальное свободно расстояние.

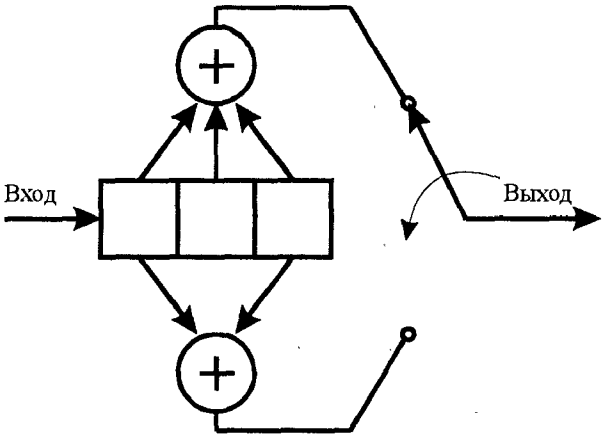


Рис. Р.8.30

8.31. Нарисуйте схемы свёрточных кодеров по следующим параметрам:

a) Скорость кода 1/2, K=5, максимальное свободное расстояние (табл. 8.2.1).

b) Скорость кода 1/3, K=5, максимальное свободное расстояние (табл. 8.2.2).

c) Скорость кода 2/3, K=2, максимальное свободное расстояние (табл. 8.2.8).

8.32. Получите диаграмму состояний для свёрточного кода со скоростью 2/3, K=5, указанных в задаче 8.31(c) и покажите для каждого перехода выходную последовательность и расстояние выходной последовательности от последовательности состоящей из одних нулей.

8.33. Рассмотрите свёрточный код со скоростью 1/2, K=3, показанный на рисунке Р8.30. Предположим, что код используется в двоичном симметричном канале и принимаемая последовательность для первых восьми ветвей такова 0001100000001001. Постройте диаграмму решётки. Проследите за решением по диаграмме решётки и пометьте у выживших путей расстояние Хемминга на каждом узле. Если возникнет неувязка с метриками, требуемыми для решения, выберете верхний путь (произвольный выбор).

8.34. Используйте передаточную функцию, полученную в задаче 8.30 для свёрточного кода со скоростью R_c=1/2, K=3, для расчета вероятности ошибки на бит в канале с АБГШ при декодировании жёстких решений (*a*) и мягких решений (*b*). Сравните качество, построив результаты расчетов на одном графике.

8.35. Используйте генераторы, данные (8.2.36) для получения кодера для функции 3 – дуального свёрточного кода со скоростью 1/2. Определите диаграмму состояний и получите передаточную функцию $T(N,D,J)$.

8.36. Получите диаграмму состояний для свёрточного кода, генерируемого кодером, показанным на рисунке Р8.36 и затем определите является ли код катастрофическим или нет. Также дайте пример свёрточного кода со скоростью 1/2, K=4, который проявляет катастрофические распространения ошибок.

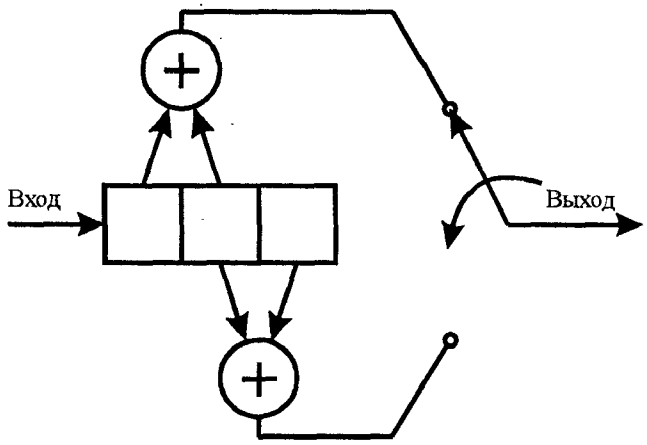


Рис. Р.8.36

8.37. Решётчато-кодированный сигнал формируется так, как показано на рисунке Р8.37 путём кодирования одного символа посредством свёрточного кода со скоростью 1/2 а три дополнительных

информационных символа остаются некодированными. Образуйте расчленение ансамбля созвездия 32КАМ (крест) и укажите подобразы в расчленении. Насколько увеличится расстояние между соседними сигнальными точками в результате расчленения?

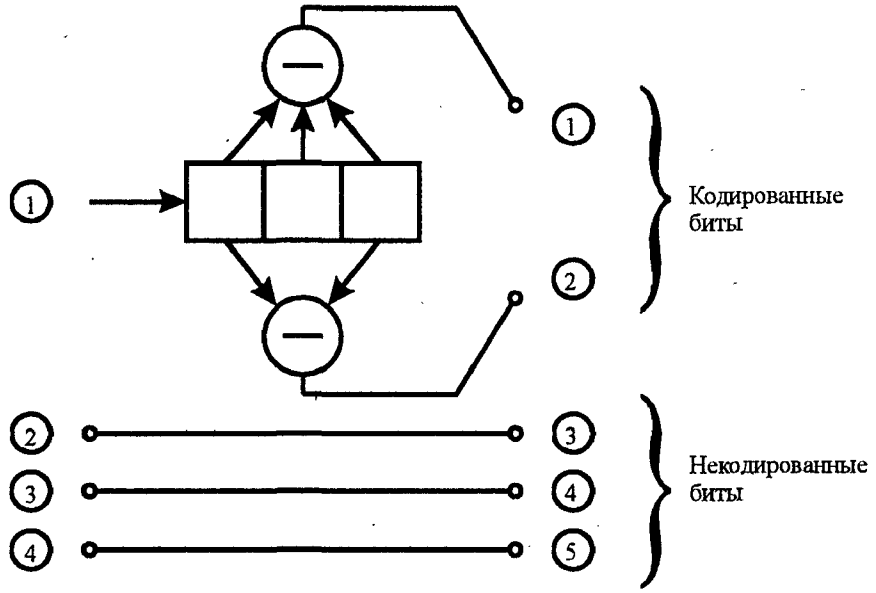


Рис. Р.8.37

8.38. Пусть X_1 и X_2 – два кодовых слова длины n с расстоянием d и предположите, что эти два кодовых слова передаются по двоичному симметричному каналу с переходной вероятностью ошибки p. Пусть $P(d)$ обозначает вероятность ошибки при передаче этих кодовых слов.

a) Покажите, что

$$P(d) \le \sum_{i=1}^{2^n} \sqrt{P(y_i|x_1)P(y_i|x_2)},$$

где суммирование ведется по всем двоичным последовательностям y_i

b из вышесказанного заключите, что

$$P(d) \le \left[4p(1-p)\right]^{d/2}.$$

9

СИНТЕЗ СИГНАЛОВ ДЛЯ ОГРАНИЧЕННЫХ ПО ПОЛОСЕ КАНАЛОВ

В предыдущих главах мы рассмотрели передачу цифровой информации по каналу с аддитивным гауссовским шумом. Причём при синтезе сигналов и систем связи никакие ограничения на полосу частот не были наложены.

В этой главе мы рассмотрим проблему синтеза сигналов, когда канал ограничен по спектру некоторой определённой полосой W Гц. При этом условии канал можно моделировать линейным фильтром, имеющем эквивалентную низкочастотную характеристику $C(f)$, которая равна нулю для $|f| > W$.

Первая тема, которая обсуждается при синтезе сигнального импульса $g(t)$ в линейно-модулированном сигнале, представленном в виде

$$\upsilon(t) = \sum_{n=0}^{\infty} I_n g(t - nT),$$

сводится к эффективному использованию общей возможной полосы канала W. Мы можем видеть, что если канал идеальный для $|f| \le W$, то сигнальный импульс можно синтезировать так, чтобы обеспечить скорость передачи символов, сравнимую или превосходящую полосу канала W. С другой стороны, если канал не идеален, передача сигналов со скоростью передачи символов, равной или большей W, приводит к межсимвольной интерференции (МСИ) среди ряда соседних символов.

Вторая тема, которая обсуждается в этой главе, – это использование кодирования для формирования спектра передаваемого сигнала и преодоление, таким образом, проблемы МСИ. Мы начнем наше обсуждение с общей характеристики частотно-ограниченных линейных фильтровых каналов.

9.1. ХАРАКТЕРИСТИКА ОГРАНИЧЕННЫХ ПО ПОЛОСЕ КАНАЛОВ

Из различных каналов, имеющихся в распоряжении для цифровой связи, телефонные каналы на сегодня используются наиболее широко. Такие каналы характеризуются как *ограниченные по полосе линейные фильтры*. Это, конечно, подходящая характеристика, когда используется множественный доступ с частотным разделением (FDMA, МДЧР), как средство установления доступа абонентов к каналам телефонной сети. Современное добавление к телефонным сетям – использование импульсно-кодовой модуляции (ИКМ) для цифровизации и кодирования аналоговых сигналов и множественный доступ с временным разделением (TDMA, МДВР) для образования многих каналов. Тем не менее, частотная фильтрация аналоговых сигналов до сих пор используется до стробирования и кодирования. Следовательно, хотя сейчас телефонные сети используют совмещение FDMA и TDMA для передачи, линейная фильтровая модель для телефонных каналов все еще приёмлема.

Для наших целей ограниченный по полосе канал, такой как телефонный канал, может характеризоваться как линейный фильтр, имеющий эквивалентную низкочастотную частотную характеристику $C(f)$. Его эквивалентная низкочастотная импульсная характеристика обозначается $c(t)$.Тогда, если сигнал вида

$$s(t) = \mathrm{Re}\left[\upsilon(t)e^{j2\pi f_c t}\right] \tag{9.1.1}$$

передается по полосовому телефонному каналу, принимаемый эквивалентный низкочастотный сигнал равен

$$r_l(t) = \int_{\infty}^{-\infty} \upsilon(\tau)c(t-\tau)d\tau + z(t), \tag{9.1.2}$$

где интеграл представляет свертку $c(t)$ и $\upsilon(t)$, а $z(t)$ означает эквивалентный низкочастотный аддитивный шум. Альтернативно, сигнальную составляющую можно представить в частотной области как $V(f)C(f)$, где $V(f)$ – преобразование Фурье от $\upsilon(t)$.

Если канал ограничен по полосе частотой W Гц, тогда $C(f)=0$ для $|f|>W$. Как следствие, любая частотная компонента в $V(f)$, лежащая выше $|f|=W$, не пройдет по каналу. Из этих соображений мы ограничиваем полосу передаваемых сигналов частотой W Гц.

Внутри полосы канала мы можем выразить его частотную характеристику $C(f)$ так:

$$C(f) = |C(f)|e^{j\theta(f)}, \tag{9.1.3}$$

где $|C(f)|$ – амплитудная характеристика (АЧХ), а $\theta(f)$ – фазовая характеристика (ФЧХ). Далее, характеристика, определяющая задержку по огибающей, равна

$$\tau(f) = -\frac{1}{2\pi}\frac{d\theta(f)}{df}. \tag{9.1.4}$$

Говорят, что канал *без искажений* или *идеальный*, если его амплитудная характеристика $|C(f)|$ постоянна для всех $|f|\le W$, а фазовая характеристика $\theta(f)$ является линейной функцией частоты, т.е. $\tau(f)$ является константой для всех $|f|\le W$. С другой стороны, если $|C(f)|$ не постоянна для всех $|f|\le W$, мы говорим, что канал *искажает передаваемый сигнал* $V(f)$ *по амплитуде*, а если $\tau(f)$ не постоянна для всех $|f|\le W$, мы говорим, что канал *искажает сигнал* $V(f)$ *по задержке*.

Как результат искажений сигнала по амплитуде и задержке, вызванных неидеальностью частотной характеристики канала $C(f)$, соседние импульсы, передаваемые по каналу со скоростями, сравнимыми с W, расплываются до такой степени, что они больше не различимы на приёмной стороне. Наоборот, они перекрываются и, следовательно, имеем межсимвольную интерференцию. Для примера влияния искажений по задержке на передаваемый импульс на рис. 9.1.1(a) показан ограниченный по полосе импульс, имеющий периодические нули в точках времени $\pm T$, $\pm 2T$ и т.д.

Если информация заложена в амплитуде импульса, как, например, при АМ, тогда можно передавать последовательность импульсов, каждый из которых имеет пик в точках периодических нулей других импульсов. Однако передача импульса по каналу, моделируемому линейной зависимостью задержки огибающей $\tau(f)$ от частоты [квадратичной фазовой характеристикой $\theta(f)$], приводит к принимаемому импульсу, показанному на рис. 9.1.1(b), имеющему расположение нулей, которое теперь имеет непериодический характер. Следовательно, соседние импульсы «растекаются» один в другой, и пики отдельных импульсов больше не различимы. Таким образом искажения по

задержке канала приводят к межсимвольной интерференции. В главе 10 мы покажем, что существует возможность компенсировать неидеальную частотную характеристику канала при помощи фильтра или выравнивателя (эквалайзера) в демодуляторе. Рис. 9.1.1(c) иллюстрирует выход линейного выравнивателя (эквалайзера), который компенсирует линейные искажения в канале.

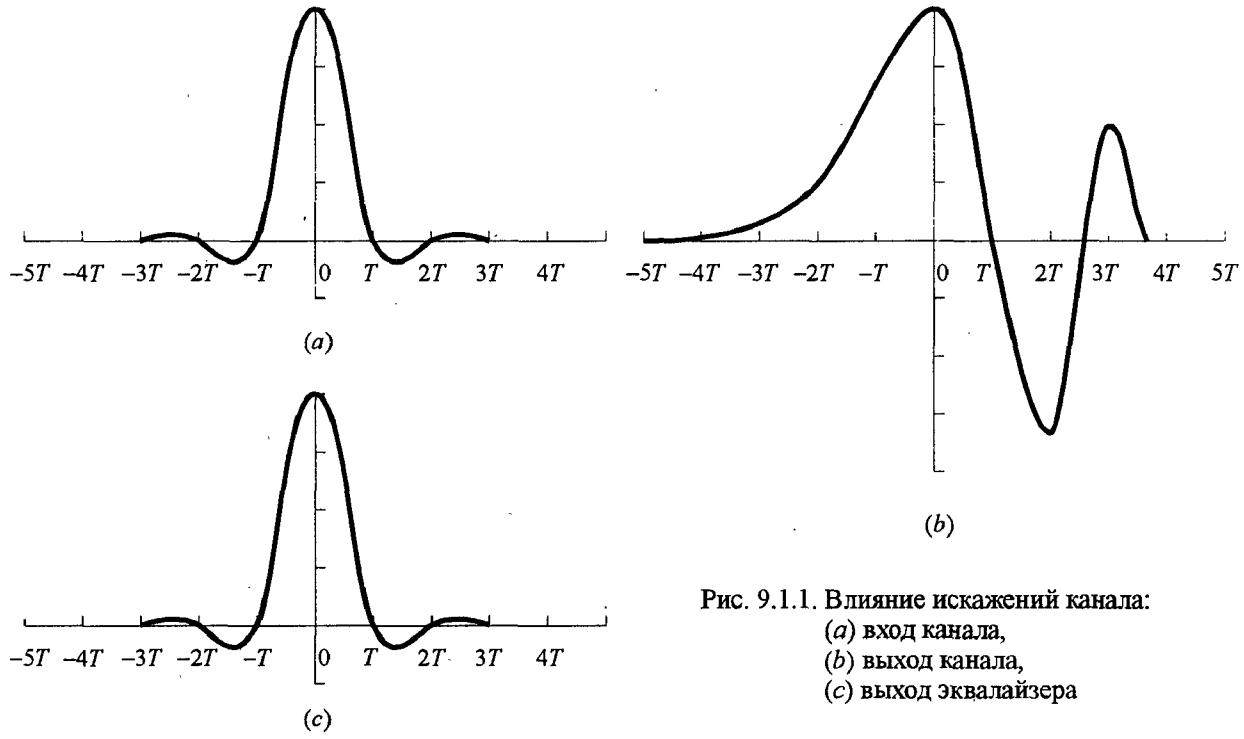


Рис. 9.1.1. Влияние искажений канала:
(*a*) вход канала,
(*b*) выход канала,
(*c*) выход эквалайзера

Существование межсимвольной интерференции в телефонном канале можно ощутить наблюдением частотной характеристики канала. Рис. 9.1.2 иллюстрирует измеренные средние значения амплитуды и задержки, как функции частоты для телефонного канала средней протяжённости в коммутируемых телефонных сетях, данные Даффи и Трэтчером (1971).

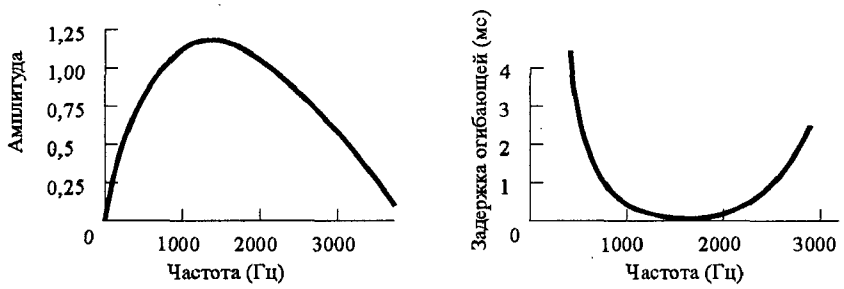


Рис. 9.1.2. Средние характеристики обычного телефонного канала:
(*a*) амплитудно-частотная характеристика (АЧХ);
(*b*) характеристика группового времени передачи (ГВП)

Обычно используемая полоса частот канала простирается от 300 Гц до примерно 3000 Гц. Соответствующая импульсная характеристика такого канала показана на рис. 9.1.3. Ее протяженность около 10 мс. Скорость передачи символов по такому каналу может быть порядка 2500 импульсов или символов в секунду. Следовательно, межсимвольная интерференция в канале может простираться на 20–30 символов.

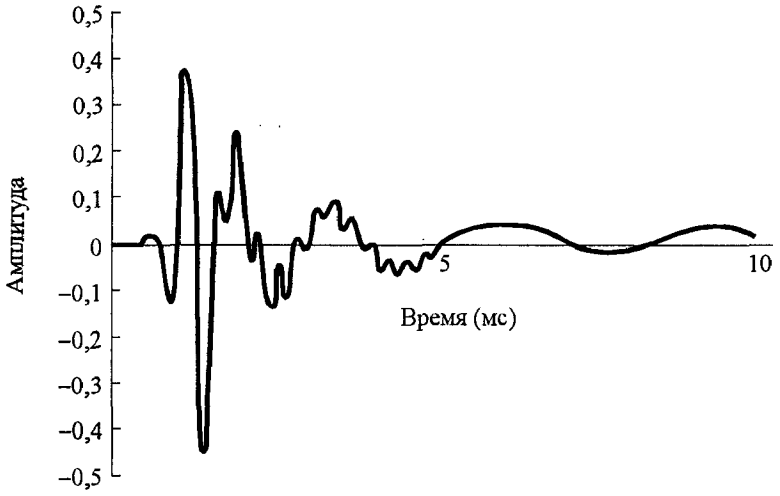


Рис. 9.1.3. Импульсная характеристика канала, частотная харакеристика которого дана на рис. 9.1.2

В дополнение к линейным искажениям (нарушениям) сигналы, передаваемые через телефонные каналы, подвергаются иным искажениям, в частности, нелинейным искажениям, смещению частоты, фазовому джиттеру, воздействию импульсных и тепловых шумов.

Нелинейные искажения в телефонных каналах происходят от нелинейности усилителей и компандеров, используемых в телефонных системах. Этот вид искажений обычно невелик, и его очень трудно корректировать.

Небольшое *смещение частоты*, обычно меньшее, чем 5 Гц, обусловлено использованием аппаратуры несущей в телефонном канале. Такое смещение нельзя терпеть в высокоскоростных цифровых системах передачи, которые используют синхронную фазово-когерентную демодуляцию. Смещение, обычно, компенсируется при помощи схемы восстановления несущей в демодуляторе.

Фазовый джиттер, по существу, представляет собой частотную модуляцию передаваемого сигнала с малым индексом низкочастотными гармониками частоты питания (50...60 Гц). Фазовый джиттер создает серьезные проблемы при цифровой передаче с большими скоростями. Однако его можно отслеживать и в определенной степени компенсировать до демодулятора.

Импульсный шум является аддитивным искажением. Он возникает прежде всего от коммутационного оборудования в телефонных системах.

Тепловой гауссовский шум также присутствует на уровнях 20...30 дБ ниже сигнала.

Степень нашей обеспокоенности этими нарушениями в канале зависит от скорости передачи по каналу и техники модуляции. Для скоростей передачи ниже 1800 бит/с ($R/W < 1$) можно выбрать технику модуляции, например, ФМ, которая относительно нечувствительна к величинам искажений, которые встречаются в типовых телефонных каналах от всех источников, указанных выше. Для скоростей между 1800 и 2400 бит/с ($R/W = 1$) обычно, используется более эффективная по полосе частот техника модуляции, такая как четырехфазовая ФМ. На этих скоростях для компенсации искажений средних амплитуд и задержек в канале часто используются некоторые виды компромиссного выравнивания. Дополнительно синтезируется метод восстановления несущей, позволяющий компенсировать смещение частоты. Другие канальные нарушения несущественно влияют на этих скоростях на характеристику частости ошибок. На скоростях передачи выше 2400 бит/с ($R/W > 1$) используется эффективная по полосе

частот техника кодированной модуляции, такая как решётчато-кодированная КАМ, АМ и ФМ. Для этих скоростей особое внимание надо уделять линейным искажениям, смещению частоты и фазовому джиттеру. Линейные искажения обычно компенсируются посредством адаптивных выравнивателей (эквалайзеров). Фазовый джиттер преодолевается комбинированием синтеза сигнала и некоторых видов фазовых компенсаторов в демодуляторе. На скоростях выше 9600 бит/с специальное внимание надо уделить не только линейным искажениям, частотному смещению и фазовому джиттеру, но также иным нарушениям в канале, упомянутым выше.

К сожалению, модель канала, которая учитывает все искажения, указанные выше, сложна для анализа. В этой и следующих двух главах в качестве модели канала используется линейный фильтр, который вносит амплитудно-частотные искажения и искажения по задержке сигнала, а также добавляет гауссовский шум.

Кроме телефонных каналов, имеются другие физические каналы, в которых проявляются некоторые формы временного рассеяния (дисперсии) сигналов и, таким образом, вводится межсимвольная интерференция. Радиоканалы, такие как коротковолновые каналы ионосферного распространения (ВЧ) и тропосферного рассеяния, являются двумя примерами каналов с временной дисперсией. В этих каналах временное рассеяние и, следовательно, межсимвольная интерференция, являются результатом многопутевого распространения волн с различными задержками в отдельных путях. Число путей и относительные временные задержки по путям меняются во времени, и из этих соображений эти радиоканалы обычно называют *меняющимися во времени многопутевыми каналами*. Условия изменяющейся во времени многопутевости дают начало большому разнообразию частотных характеристик каналов. Следовательно, частотные характеристики, использованные в телефонных каналах, не подходят для меняющихся во времени многопутевых каналов. Вместо них радиоканалы характеризуются статистически, как более детально излагается в главе 14, через функцию рассеяния, которая является двумерным представлением средней мощности принимаемого сигнала как функция от относительной задержки во времени и доплеровской частоты.

С целью иллюстрации на рис. 9.1.4 показана функция рассеяния для канала тропосферного рассеяния средней протяжённости (150 миль).

Суммарная длительность во времени (из-за многопутевого рассеяния) характеристики канала в среднем примерно 0,7 мкс, а рассеяние между «точками половинной мощности» по доплеровской частоте немного меньше, чем 1 Гц для наиболее мощного пути и немного больше для других путей. Типично, что при скорости передачи по такому каналу 10^7 символов/с многопутевое рассеяние 0,7 мкс приводит к межсимвольной интерференции, которая простирается на семь символов.

В этой главе мы имеем дело исключительно с линейной, неизменной во времени, фильтровой моделью для ограниченного по полосе канала. Адаптивная техника выравнивания, представленная в главах 10 и 11 для борьбы с межсимвольной интерференцией, также применима для меняющихся во времени многопутевых каналов при условии, что изменения во времени канала относительно медленные по сравнению с общей полосой канала или, что эквивалентно, со скоростью передачи символов по каналу.

9.2. СИНТЕЗ СИГНАЛОВ ДЛЯ ОГРАНИЧЕННЫХ ПО ПОЛОСЕ КАНАЛОВ

В главе 4 было показано, что передаваемый эквивалентный низкочастотный сигнал для несколько различных видов техники цифровой модуляции имеет общую формулу

$$\upsilon(t)=\sum_{n=0}^{\infty} I_n g(t-nT), \tag{9.2.1}$$

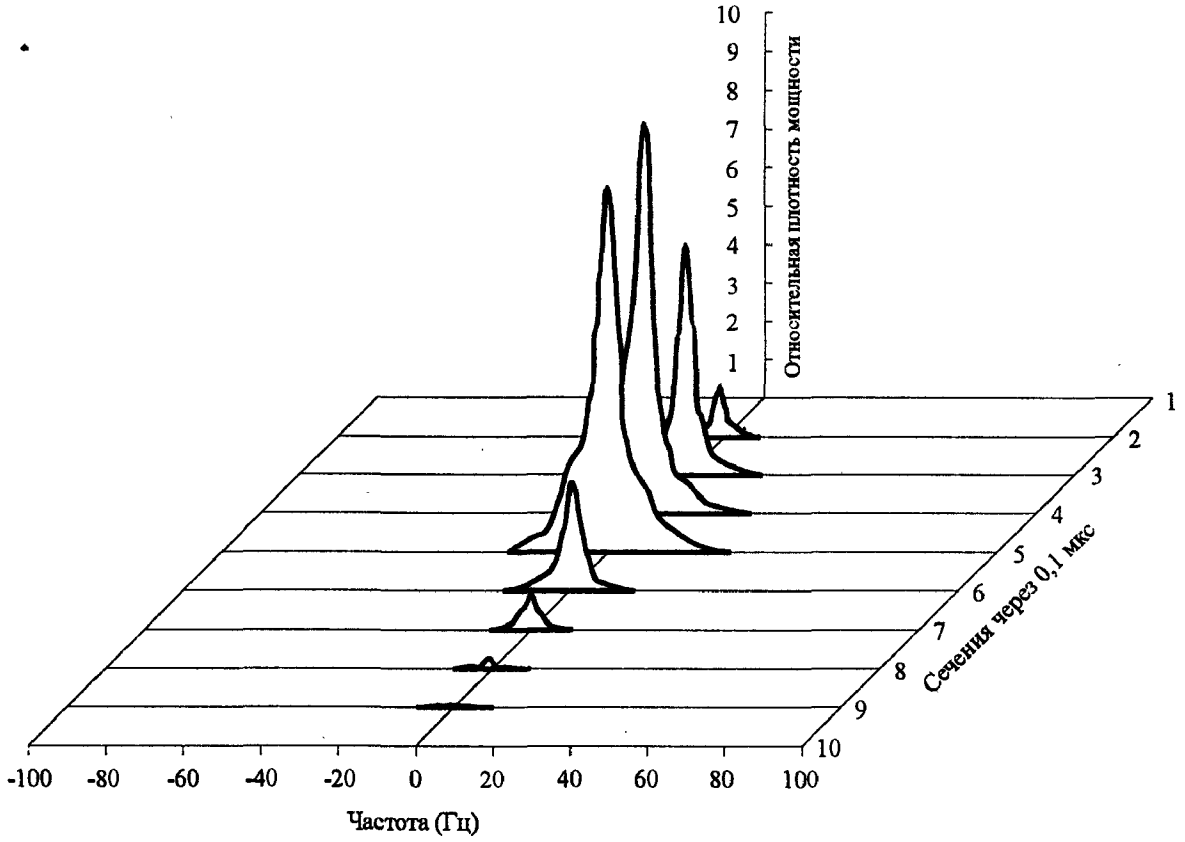


Рис. 9.1.4. Функция рассеяния канала с тропосферным рассеянием

где $\{I_n\}$ представляют последовательность символов, несущую информацию, а $g(t)$ – импульс, который для целей дальнейшего обсуждения предполагается с ограниченным по полосе частотной спектром $G(f)$, т.е. $G(f)=0$ для $|f|>W$. Этот сигнал передается по каналу, имеющему частотную характеристику $C(f)$, также ограниченную по полосе $|f|\le W$. Следовательно, принимаемый сигнал можно представить так:

$$r_l(t)=\sum_{n=0}^{\infty} I_n h(t-nT)+z(t), \tag{9.2.2}$$

где

$$h(t)=\int_{-\infty}^{\infty} g(\tau)c(t-\tau)d\tau, \tag{9.2.3}$$

а $z(t)$ представляют аддитивный гауссовский шум.

Предположим, что принимаемый сигнал сначала пропускается через фильтр, а затем стробируется со скоростью $1/T$ отсчетов в секунду. Мы показали в соответствующем разделе, что оптимальный фильтр при детектировании сигнала – это фильтр, согласованный с принимаемым импульсом. Это значит, что частотная характеристика фильтра на приёме равна $H^*(f)$ ($H(f)$ – преобразование Фурье от $h(t)$). Обозначим выход фильтра приёмника так

$$y(t)=\sum_{n=0}^{\infty} I_n x(t-nT)+v(t), \tag{9.2.4}$$

где $x(t)$ – отклик фильтра на приёме на входной импульс $h(t)$, а $v(t)$ – отклик фильтра на шум $z(t)$

Теперь, если $y(t)$ стробируется во времени в точках $t = kT + \tau_0$, $k = 0, 1, \ldots$, мы имеем

$$y(kT+\tau_0) \equiv y_k = \sum_{n=0}^{\infty} I_n x(kT - nT + \tau_0) + v(kT+\tau_0), \qquad (9.2.5)$$

или, что эквивалентно

$$y_k = \sum_{n=0}^{\infty} I_n x_{k-n} + v_k, \quad k = 0, 1, \ldots, \qquad (9.2.6)$$

где τ_0 – задержка при передаче по каналу. Величины отсчетов можно выразить так

$$y_k = x_0 \left(I_k + \frac{1}{x_0} \sum_{\substack{n=0 \\ n \neq k}}^{\infty} I_n x_{k-n} \right) + v_k, \quad k = 0, 1, \ldots. \qquad (9.2.7)$$

Мы считаем x_0 произвольным (известным) скалярным множителем, который, для удобства, примем равным единицы. Тогда

$$y_k = I_k + \sum_{\substack{n=0 \\ n \neq k}}^{\infty} I_n x_{k-n} + v_k. \qquad (9.2.8)$$

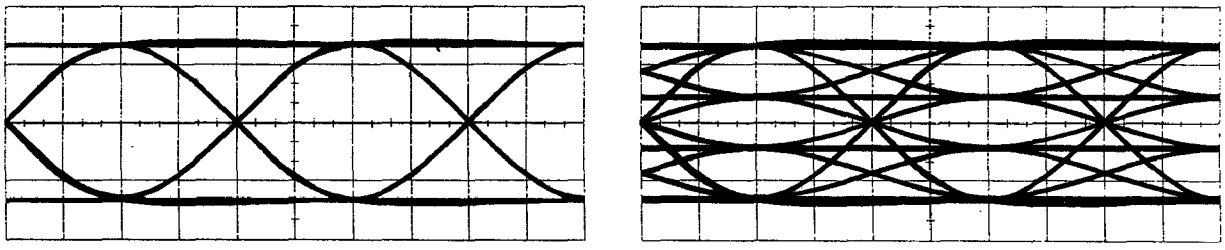
Слагаемое I_k представляет желательный информационный символ в k-й отсчетной точке, слагаемое

$$\sum_{\substack{n=0 \\ n \neq k}}^{\infty} I_n x_{k-n}$$

представляет межсимвольную интерференцию (МСИ), а v_k – аддитивная гауссовская шумовая величина в k-ой отсчётной точке.

Уровень МСИ и шум в цифровых системах связи можно наблюдать на осциллографе. Для сигналов АМ мы можем наблюдать принимаемый сигнал $y(t)$ на вертикальном входе (по вертикальной оси) при периоде горизонтальной развёртки, кратном T. Результирующая осциллограмма на дисплее называется *глазковой диаграммой* из-за её сходства с человеческим глазом.

Для примера, рис.9.2.1 иллюстрирует глазковую диаграмму для двоичной и четырехуровневой АМ



двоичная 4-позиционная

Рис. 9.2.1. Примеры глазковых диаграмм для двоичной и 4-позиционной АМ

Влияние МСИ проявляется в закрытии глазка, тем самым уменьшается допуск на величину аддитивного шума, вызывающую ошибку. Рис. 9.2.2 графически иллюстрирует

влияние МСИ на сокращение открытости двоичного глазка.

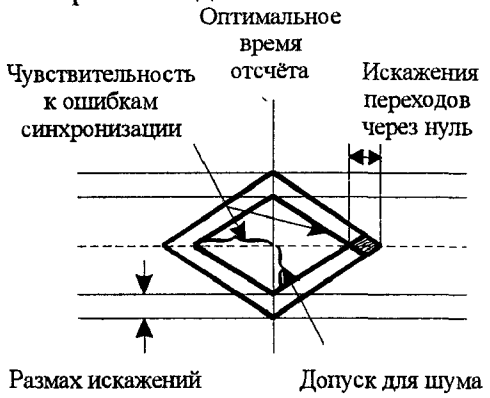


Рис. 9.2.2. Влияние МСИ на раскрытие глазковой диаграммы

Заметим, что МСИ искажает положение переходов через нуль и вызывает уменьшение открытости глазка. Тем самым она обуславливает большую чувствительность системы к ошибкам временной синхронизации.

Для ФМ и КАМ привычно рассматривать «глазковую диаграмму» как двухмерную диаграмму рассеяния, иллюстрирующая величины отсчетов $\{y_k\}$, которые представляют величины решений в отсчетных точках. Рис. 9.2.3 иллюстрирует такую глазковую диаграмму для сигнала 8 ФМ.

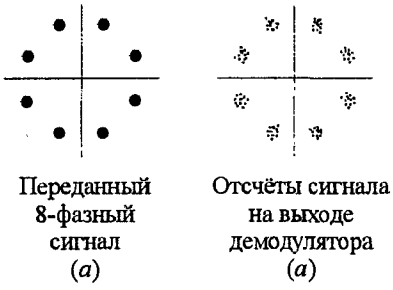


Рис. 9.2.3. Двухмерные цифровые глазковые диаграммы

В отсутствие МСИ и шума переданный сигнал в отсчётные моменты времени порождает в месте приёма восемь различимых точек, соответствующих восьми переданным значениям фаз сигнала. МСИ и шум приводят к отклонению принимаемых отсчетов $\{y_k\}$ от желаемых сигналов 8 ФМ. Чем больше МСИ и шум, тем больше рассеяние отсчетов принимаемых сигналов относительно точек передаваемых сигналов.

Ниже мы рассмотрим проблему синтеза сигналов при условии, что в отсчётных точках нет МСИ.

9.2.1. Синтез ограниченных по полосе сигналов при отсутствии межсимвольной интерференции – критерий Найквиста

В этом разделе и в 9.2.2 мы предположим, что ограниченные по полосе каналы имеют идеальную частотную характеристику, т.е. $C(f)=1$ для $|f|\le W$. Тогда импульс $x(t)$ имеет спектральную характеристику $X(f)=|G(f)|^2$, причём

$$x(t)=\int_{-W}^{W} X(f)\mathrm{e}^{j2\pi ft}df\,. \qquad (9.2.9)$$

Мы интересуемся спектральными свойствами импульса $x(t)$ и затем передаваемого импульса $g(t)$, когда нет межсимвольной интерференции. Поскольку

$$y_k = I_k + \sum_{\substack{n=0 \\ n\neq k}}^{\infty} I_n x_{k-n} + v_k . \tag{9.2.10}$$

то условие отсутствия МСИ можно записать так

$$x(t = kT) \equiv x_k = \begin{cases} 1 & (k = 0) \\ 0 & (k \neq 0). \end{cases} \tag{9.2.11}$$

Ниже мы определим необходимые и достаточные условия для $X(f)$ для того, чтобы импульс $x(t)$ удовлетворял бы вышеуказанному соотношению. Это условие известно, как *критерий отсчётности сигнала Найквиста или условие Найквиста для нулевой МСИ*. Оно формулируется следующей теоремой.

Теорема (Найквиста). Необходимое и достаточное условие для того, чтобы $x(t)$ удовлетворяло условиям

$$x(nT) = \begin{cases} 1 & (n = 0) \\ 0 & (n \neq 0) \end{cases} \tag{9.2.12}$$

сводится к тому, чтобы преобразование Фурье $X(f)$ удовлетворяло условию

$$\sum_{m=-\infty}^{\infty} X(f + m/T) = T . \tag{9.2.13}$$

Доказательство.

В общем $x(t)$ определяется обратным преобразование Фурье $X(f)$. Следовательно,

$$x(t) = \int_{-\infty}^{\infty} X(f)\, e^{j2\pi fT} df . \tag{9.2.14}$$

В точках отсчета $t = nT$ это соотношение принимает вид

$$x(nT) = \int_{-\infty}^{\infty} X(f) e^{j2\pi fnT} df . \tag{9.2.15}$$

Разобьем интеграл в (9.2.15) на интегралы, перекрывающие ограниченные области частот $1/T$. Тогда получаем

$$\begin{aligned} x(nT) &= \sum_{m=-\infty}^{\infty} \int_{(2m-1)/2T}^{(2m+1)/2T} X(f) e^{j2\pi fnT} df = \sum_{m=-\infty}^{\infty} \int_{-1/2T}^{1/2T} X(f + m/T) e^{j2\pi fnT} df = \\ &= \int_{-1/2T}^{1/2T} \left[\sum_{m=-\infty}^{\infty} X(f + m/T) \right] e^{j2\pi fnT} df = \int_{-1/2T}^{1/2T} B(f) e^{j2\pi fnT} df , \end{aligned} \tag{9.2.16}$$

где мы определили $B(f)$ так:

$$B(f) = \sum_{m=-\infty}^{\infty} X(f + m/T) . \tag{9.2.17}$$

Очевидно, что $B(f)$ является периодической функцией с периодом $1/T$ и, следовательно, ее можно представить рядом Фурье

$$B(f) = \sum_{n=-\infty}^{\infty} b_n e^{j2\pi nfT} , \tag{9.2.18}$$

где

$$b_n = T \int_{-1/2T}^{1/2T} B(f) e^{-j2\pi nfT} df . \tag{9.2.19}$$

Сравнив (9.2.19) и (9.2.16) находим

$$b_n = Tx(-nT). \tag{9.2.20}$$

Следовательно, достаточное и необходимое условие для (9.2.11) сводится к тому, что

$$b_n = \begin{cases} T & (n=0) \\ 0 & (n \neq 0). \end{cases} \tag{9.2.21}$$

Подстановка этого условия в (9.2.18) дает

$$B(f) = T \tag{9.2.22}$$

или, что эквивалентно

$$\sum_{m=-\infty}^{\infty} X(f + m/T) = T. \tag{9.2.23}$$

Этим заканчивается доказательство теоремы.

Теперь предположим, что канал имеет полосу W. Тогда $C(f) \equiv 0$ для $|f| > W$ и, следовательно, $X(f) = 0$ для $|f| > W$. Мы рассмотрим три случая.

1. Когда $T < 1/2W$ или, что эквивалентно $1/T > 2W$. Поскольку $B(f) = \sum_{n=-\infty}^{\infty} X(f + n/T)$ состоит в этом случае из неперекрывающихся копий $X(f)$, разделенных интервалом $1/T$, как показано на рис. 9.2.4, то нет возможности выбрать $X(f)$ так, чтобы удовлетворить условию $B(f) \equiv T$, и таким путем мы не сможем синтезировать систему без МСИ.

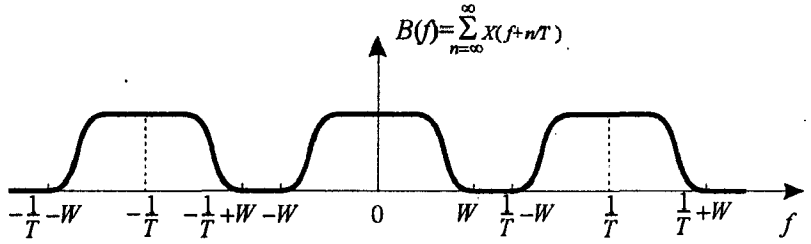


Рис. 9.2.4. Кривая $B(f)$ для случая $T<1/2F$

2. Когда $T = 1/2W$ или, что эквивалентно, $1/T = 2W$ (скорость Найквиста), повторения $X(f)$, разделенные интервалом $1/T$, показаны на рис. 9.2.5. Ясно, что в этом случае существует только одна характеристика $X(f)$, которая удовлетворяет условию $B(f) \equiv T$, именно

$$X(f) = \begin{cases} T & (|f| < W) \\ 0 & \text{при других } f. \end{cases} \tag{9.2.24}$$

что ведёт к импульсу

$$x(t) = \frac{\sin(\pi t/T)}{\pi t/T} = \operatorname{sinc}\left(\frac{\pi t}{T}\right) \tag{9.2.25}$$

Это значит, что наименьшее значение T, при котором возможна передача с нулевой МСИ, это $T = 1/2W$, и при этом условии $x(t)$ является функцией sinc. Трудности такого выбора функции $x(t)$ заключаются в том, что она некаузальная и, следовательно, не реализуемая.

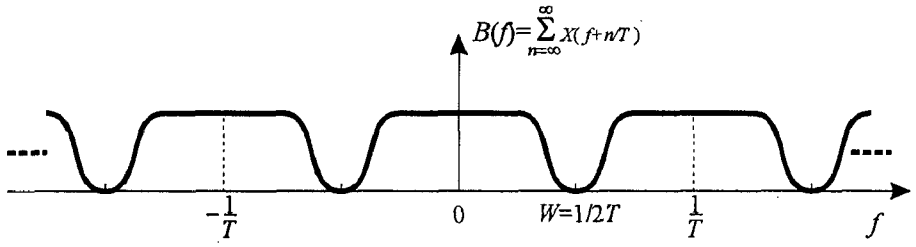


Рис. 9.2.5. Кривая $B(f)$ для случая $T=1/2F$

Чтобы сделать её реализуемой, обычно вводят задержку, т.е. используется функция $\mathrm{sinc}[\pi(t-t_0)/T]$, а t_0 выбирается так, чтобы при $t<0$ иметь $\mathrm{sinc}[\pi(t-t_0)/T]\approx 0$. Конечно, при таком выборе $x(t)$ время стробирования также надо сдвигать до $mT+t_0$. Вторая трудность с такой огибающей импульса заключается в том, что скорость её сходимости к нулю медленная. «Хвосты» от $x(t)$ затухают, как $1/T$, следовательно, малая ошибка в моменте взятия отсчёта на выходе согласованного фильтра демодулятора ведет к неограниченному ряду компонент МСИ. Такой ряд абсолютно не сходится из-за низкой скорости затухания импульса $1/T$ и, следовательно, результирующая МСИ в принципе не сходится.

3. Когда $T>1/2W$, то $B(f)$ состоит из перекрывающихся копий $X(f)$, разделенных интервалом $1/T$, как показано на рис. 9.2.6. В этом случае существует несчётное число выборов $X(f)$, при которых $B(f)\equiv T$.

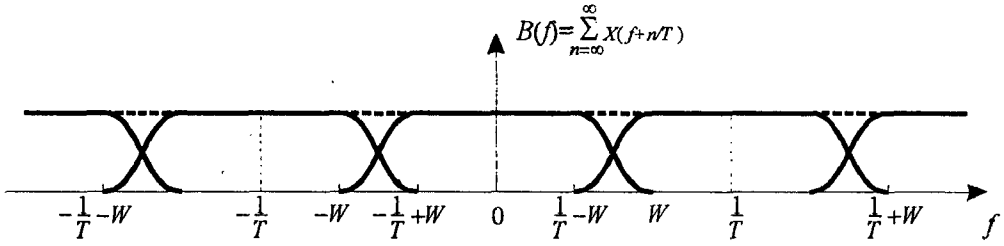


Рис. 9.2.6. Кривая $B(f)$ для случая $T>1/2F$

Частный спектр импульса для случая $T>1/2W$, который имеет требуемые спектральные свойства и который широко используется на практике – это спектр приподнятого косинуса. Частотная характеристика приподнятого косинуса (см. задачу 9.11) определяется так:

$$X_{rc}(f)=\begin{cases} T. & \left(0\le|f|\le\dfrac{1-\beta}{2T}\right) \\ \dfrac{T}{2}\left\{1+\cos\left[\dfrac{\pi T}{\beta}\left(|f|-\dfrac{1-\beta}{2T}\right)\right]\right\} & \left(\dfrac{1-\beta}{2T}\le|f|\le\dfrac{1+\beta}{2T}\right), \\ 0 & \left(|f|>\dfrac{1+\beta}{2T}\right) \end{cases} \qquad (9.2.26)$$

где β называется *коэффициентом ската* и принимает значение в области $0\le\beta\le 1$. Полоса, которую сигнал занимает вне полосы Найквиста $1/2T$, называют *излишком* полосы, и он

обычно выражается в процентах от полосы частот Найквиста. Например, если $\beta = \frac{1}{2}$, излишек полосы равен 50%, а когда $\beta = 1$, излишек полосы равен 100%. Импульс $x(t)$, имеющий спектр приподнятого косинуса, определяется так:

$$x(t) = \frac{\sin(\pi t/T)}{\pi t/T}\frac{\cos(\pi\beta t/T)}{1-4\beta^2 t^2/T^2} = \mathrm{sinc}\,(\pi t/T)\frac{\cos(\pi\beta t/T)}{1-4\beta^2 t^2/T^2}. \qquad (9.2.27)$$

Заметим, что $x(t)$ нормирован так, что $x(0) = 1$. Рис. 9.2.7 иллюстрирует спектральные характеристики приподнятого косинуса и соответствующие импульсы для $\beta = 0$, $\frac{1}{2}$ и 1.

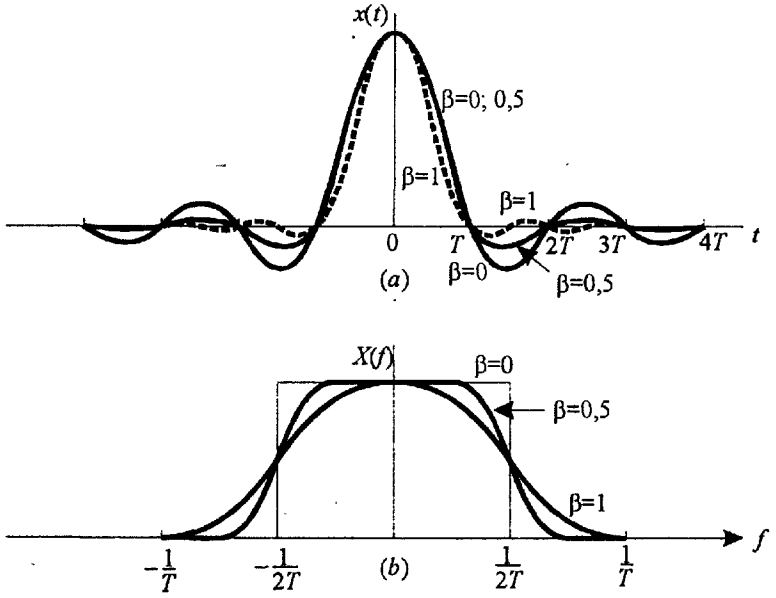


Рис 9.2.7. Импульсы, имеющие спектр типа приподнятого косинуса

Заметим, что при $\beta = 0$ импульс вырождается в $x(t) = \mathrm{sinc}\,(\pi t/T)$, а скорость передачи символов равна $1/T = 2W$. Если $\beta = 1$ скорость передачи символов $1/T = W$. В общем, хвосты $x(t)$ убывают как $1/t^3$ для $\beta > 0$. Следовательно, ошибка при стробировании отсчётов ведет к ряду компонент МСИ, который сходится к конечной величине.

С учетом «гладкостных» характеристик спектра приподнятого косинуса возможно синтезировать реализуемые фильтры для передатчика и приёмника, которые аппроксимируют желательные суммарные частотные характеристики. В частном случае, когда канал идеален, т.е. $C(f) = 1$, $|f| \le W$, имеем

$$X_{rc}(f) = G_T(f)G_R(f), \qquad (9.2.28)$$

где $G_T(f)$ и $G_R(f)$ – частотные характеристики двух указанных фильтров. В том случае, когда фильтр на приёме согласован с фильтром на передаче, имеем

$$X_{rc}(f) = G_T(f)G_R(f) = |G_T(f)|^2.$$

В идеале

$$G_T(f) = \sqrt{|X_{rc}(f)|}\,e^{-j2\pi f t_0}, \qquad (9.2.29)$$

и $G_R(f) = G_T^*(f)$, а t_0 – некоторая номинальная задержка, которая должна удовлетворить физической реализуемости фильтра. Таким образом, суммарная спектральная характеристика приподнятого косинуса расщепляется поровну между фильтрами передатчика и приёмника. Подчеркнем также, что необходима дополнительная задержка для обеспечения физической реализуемости фильтра на приёме.

9.2.2. Синтез ограниченных по полосе символов с контролируемой МСИ – сигналы с парциальным откликом

Как мы видели из нашего обсуждения синтеза сигналов, для получения нулевой МСИ необходимо сократить скорость передачи символов $1/T$ ниже скорости Найквиста $2W$ символов/с для практической реализации фильтров на передаче и приёме. С другой стороны, предположим, что мы снимем требование нулевой МСИ и, таким образом, достигнем скорости передачи символов $2W$ символов/с. Допуская контролируемую величину МСИ, мы можем достичь эту скорость передачи символов.

Мы уже видели, что условием для нулевой МСИ является $x(nT) = 0$ для $n \neq 0$. Однако предположим, что мы синтезируем ограниченный по полосе сигнал с контролируемой МСИ в определённый момент времени. Это означает, что мы допускаем некоторую дополнительную ненулевую величину МСИ в отсчётах $\{x(nT)\}$, которую мы вводим. Она детерминирована или «контролируема» и может принять определённую величину в приёмнике, что обсудим ниже.

Один частный случай, который ведет (приближённо) к физически реализуемым фильтрам передатчика и приёмника, определяется нормированными отсчетами[1]

$$x(nT) = \begin{cases} 1 & (n = 0,\ 1) \\ 0 & (\text{при других } n). \end{cases} \tag{9.2.30}$$

Теперь, используя (9.2.20), мы получим

$$b_n = \begin{cases} T & (n = 0,\ -1) \\ 0 & (\text{при других } n), \end{cases} \tag{9.2.31}$$

что, при подстановки в (9.2.18) дает

$$B(f) = T + T\mathrm{e}^{-j2\pi fT}. \tag{9.2.32}$$

Как и в предыдущем разделе, невозможно удовлетворить этому уравнению при $T < 1/2W$. Однако для $T = 1/2W$ мы получим

$$X(f) = \begin{cases} \dfrac{1}{2W}\left(1 + \mathrm{e}^{-j\pi f/W}\right) & (|f| < W) \\ 0 & (\text{при других } f) \end{cases} = \begin{cases} \dfrac{1}{W}\mathrm{e}^{-j\pi f/2W}\cos\dfrac{\pi f}{2W} & (|f| < W) \\ 0 & (\text{при других } f). \end{cases} \tag{9.2.33}$$

Следовательно, $x(t)$ определится так

$$x(t) = \mathrm{sinc}(2\pi Wt) + \mathrm{sinc}\left[2\pi\left(Wt - \tfrac{1}{2}\right)\right]. \tag{9.2.34}$$

Этот импульс называется дуобинарным сигнальным импульсом. Он иллюстрируется вместе со спектром амплитуд на рис. 9.2.8.

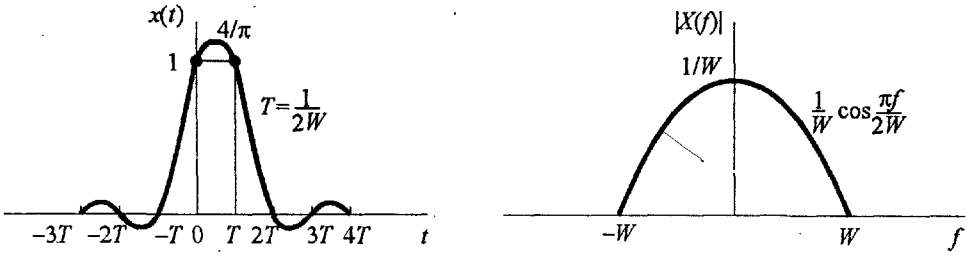


Рис. 9.2.8. Характеристика во временной и частотной областях дуобинарного сигнала

[1] Удобно иметь дело с отсчетами $x(t)$, которые нормированы к единице для $n = 0, 1$.

Заметим, что спектр падает до нуля плавно, что означает, что можно синтезировать физически реализуемые фильтры, которые аппроксимируют этот спектр очень плотно. Таким образом достигается скорость передачи $2W$.

Другой частный случай, который ведет (приближенно) к физически реализуемым фильтрам передатчика и приёмника, определяется отсчётами

$$x\left(\frac{n}{2W}\right) = x(nT) = \begin{cases} 1 & (n=-1) \\ -1 & (n=1) \\ 0 & \text{при других } n. \end{cases} \tag{9.2.35}$$

Соответствующий импульс $x(t)$ определяется так:

$$x(t) = \operatorname{sinc}\left[\frac{\pi(t+T)}{T}\right] - \operatorname{sinc}\left[\frac{\pi(t-T)}{T}\right], \tag{9.2.36}$$

а его спектр

$$X(f) = \begin{cases} \dfrac{1}{2W}\left(e^{j\pi f/W} - e^{-j\pi f/W}\right) = \dfrac{j}{W}\sin\dfrac{\pi f}{W} & |f| \le W \\ 0 & |f| > W. \end{cases} \tag{9.2.37}$$

Этот импульс и его спектр амплитуд иллюстрируются на рис. 9.2.9. Он называется *модифицированным дуобинарным импульсом* сигнала. Интересно отметить, что спектр этого сигнала равен нулю при $f = 0$, что делает его подходящим для передачи по каналу, который не пропускает постоянную составляющую.

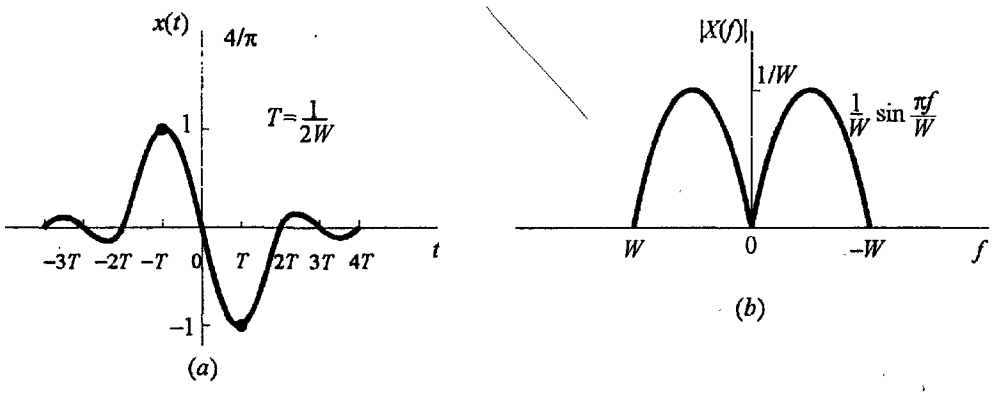


Рис. 9.2.9. Характеристика во временной и частотной областях модифицированного дуобинарного сигнала

Можно получить другие интересные и физически реализуемые характеристики фильтров, как показано Кречмером (1966) и Лакки и др. (1968), выбирая различные значения для отсчетов $\{x(n/2W)\}$ и больше, чем два ненулевых отсчета. Однако, если мы выберем больше ненулевых отсчетов, то проблема отслеживания контролируемой МСИ становится более трудной и практически неразрешимой.

В общем, класс ограниченных по полосе импульсов сигналов, имеющих форму

$$x(t) = \sum_{n=-\infty}^{\infty} x\left(\frac{n}{2W}\right)\operatorname{sinc}\left[2\pi W\left(t - \frac{n}{2W}\right)\right] \tag{9.2.38}$$

и соответствующий спектр

$$X(f) = \begin{cases} \dfrac{1}{2W}\displaystyle\sum_{n=-\infty}^{\infty} x\left(\frac{n}{2W}\right)e^{-jn\pi f/W} & (|f| \le W) \\ 0 & (|f| > W) \end{cases} \tag{9.2.39}$$

называют сигналами с парциальным откликом, когда контролируемая МСИ намеренно вводиться отбором двух или больше ненулевых отсчетов из ансамбля $\{x(n/2W)\}$. Результирующий сигнальный импульс позволяет нам передавать информационные символы со скоростью Найквиста $2W$ символов/с. Детектирование принимаемых символов в присутствии контролируемой МСИ описывается ниже.

Альтернативное представление сигналов с парциальным откликом. Мы включили этот подраздел для представления других интерпретаций сигналов с парциальным откликом. Предположим, что сигнал с парциальным откликом генерируется так, как показано на рис. 9.2.10, путём прохождения последовательности $\{I_n\}$ с дискретным временем через линейный фильтр с дискретным временем и с коэффициентами $x_n \equiv x(n/2W)$, $n = 0, 1, ..., N-1$ и использовании выходной последовательности $\{B_n\}$ этого фильтра для периодической подаче ... на аналоговый фильтр с импульсной характеристикой $\mathrm{sinc}(2\pi Wt)$.

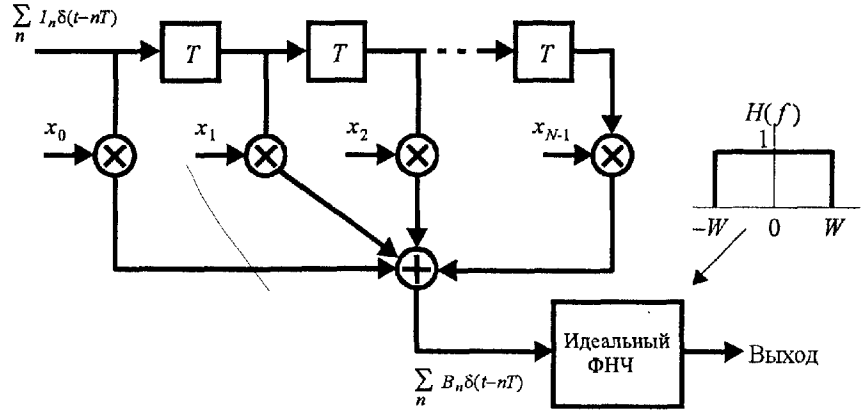


Рис. 9.2.10. Альтернативный метод формирования сигнала с парциальным откликом

Результирующий выходной сигнал фильтра идентичен сигналу с парциальным откликом (9.2.38)

Поскольку

$$B_n = \sum_{k=0}^{N-1} x_k I_{n-k}\,, \tag{9.2.40}$$

то последовательность символов $\{B_n\}$ коррелированна вследствие фильтрации последовательности $\{I_n\}$. Действительно, автокорреляционная функция последовательности $\{B_n\}$ равна

$$\phi(m) = E(B_n B_{n+m}) = \sum_{k=0}^{N-1}\sum_{l=0}^{N-1} x_k x_l E(I_{n-k} I_{n+m-l})\,. \tag{9.2.41}$$

Когда входная последовательность имеет нулевое среднее и равномерный спектр, то

$$E(I_{n-k} I_{n+m-l}) = \delta_{m+k-l}\,, \tag{9.2.42}$$

где мы использовали нормирование $E(I_n^2) = 1$. Подстановка (9.2.42) в (9.2.41) приводит к желательной автокорреляционной функции $\{B_n\}$ в виде

$$\phi(m) = \sum_{k=0}^{N-1-|m|} x_k x_{k+|m|}, \qquad m = 0, \pm 1, ..., \pm(N-1)\,. \tag{9.2.43}$$

Соответствующая спектральная плотность мощности равна

$$\Phi(f) = \sum_{m=-(N-1)}^{N-1} \phi(m) e^{-j2\pi fmT} = \left| \sum_{m=0}^{N-1} x_m e^{-j2\pi fmT} \right|^2, \qquad (9.2.44)$$

где $T = 1/2W$ и $f \le 1/2T = W$.

9.2.3. Детектирование данных при контролируемой МСИ

В этом разделе мы опишем два метода детектирования информационных символов на приёме, когда принимаемый сигнал содержит контролируемую МСИ. Одним из них является посимвольный метод детектирования, который относительно легко реализовать. Второй метод базируется на правиле максимального правдоподобия для детектирования последовательности символов. Последний метод минимизирует вероятность ошибки, но он немного сложнее для реализации. В частности, мы рассмотрим детектирование дуобинарных и модифицированных дуобинарных сигналов с парциальным откликом. В обоих случаях мы предполагаем, что желательная спектральная характеристика $X(f)$ для сигнала с парциальным откликом распределена поровну между фильтрами передатчика и приёмника, т.е. $|G_T(f)| = |G_R(f)| = |X(f)|^{1/2}$. Наша трактовка базируется на сигналах АМ, но она легко обобщается на КАМ и ФМ.

Посимвольное субоптимальное детектирование. Для дуобинарного импульса сигнала $x(nT) = 1$ для $n = 0, 1$ и $x(nT) = 0$ в других точках. Следовательно, отсчёты на выходе фильтра приёмника (демодулятора) имеют вид

$$y_m = B_m + \nu_m = I_m + I_{m-1} + \nu_m, \qquad (9.2.45)$$

где $\{I_m\}$ передаваемая последовательность амплитуд, а $\{\nu_m\}$ – последовательность отсчетов аддитивного гауссовского шума. Пренебрежем на время шумом и рассмотрим двоичный случай, когда $I_m = \pm 1$ с равной вероятностью. Тогда B_m принимает одно из трех возможных значений, именно $B_m = -2, 0, 2$ с соответствующими вероятностями $1/4$, $1/2$, $1/4$. Если I_{m-1} – продетектированный символ на $(m-1)$-м тактовом интервале, то он влияет на B_m, и принимаемый сигнал на m-м тактовом интервале можно восстановить вычитанием, что позволяет продетектировать I_m. Этот процесс можно повторять последовательно для каждого принимаемого символа.

Важнейшая проблема при использовании этой процедуры заключается в том, что ошибки, возникающие от действия шума, имеют тенденцию размножаться. Например, если I_{m-1} принят с ошибкой, то его влияние на B_m усиливается из-за неправильного вычитания. Следовательно, высока вероятность того, что и детектирование B_m будет ошибочным.

Размножение ошибок может быть преодолено путем предварительного кодирования данных на передаче вместо ограничения контролируемого МСИ путем вычитания на приёме. Предварительное кодирование выполняется над последовательностью двоичных данных до модуляции. Из последовательности $\{D_n\}$ из 1 и 0, которые должны быть преданы, генерируется новая последовательность $\{P_n\}$, называемая предварительно кодированной последовательностью. Для дуобинарного сигнала, предварительно кодированная последовательность определяется так

$$P_m = D_m \ominus P_{m-1}, \quad m = 1, 2, \ldots, \qquad (9.2.46)$$

где $\ominus$ означает вычитание по модулю 2[1]. Тогда мы полагаем $I_m = -1$, если $P_m = 0$ и $I_m = 1$, если $P_m = 1$, т.е. $I_m = 2P_m - 1$. Заметим, что эта операция предварительного кодирования идентична той, которая описана в разделе 4.3.2 в контексте нашего обсуждения двоичного сигнала без возврата к нулю с памятью (ДБНП или NRZI).

Свободные от шума отсчеты на выходе фильтра приёмника равны

$$B_m = I_m + I_{m-1} = (2P_m - 1) + (2P_{m-1} - 1) = 2(P_m + P_{m-1} - 1). \qquad (9.2.47)$$

Следовательно,

$$P_m + P_{m-1} = \tfrac{1}{2}B_m + 1. \qquad (9.2.48)$$

Поскольку $D_m = P_m \oplus P_{m-1}$, то следует, что последовательность D_m получается из B_m посредством отношения

$$D_m = \tfrac{1}{2}B_m + 1 \pmod 2. \qquad (9.2.49)$$

Следовательно, если $B_m = \pm 2$ тогда $D_m = 0$ и, если $B_m = 0$, то $D_m = 1$.

Пример, который использует операции предварительного кодирования и декодирования дан в табл. 9.2.1.

Табл. 9.2.1. Передача двоичных символов посредством дуобинарных импульсов.

Данные последовательности D_n		1	1	1	0	1	0	0	1	0	0	0	1	1	0	1
Предварительно кодированная последовательность P_n	0	1	0	1	1	0	0	0	1	1	1	1	0	1	1	0
Переданная последовательность I_m	-1	1	-1	1	1	-1	-1	-1	1	1	1	1	-1	1	1	-1
Принятая последовательность B_n		0	0	0	2	0	-2	-2	0	2	2	2	0	0	2	0
Декодированная последовательность P_n		1	1	1	0	1	0	0	1	0	0	0	1	1	0	1

В присутствии аддитивного шума, выходные отсчеты фильтра приёмника определяются согласно (9.2.45). В этом случае $y_m = B_m + \nu_m$ сравнивается с двумя пороговыми уровнями $+1$ и -1. Последовательность данных $\{D_n\}$ получается согласно правилу детектирования

$$D_m = \begin{cases} 1 & (|y_m| < 1) \\ 0 & (|y_m| \geq 1) \end{cases} \qquad (9.2.50)$$

Расширение от двоичной АМ до многоуровневой АМ с использованием дуобинарных импульсов выполняется непосредственно. В этом случае последовательность $\{I_m\}$ с M-уровневыми амплитудами приводит к последовательности (при отсутствии шума)

$$B_m = I_m + I_{m-1}, \quad m = 1, 2, \ldots \qquad (9.2.51)$$

с $2M - 1$ возможными равноудаленными уровнями. Уровни амплитуд определяются из отношения

$$I_m = 2P_m - (M - 1), \qquad (9.2.52)$$

[1] Хотя это идентично суммированию по mod 2, удобно рассмотреть операцию предварительного кодирования для дуобинарного сигнала как вычитание по mod 2.

где $\{P_m\}$ – предварительно кодированная последовательность, которая получается из M -уровневой последовательности данных $\{D_m\}$ согласно отношению

$$P_m = D_m \ominus P_{m-1} \pmod{M}, \tag{9.2.53}$$

где возможные значения последовательности $\{D_m\}$ равны 0,1,2...M–1.

В отсутствии шума, отсчёты на выходе фильтра приёмника можно выразить так

$$B_m = I_m + I_{m-1} = 2[P_m + P_{m-1} - (M-1)]. \tag{9.2.54}$$

Следовательно,

$$P_m + P_{m-1} = \tfrac{1}{2} B_m + (M-1). \tag{9.2.55}$$

Поскольку $D_m = P_m \oplus P_{m-1} \pmod{M}$ следует

$$D_m = \tfrac{1}{2} B_m + (M-1) \pmod{M}. \tag{9.2.56}$$

Пример, иллюстрирующий многоуровневое предварительное кодирование и декодирование, дан в табл. 9.2.2.

Табл.9.2.2. Передача 4-позиционных символов посредством дуобинарных импульсов

Данные последовательности D_n		0	0	1	3	1	2	0	3	3	2	0	1	0
Предварительная кодированная последовательность P_n	0	0	0	1	2	3	3	1	2	1	1	3	2	2
Переданная последовательность I_m	-3	-3	-3	-1	1	3	3	-1	1	-1	-1	3	1	1
Принятая последовательность B_n		-6	-6	-4	0	4	6	2	0	0	-2	2	4	2
Декодированная последовательность P_n		0	0	1	3	1	2	0	3	3	2	0	1	0

В присутствии шума принимаемый сигнал квантуется к ближайшим возможным уровням сигнала и правило, данное выше, применяется к квантовым значениям для восстановления последовательности данных.

В случае модифицированного дуобинарного импульса, контролируемая МСИ определяется величинами $x(n/2W) = -1$ для $n = 1$, $x(n/2W) = 1$ для $n = -1$ и нулю в других точках. Следовательно, свободные от шума отсчеты выхода фильтра приёмника определяются так

$$B_m = I_m - I_{m-2}, \tag{9.2.57}$$

где M -уровневая последовательность $\{I_m\}$ получается путем отображения предварительно кодированной последовательности согласно отношению (9.2.52) и

$$P_m = D_m \oplus P_{m-2} \pmod{M}. \tag{9.2.58}$$

Из этих соотношений легко показать, что правило детектирования для восстановления последовательности данных $\{D_m\}$ по $\{B_m\}$ в отсутствии шума таково

$$D_m = \tfrac{1}{2} B_m \pmod{M}. \tag{9.2.59}$$

Как показано выше, предварительное кодирование данных на передаче делает возможным детектировать принимаемые данные посимвольно без необходимости

оглядываться назад на предшествующие детектированные символы. Таким образом преодолевается размножение ошибок.

Посимвольное правило детектирования, описанное выше, не реализует оптимальную схему детектирования для сигналов с парциальным откликом, из-за памяти, свойственной принимаемому сигналу. Все же посимвольное детектирование относительно просто реализуется и оно применяется во многих практических приложениях, использующих дуобинарный и модифицированный сигнальные импульсы. Его качество вычисляется в следующем разделе.

Максимально-правдоподобное последовательное детектирование. Ясно из предыдущего обсуждения, что сигналы с парциальным откликом являются сигналами с памятью. Эту память удобно представить решеткой. Например, на рис. 9.2.11 представлена решетка для дуобинарного сигнала с парциальным откликом для передачи двоичных данных.

Для двоичной модуляции решетка содержит два состояния, соответствующие двум возможным выходным величинам I_m, т.е. $I_m = \pm 1$. Каждая ветвь в решетке обозначается двумя числами. Первое число слева – это новый символ, т.е. $I_{m+1} = \pm 1$. Это число определяет переход в новое состояние. Число справа – это уровень принимаемого сигнала.

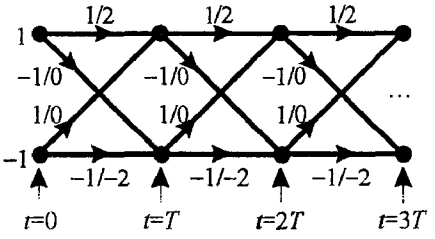


Рис. 9.2.11. Решётка для дуобинарного сигнала с парциальным откликом

Дуобинарный сигнал имеет память длины $L = 1$.. Следовательно, для двоичной модуляции решетка имеет $S_t = 2$ состояния. В общем, для M-ичной модуляции, число состояний решетки равно M^L.

Оптимальный детектор максимального правдоподобия (МП) последовательности выбирает наиболее правдоподобный путь по решетке после наблюдения последовательности принимаемых данных $\{y_m\}$.в отсчетных моментах $t = mT, \quad m = 1, 2, \ldots$ В общем, каждый узел решетки имеет M входящих путей и M соответствующих метрик. Основываясь на величинах метрик, один вход из M приходящих путей выбирается как наиболее вероятный, а остальные $M-1$ путей и их метрики отбрасываются. Выживший путь в каждом узле затем продолжается в M новых путях, один для каждого из M возможных входных символов, и процесс поиска продолжается. Это в своей основе алгоритм Витерби для выполнения поиска по решетке.

Для класса сигналов с парциальным откликом принимаемая последовательность $\{y_m, 1 \le m \le N\}$ в общем описывается статически совместной ФПВ $f(\mathbf{y}_N|\mathbf{I}_N)$, где $\mathbf{y}_N = [y_1 \; y_2 \ldots y_N]^T$ и $\mathbf{I}_N = [I_1 \; I_2 \ldots I_N]^T$ и $N > L$. Если аддитивный шум гауссовский с нулевым средним, $f(\mathbf{y}_N|\mathbf{I}_N)$ является многомерной гауссовской ФПВ, т.е.

$$f(\mathbf{y}_N|\mathbf{I}_N) = \frac{1}{(2\pi \det \mathbf{C})^{N/2}} \exp\left[-\tfrac{1}{2}(\mathbf{y}_N - \mathbf{B}_N)^T \mathbf{C}^{-1}(\mathbf{y}_N - \mathbf{B}_N)\right], \qquad (9.2.60$$

где $\mathbf{B}_N = [B_1 \; B_2 \ldots B_N]^T$ – среднее значение вектора $\mathbf{y}_N$, а $\mathbf{C}$ является $N \times N$ матрицей ковариации для $\mathbf{y}_N$. Затем МП детектор последовательности выбирает последовательность по решетке, которая максимизирует ФПВ $f(\mathbf{y}_N|\mathbf{I}_N)$.

Расчёт для нахождения наиболее правдоподобной последовательности по решетке упрощается, если взять натуральный логарифм от $f(\mathbf{y}_N|\mathbf{I}_N)$. Тогда

$$\ln f(\mathbf{y}_N|\mathbf{I}_N) = -\tfrac{1}{2}N\ln(2\pi\det\mathbf{C}) - \tfrac{1}{2}(\mathbf{y}_N - \mathbf{B}_N)^{\mathrm{T}}\mathbf{C}^{-1}(\mathbf{y}_N - \mathbf{B}_N). \qquad (9.2.61)$$

При заданной принимаемой последовательности $\{y_m\}$ нахождение последовательности $\{I_m\}$, которая максимизирует $\ln f(\mathbf{y}_N|\mathbf{I}_N)$ идентично нахождению последовательности $\{I_m\}$, которая минимизирует $(\mathbf{y}_N - \mathbf{B}_N)^{\mathrm{T}}\mathbf{C}^{-1}(\mathbf{y}_N - \mathbf{B}_N)$, т.е.

$$\hat{\mathbf{I}}_N = \arg\min_{\mathbf{I}_N}\left[(\mathbf{y}_N - \mathbf{B}_N)^{\mathrm{T}}\mathbf{C}^{-1}(\mathbf{y}_N - \mathbf{B}_N)\right]. \qquad (9.2.62)$$

Вычисление метрик при поиске по решетке усложняется из-за корреляции отсчетов шума на выходе согласованного фильтра для сигнала с парциальным откликом. Для примера, в случае дуобинарного сигнала, корреляция шумовой последовательности $\{\nu_m\}$ простирается на два соседних сигнала. Следовательно, ν_m и ν_{m+k} коррелированны для $k=1$ и некоррелированы для $k>1$. В общем, сигнал с парциальным откликом с памятью L приводит к коррелированной шумовой последовательности на выходе согласованного фильтра, которая удовлетворяет условию $E(\nu_m\nu_{m+k})=0$ для $k>L$. В этом случае алгоритм Витерби для выполнения поиска по решетке может быть модифицирован, как описывается в главе 10.

Некоторое упрощение в вычислении метрик получается, если мы пренебрежем корреляцией шума, предположив, что $E(\nu_m\nu_{m+k})=0$ для $k>0$. Тогда, при нашем предположении матрица ковариаций $\mathbf{C}=\sigma_\nu^2\mathbf{1}_N$, где $\sigma_\nu^2 = E(\nu_m^2)$, а $\mathbf{1}_N$ является $N\times N$ единичной матрицей[1]. В этом случае (9.2.62) упрощается

$$\hat{\mathbf{I}}_N = \arg\min_{\mathbf{I}_N}\left[(\mathbf{y}_N - \mathbf{B}_N)^{\mathrm{T}}(\mathbf{y}_N - \mathbf{B}_N)\right] = \arg\min_{\mathbf{I}_N}\left[\sum_{m=1}^{N}\left(y_m - \sum_{k=0}^{L}x_k I_{m-k}\right)^2\right], \qquad (9.2.63)$$

где $B_m = \sum_{k=0}^{L}x_k I_{m-k}$, а $x_k = x(kT)$ – отсчётные значения сигналов с парциальным откликом. В этом случае вычисление метрик в каждом узле решетки выполняется так

$$DM_m(I_m) = DM_{m-1}(I_{m-1}) + \left(y_m - \sum_{k=0}^{L}x_k I_{m-k}\right)^2, \qquad (9.2.64)$$

где $DM_m(\mathbf{I}_m)$ – дистанционные метрики в момент $t=mT$, $DM_{m-1}(\mathbf{I}_{m-1})$ – дистанционные метрики в момент $t=(m-1)T$, а второе слагаемое в правой части (9.2.64) представляет новое приращение к метрике, основанное на новый принимаемый отсчет y_m.

Как указано в разделе 5.1.4, МП декодирование последовательности приводит, вообще говоря, к переменной задержке при декодировании каждого кодового информационного символа. На практике, переменная задержки преодолевается путем усечения выживших последовательностей до N_t последних символов, где $N_t >> 5L$, так достигается фиксированная задержка решения. Для случая, когда M^L выживших последовательностей в момент $t=mT$ определяют разные значения для символа I_{m-N_t}, то символ может быть выбран по наиболее вероятной выжившей последовательности. Потери в качестве, связанные с таким усечением, пренебрежимо малы, если $N_t > 5L$.

9.2.4. Синтез сигналов для каналов с искажениями

В разделах 9.2.1. и 9.2.2 мы описали правило синтеза сигналов (фильтров) для фильтра

[1] Мы использовали обозначение $\mathbf{1}_N$, чтобы не спутать с $\mathbf{I}_N$.

модулятора на передаче и фильтра демодулятора на приёме для случая идеального канала. В этом разделе мы выполним синтез сигналов при условии, что канал искажает передаваемый сигнал. Мы предполагаем, что частотная характеристика канала $C(f)$ известна для $|f| \le W$ и что $C(f) = 0$ для $|f| > W$. Правило оптимизации характеристик фильтров $G_T(f)$ и $G_R(f)$ обеспечивает максимизацию ОСШ на выходе фильтра демодулятора или, что эквивалентно, на входе детектора. Аддитивный шум канала предполагается гауссовским со спектральной плотностью мощности $\Phi_{nn}(f)$. Рис. 9.2.12 иллюстрирует всю рассматриваемую систему.

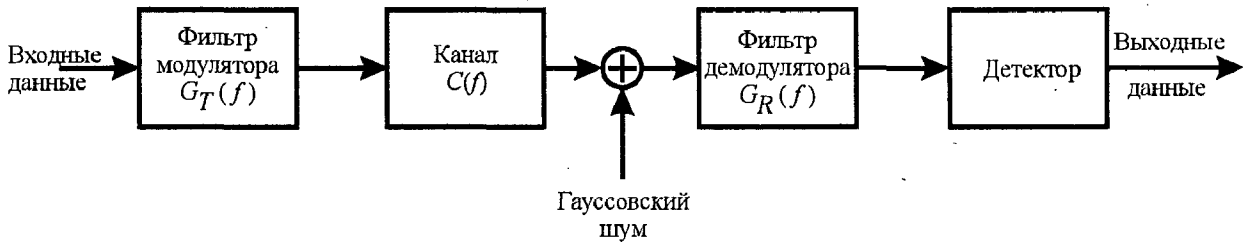


Рис. 9.2.12. Модель системы для синтеза фильтров модулятора и демодулятора

Сигнальные компоненты на выходе фильтра демодулятора должны удовлетворять условию

$$G_T(f)C(f)G_R(f) = X_d(f)\,e^{-j2\pi f t_0}, \quad |f| \le W, \tag{9.2.65}$$

где $X_d(f)$ является желательной частотной характеристикой каскада из модулятора, канала и демодулятора, а t_0 – время задержки, которое необходимо для удовлетворения физической реализуемости фильтров модулятора и демодулятора. Желательную частотную характеристику $X_d(f)$ можно выбрать так, чтобы обеспечить или нулевую МСИ или контролируемую МСИ в точках отсчёта. Мы выполним оптимизацию для нулевой МСИ, выбирая $X_d(f) = X_{rc}(f)$, где $X_{rc}(f)$ – спектр приподнятого косинуса с произвольным коэффициентом ската.

Шум на выходе фильтра демодулятора можно выразить так

$$v(t) = \int_{-\infty}^{\infty} n(t-\tau)g_R(\tau)d\tau, \tag{9.2.66}$$

где $n(t)$ – шум на входе фильтра. Поскольку $n(t)$ – гауссовский процесс с нулевым средним, то и $v(t)$ – гауссовский процесс с нулевым средним и со спектральной плотностью мощности

$$\Phi_{vv}(f) = \Phi_{nn}(f)|G_R(f)|^2. \tag{9.2.67}$$

Для простоты мы рассмотрим передачу двоичной АМ. Тогда выходные отсчёты согласованного фильтра

$$y_m = x_0 I_m + v_m = I_m + v_m, \tag{9.2.68}$$

где x_0 нормирована[1] к единице, $I_m = \pm d$, а v_m представляет слагаемое шума, которое является гауссовским с нулевым средним и дисперсией

$$\sigma_v^2 = \int_{-\infty}^{\infty} \Phi_{nn}(f)|G_R(f)|^2 df. \tag{9.2.69}$$

Следовательно, вероятность ошибки равна

[1] Взяв $x_0 = 1$ и $I_m = \pm d$ имеем в виду, что масштаб x_0 учтён в параметре d.

$$P_2 = \frac{1}{\sqrt{2\pi}} \int_{d/\sigma_v}^{\infty} e^{-y^2/2} dy = Q\left(\sqrt{d^2/\sigma_v^2}\right). \qquad (9.2.70)$$

Вероятность ошибки минимизируется при максимизации $\text{ОСШ} = d^2/\sigma_v^2$ или, что эквивалентно, минимизации отношение шум/сигнал σ_v^2/d^2. Но d^2 связан со средней мощностью переданного сигнала так:

$$P_{\text{ср}} = \frac{E(I_m^2)}{T} \int_{-\infty}^{\infty} g_T^2(t)dt = \frac{d^2}{T} \int_{-\infty}^{\infty} g_T^2(t)dt, \quad \frac{1}{d^2} = \frac{1}{P_{\text{ср}} T} \int_{-\infty}^{\infty} |G_T(f)|^2 df. \qquad (9.2.71)$$

$G_T(f)$ должно быть выбрано так, чтобы удовлетворять условию нулевой МСИ. Следовательно,

$$|G_T(f)| = \frac{|X_{rc}(f)|}{|C(f)||G_R(f)|}, \qquad |f| \le W \qquad (9.2.72)$$

и $G_T(f) = 0$ для $|f| \ge W$. Следовательно,

$$\frac{1}{d^2} = \frac{1}{P_{\text{ср}} T} \int_{-W}^{W} \frac{|X_{rc}(f)|^2}{|C(f)|^2 |G_R(f)|^2} df. \qquad (9.2.73)$$

Таким образом, отношение шум/сигнал, которое должно быть минимизировано по $|G_R(f)|$ для $|f| \le W$, равно

$$\frac{\sigma_v^2}{d^2} = \frac{1}{P_{\text{ср}} T} \int_{-W}^{W} \Phi_{nn}(f) |G_R(f)|^2 df \cdot \int_{-W}^{W} \frac{|X_{rc}(f)|^2}{|C(f)|^2 |G_R(f)|^2} df. \qquad (9.2.74)$$

Оптимальное значение $|G_R(f)|$ можно найти, использовав неравенство Коши-Шварца

$$\int_{-\infty}^{\infty} |U_1(f)|^2 df \cdot \int_{-\infty}^{\infty} |U_2(f)|^2 df \ge \left[\int_{-\infty}^{\infty} |U_1(f)||U_2(f)| df \right]^2, \qquad (9.2.75)$$

где $|U_1(f)|$ и $|U_2(f)|$ определены так:

$$|U_1(f)| = \left|\sqrt{\Phi_{nn}(f)}\right| |G_R(f)|, \qquad |U_2(f)| = \frac{|X_{rc}(f)|}{|C(f)||G_R(f)|}. \qquad (9.2.76)$$

Минимальное значение (9.2.74) получается, если $|U_1(f)|$ пропорционально $|U_2(f)|$ или, что эквивалентно, когда

$$|G_R(f)| = K \frac{|X_{rc}(f)|^{1/2}}{|\Phi_{nn}(f)|^{1/4} |C(f)|^{1/2}}, \quad |f| \le W, \qquad (9.2.77)$$

где K – произвольная константа. Соответствующий фильтр модулятора имеет амплитудную характеристику

$$|G_T(f)| = \frac{1}{K} \frac{|X_{rc}(f)|^{1/2} |\Phi_{nn}(f)|^{1/4}}{|C(f)|^{1/2}}, \quad |f| \le W. \qquad (9.2.78)$$

Наконец, максимум ОСШ, достигаемый этими оптимальными на передаче и приёме фильтрами

$$\frac{d^2}{\sigma_v^2} = \frac{P_{\text{ср}} T}{\left\{ \int_{-W}^{W} |X_{rc}(f)| [\Phi_{nn}(f)]^{1/2} |C(f)|^{-1} df \right\}^2}. \qquad (9.2.79)$$

Заметим, что оптимальные фильтры модулятора и демодулятора определены только по амплитудно-частотной характеристике. Фазовые характеристики $G_T(f)$ и $G_R(f)$ можно выбрать так, чтобы удовлетворить условию (9.2.65), т.е.

$$\Theta_T(f)+\Theta_c(f)+\Theta_R(f)=2\pi f t_0, \qquad (9.2.80)$$

где $\Theta_T(f)$, $\Theta_c(f)$, $\Theta_R(f)$ – фазовые характеристики фильтра модулятора, канала и фильтра демодулятора, соответственно.

В частном случае, когда аддитивный шум на входе демодулятора гауссовский и белый со спектральной плотностью мощности $\frac{1}{2}N_0$, оптимальные характеристики фильтров, определяемые (9.2.77) и (9.2.78), выражаются так

$$|G_R(f)|=K_1\frac{|X_{rc}(f)|^{1/2}}{|C(f)|^{1/2}}, \quad |f|\le W$$
$$|G_T(f)|=K_2\frac{|X_{rc}(f)|^{1/2}}{|C(f)|^{1/2}}, \quad |f|\le W, \qquad (9.2.81)$$

где K_1 и K_2 – произвольные скалярные множители. Заметим, что в этом случае $|G_R(f)|$ – АЧХ фильтра, согласованного с фильтром, имеющим характеристику $|G_T(f)|$. Соответствующее ОСШ детектора, определяемые (9.2.79), приводится к виду

$$\frac{d^2}{\sigma_v^2}=\frac{2P_{\text{ср}}T}{N_0}\left[\int_{-W}^{W}\frac{|X_{rc}(f)|}{|C(f)|}df\right]^{-2}. \qquad (9.2.82)$$

Пример 9.2.1

Определим оптимальные фильтры на передаче и приёме для двоичной системы связи, которая передает данные со скоростью 4800 бит/с по каналу с частотной характеристикой

$$|C(f)|=\frac{1}{\sqrt{1+(f/W)^2}}, \quad |f|\le W, \qquad (9.2.83)$$

где $W=4800$ Гц. Аддитивный шум гауссовский, с нулевым средним и спектральной плотностью мощности $\frac{1}{2}N_0=10^{-15}$ Вт/Гц.

Поскольку $W=1/T=4800$, мы используем импульс сигнала со спектром приподнятого косинуса с $\beta=1$. Таким образом,

$$X_{rc}(f)=\tfrac{1}{2}T\left[1+\cos(\pi T|f|)\right]=T\cos^2\left(\frac{\pi|f|}{9600}\right). \qquad (9.2.84)$$

Затем

$$|G_T(f)|=|G_R(f)|=\left[1+\left(\frac{f}{4800}\right)^2\right]^{1/4}\cos\left(\frac{\pi|f|}{9600}\right), \quad |f|\le 4800 \qquad (9.2.85)$$

и $|G_T(f)|=|G_R(f)|=0$ в другой области частот. Рис. 9.2.13 даёт АЧХ фильтра $G_T(f)$.

Можно теперь использовать эти оптимальные фильтры для определения величины передаваемой энергии $\mathscr{E}$, требуемой для достижения заданной вероятности ошибки. Эта задача оставлена в качестве упражнения для читателя.

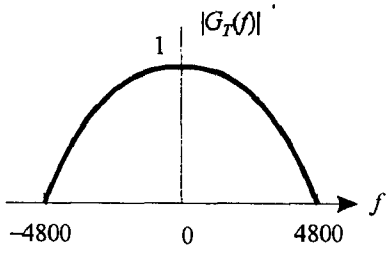


Рис. 9.2.13. Частотная характеристика оптимального фильтра передачи

9.3. ВЕРОЯТНОСТЬ ОШИБКИ ПРИ ДЕТЕКТИРОВАНИИ АМ

В этом разделе мы рассчитаем качество демодуляции и детектирования M-ичного сигнала АМ в присутствии АБГШ на входе приёмника. Сначала рассмотрим случай, когда фильтры передатчика и приёмника $G_R(f)$ и $G_T(f)$ синтезированы для нулевой МСИ. Затем мы рассмотрим случай, когда $G_R(f)$ и $G_T(f)$ синтезированы так, что $x(t) = g_T(t) \otimes g_R(t)$ является или дуобинарным или модифицированным дуобинарным двоичным сигналом.

9.3.1. Вероятность ошибки при детектировании АМ с нулевой МСИ

В отсутствии МСИ отсчеты принимаемого сигнала на выходе согласованного фильтра приёмника имеют вид

$$y_m = x_0 I_m + v_m, \tag{9.3.1}$$

где

$$x_0 = \int_{-W}^{W} |G_T(f)|^2 df = \mathscr{E}_g, \tag{9.3.2}$$

а v_m – аддитивный гауссовский шум с нулевым средним и дисперсией

$$\sigma_v^2 = \tfrac{1}{2}\mathscr{E}_g N_0. \tag{9.3.3}$$

В общем I_m принимает одно из M возможных равно удаленных значений амплитуды с равной вероятностью. При заданном частном значении уровня амплитуды задача сводится к определению вероятности ошибки.

Задача расчета вероятности ошибки для цифровой АМ в ограниченном по полосе канале с АБГШ в отсутствии МСИ идентична расчету вероятности ошибки M-ичных АМ, данному в разделе 5.2. Окончательный результат, который был получен, таков

$$P_M = \frac{2(M-1)}{M} Q\left(\sqrt{\frac{2\mathscr{E}_g}{N_0}}\right). \tag{9.3.4}$$

Но $\mathscr{E}_g = 3\mathscr{E}_{\text{ср}}/(M^2-1)$, а $\mathscr{E}_{\text{ср}} = k\mathscr{E}_{b\text{ср}}$ – средняя энергия на символ, а $\mathscr{E}_{b\text{ср}}$ – средняя энергия на бит.

Следовательно

$$P_M = \frac{2(M-1)}{M} Q\left(\sqrt{\frac{6(\log_2 M)\mathscr{E}_{b\text{ср}}}{(M^2-1)N_0}}\right). \tag{9.3.5}$$

Это в точности форма для вероятности ошибки в M-ичной АМ, полученная в разделе 5.2 (смотри (5.2.46)). В трактовке АМ настоящей главы мы навязываем дополнительное условие, что передаваемый сигнал ограничен полосой, пропускаемой каналом. Следовательно, передаваемые импульсы сигнала синтезированы как ограниченные по полосе и приводящие к нулевой МСИ.

В противоположность этому, ограничение по полосе не было навязано сигналам АМ, рассмотренным в разделе 5.2. Тем не менее, приёмники (демодуляторы и детекторы) в обоих случаях оптимальны (согласованные фильтры) для соответствующих переданных сигналов. Следовательно, нет потерь в качестве (вероятности ошибки), обусловленных ограничением полосы, если импульс сигнала синтезирован для нулевой МСИ, а канал не искажает передаваемый сигнал.

9.3.2. Вероятность ошибки при детектировании сигналов с парциальным откликом

В этом разделе мы определим вероятность ошибки при детектировании сигналов M-ичной АМ, используя дуобинарные и модифицированные дуобинарные импульсы. Канал считается идеальным, ограниченным по полосе, с АБГШ. Модель для системы связи показана на рис. 9.3.1.

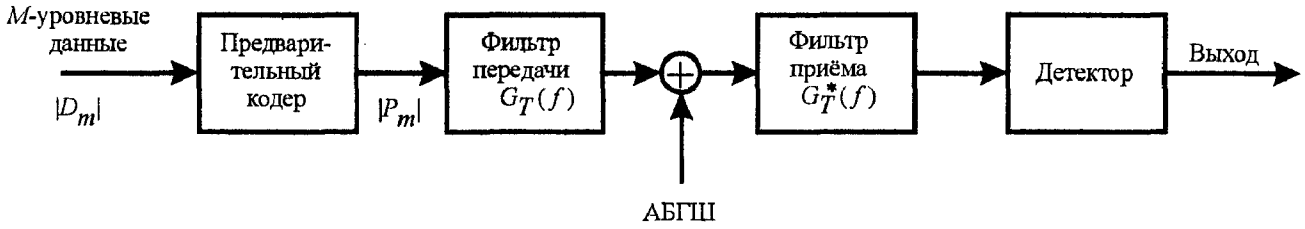


Рис. 9.3.1 Блок-схема модулятора и демодулятора для сигнала с парциальным откликом

Мы рассмотрим два типа детекторов. Первый – это посимвольный детектор, а второй – оптимальный МП детектор последовательности, описанный в предыдущем разделе.

Посимвольный детектор. На передаче M-уровневая последовательность данных $\{D_m\}$ подвергается предварительному кодированию, как описано ранее. Выход предварительного кодера отображается одним из M возможных уровней амплитуды. Затем на выходе фильтра передатчика с частотной характеристикой $G_T(f)$ появляется сигнал

$$\upsilon(t) = \sum_{n=-\infty}^{\infty} I_n g_T(t-nT). \qquad (9.3.6)$$

При парциальном отклике частотная функция $X(f)$ делится поровну между фильтрами передатчика и приёмника. Фильтр приёмника согласован с переданным импульсом, а каскадное объединение двух фильтров дает частотную характеристику

$$|G_T(f)G_R(f)| = |X(f)|. \qquad (9.3.7)$$

Выход согласованного фильтра стробируется в моменты $t = nT = n/2W$ и отсчеты питают декодер. Для дуобинарного сигнала выход согласованного фильтра в точках отсчёта можно выразить так

$$y_m = I_m + I_{m-1} + \nu_m = B_m + \nu_m, \qquad (9.3.8)$$

где ν_m – аддитивная компонента шума. Аналогично выход согласованного фильтра для модифицированного дуобинарного сигнала равен

$$y_m = I_m - I_{m-2} + \nu_m = B_m + \nu_m. \qquad (9.3.9)$$

При двоичной передачи положим $I_m = \pm d$, где $2d$ – расстояние между соседними уровнями сигнала. Тогда соответствующие значения B_m равны $(2d, 0, -2d)$. При передаче M-уровней АМ, когда $I_m = \pm d, \pm 3d, \ldots, \pm(M-1)d$, уровни принимаемого сигнала равны $B_m = 0, \pm 2d, \pm 4d, \ldots, \pm 2(M-1)d$. Таким образом число уровней принимаемых сигналов равно $2M-1$, а скалярный множитель d эквивалентен $x_0 = \mathscr{E}_g$.

Входные передаваемые символы $\{I_m\}$ считаются равновероятными. Тогда, для дуобинарной и модифицированной дуобинарного сигналов легко показать, что, в отсутствии шума, уровни принимаемых сигналов имеют (треугольное) распределение вероятностей в виде

$$P(B=2md)=\frac{M-|m|}{M^2},\quad m=0,\pm1,\pm2,\ldots,\pm(M-1), \tag{9.3.10}$$

где B означает свободные от шума принимаемые уровни, а $2d$ – расстояние между двумя соседними уровнями сигналами.

Канал искажает передаваемый сигнал путем добавления к нему гауссовского шума с нулевым средним и спектральной плотностью мощности $\frac{1}{2}N_0$.

Предположим, что ошибка символа возникает тогда, когда шумовая компонента на выходе демодулятора превышает некоторый порог d. Компонента шума ν_m является гауссовской с нулевым средним и с дисперсией

$$\sigma_\nu^2=\tfrac{1}{2}N_0\int_{-W}^{W}|G_R(f)|^2\,df=\tfrac{1}{2}N_0\int_{-W}^{W}|X(f)|\,df=2N_0/\pi \tag{9.3.11}$$

как для дуобинарного, так и для модифицированного дуобинарного сигналов. Таким образом, верхняя граница для вероятности ошибки символа определяется так:

$$\begin{aligned}P_M<&\sum_{m=-(M-2)}^{M-2}P(|y-2md|>d\,|B=2md)P(B=2md)+\\&+2P(y+2(M-1)d>d\,|\,B=-2(M-1)d)P(B=-2(M-1)d)=\\&=P(|y|>d|b=0)\left[2\sum_{m=0}^{M-1}P(B=2md)-P(B=0)-P(B=-2(M-1)d)\right]=\\&=\left(1-M^{-2}\right)P(|y|>d\,|B=0).\end{aligned} \tag{9.3.12}$$

Но

$$P(|y|>d\,|B=0)=\frac{2}{\sqrt{2\pi}\sigma_\nu}\int_d^\infty e^{-x^2/2\sigma_\nu^2}dx=2Q\left(\sqrt{\pi d^2/2N_0}\right). \tag{9.3.13}$$

Следовательно, средняя вероятность ошибки (вероятность ошибочного приёма) символа ограничена сверху так:

$$P_m<2\left(1-M^{-2}\right)Q\left(\sqrt{\pi d^2/2N_0}\right). \tag{9.3.14}$$

Скалярный множитель d в (9.3.14) может быть исключён путём его выражения через среднюю мощность сигнала, переданную по каналу. Для M-ичной АМ, уровни которой передаются равновероятно, средняя мощность на выходе фильтра передатчика равна

$$P_{ср}=\frac{E(I_m^2)}{T}\int_{-W}^{W}|G_T(f)|^2\,df=\frac{E(I_m^2)}{T}\int_{=W}^{W}|X(f)|\,df=\frac{4}{\pi T}E(I_m^2), \tag{9.3.15}$$

где $E(I_m^2)$ значение среднего квадрата M-уровневого сигнала, который равен

$$E(I_m^2)=\tfrac{1}{3}d^2(M^2-1). \tag{9.3.16}$$

Следовательно,

$$d^2=\frac{3\pi P_{ср}T}{4(M^2-1)}. \tag{9.3.17}$$

Подставив величину d^2 из (9.3.17) в (9.3.14), получим верхнюю границу для вероятности ошибки на символ

$$P_M<2\left(1-\frac{1}{M^2}\right)Q\left(\sqrt{\left(\frac{\pi}{4}\right)^2\frac{6}{M^2-1}\frac{\mathscr{E}_{ср}}{N_0}}\right), \tag{9.3.18}$$

где $\mathscr{E}_{ср}$ – средняя энергия на переданный символ, которую можно также выразить через среднюю энергию на бит (битовую энергию $\mathscr{E}_{bср}$), так: $\mathscr{E}_{ср}=k\mathscr{E}_{bср}=(\log_2 M)\mathscr{E}_{bср}$.

Выражение (9.3.18) для вероятности ошибки M -уровневой АМ справедлива как для дуобинарного, так и модифицированного дуобинарного сигналов с парциальным откликом. Если сравним этот результат с вероятностью ошибки M -уровневой АМ с нулевой МСИ, которую можно получить используя сигнальный импульс со спектром приподнятого косинуса, мы видим, что качество дуобинарного или модифицированного дуобинарного сигналов с парциальным откликом имеют энергетическую потерю $(\frac{1}{4}\pi)^2$ или 2,1 дБ. Эта потеря в ОСШ обусловлена тем, что детектор сигналов с парциальным откликом делает посимвольное решение, тем самым игнорируя внутреннюю память, которая имеется в принимаемом сигнале на входе детектора.

Последовательный детектор максимального правдоподобия. Последовательный МП детектор, который ищет по решётке наиболее вероятную переданную последовательность $\{I_m\}$, ранее описан в разделе 9.2.3. На каждом шаге процесса поиска детектор сравнивает метрики путей, которые сходятся у каждого узла и выбирает путь, который наиболее правдоподобен для каждого узла. Качество детектора можно рассчитать путём определения вероятности ошибочных событий, основываясь на метрики евклидового расстояния, как это было сделано при декодирования мягких решений свёрточных кодов. Общее исследование дано в разделе 10.1.14. Для случая дуобинарного и модифицированного дуобинарного сигналов оно показывает, что потеря в 2,1 дБ, присущая посимвольному детектору, полностью компенсируется последовательным МП детектором.

9.3.3. Вероятность ошибки для оптимальных сигналов в канале с искажением

В разделе 9.2.4 мы установили характеристики фильтров для модулятора и демодулятора, которые минимизируют ОСШ на входе детектора при искажениях в канале. Если фильтры синтезированы для нулевой МСИ в моментах отсчёта, вероятность ошибки для M -ичной АМ равна

$$P_M = \frac{2(M-1)}{M} Q\left(\sqrt{d^2/\sigma_v^2}\right). \tag{9.3.19}$$

Параметр d связан со средней переданной мощностью

$$P_{\text{ср}} = \frac{E\left[I_m^2\right]}{T}\int_{-W}^{W}|G_T(f)|^2 df = \frac{(M^2-1)d^2}{3T}\int_{-W}^{W}|G_T(f)|^2 df, \tag{9.3.20}$$

а дисперсия шума определена (9.2.69). Для АБГШ, (9.3.19) можно выразить так

$$P_M = \frac{2(M-1)}{M} Q\left(\sqrt{\frac{6\mathscr{E}_{\text{ср}}}{(M^2-1)N_0}\left[\int_{-W}^{W}\frac{|X_{rc}(f)|}{C(f)}df\right]^{-2}}\right). \tag{9.3.21}$$

Наконец, мы видим, что энергетическая потеря, обусловленная искажениями в канале, равна

$$20\lg\left[\int_{-W}^{W}\frac{|X_{rc}(f)|}{C(f)}df\right]. \tag{9.3.22}$$

Заметим, что когда $C(f)=1$ для $|f|\le W$ канал идеален и

$$\int_{-W}^{W} X_{rc}(f)df = 1, \tag{9.3.23}$$

так что в этом случае потерь нет, как следует из (9.3.22). Эти потери не зависят от фазовых искажений в канале, поскольку фазовые искажения точно компенсируются, что подразумевается в (9.2.80).

Потери, определяемые (9.3.22), обусловлены полностью искажениями амплитуд и являются мерой повышения шума, возникающего из-за фильтра, который призван компенсировать канальные искажения.

9.4. МОДУЛЯЦИОННЫЕ КОДЫ ДЛЯ ФОРМИРОВАНИЯ СПЕКТРА

Мы видели, что спектральную плотность мощности сигнала в цифровой системе связи можно контролировать (управлять) и формировать путём выбора формы импульса передаваемого сигнала $g(t)$ и введением корреляции посредством кодирования, которая используется для борьбы с искажениями в канале и шумом при передаче сигнала. Кодирование для формирования сигнала выполняется после канального кодирования, так что спектр передаваемого сигнала согласовывается со спектральными характеристиками базового или эквивалентного низкочастотного сигнала. Коды, которые используются для формирования спектра обычно называются или *модуляционными кодами*, или *линейными кодами*, или *кодами перевода данных*. Такие коды, в общем, вводят ограничение на последовательность символов, поступающих на модулятор, и таким образом, вводят корреляцию и, следовательно, память в передаваемый сигнал. Именно этот тип кодирования рассматривается в этом разделе.

Модуляционные коды обычно используются в магнитной записи, в оптической записи и цифровой связи по кабельным системам, для достижения формирования спектра и для ограничения или минимизации постоянной составляющей, имеющейся в передаваемом (или записываемом) базовом сигнале. В каналах магнитной записи модуляционный код синтезируется для увеличения расстояния между переходами в записанном сигнале и, таким образом, также уменьшается влияние МСИ.

Как пример использования модуляционного кода рассмотрим систему магнитной записи, которая состоит из элементов, показанных на блок-схеме рис. 9.4.1.

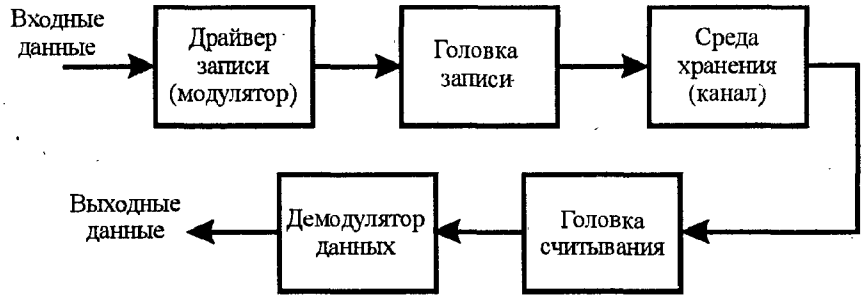


Рис 9.4.1. Блок-схема системы записи/считывания при магнитном хранении данных

Чтобы хранить двоичную последовательность данных она используется для создания (генерирования) записывающего тока. Этот ток можно рассматривать как выход «модулятора». Наиболее общий используемый метод для отображения информационной последовательности в сигнал записывающего тока – это ДБНП, который был описан в разделе 4.3.2. Напомним, что в ДБНП переход от одной амплитуды к другой (A в $-A$ или $-A$ в A) осуществляется только если информационный символ «1». Перехода нет, если информационный символ «0», т.е. в этом случае сохраняется уровень амплитуды предыдущего такта. Положительный импульс амплитуды приводит к намагничиванию среды в одной (прямой) полярности, а отрицательный импульс амплитуды намагничивает среду в противоположной полярности.

Поскольку входная последовательность данных в принципе случайная с равной вероятностью 1 и 0, мы можем встретить переходы от A до $-A$ и от $-A$ до A с

вероятностью 1/2 для каждого символа данных. Считывающий сигнал для положительного перехода (от $-A$ до A) является импульсом, который хорошо моделируется математически так

$$g(t)=\frac{1}{1+(2t/T_{50})^2}, \qquad (9.4.1)$$

где T_{50} определён как ширина импульса на уровне 50% его амплитуды, как показано на рис. 9.4.2. Аналогично, считывающий сигнал для отрицательного перехода (от A до $-A$) является импульсом $-g(t)$. Величина T_{50} определяется характеристиками среды, считывающей и записывающей головками и расстоянием от головки до среды.

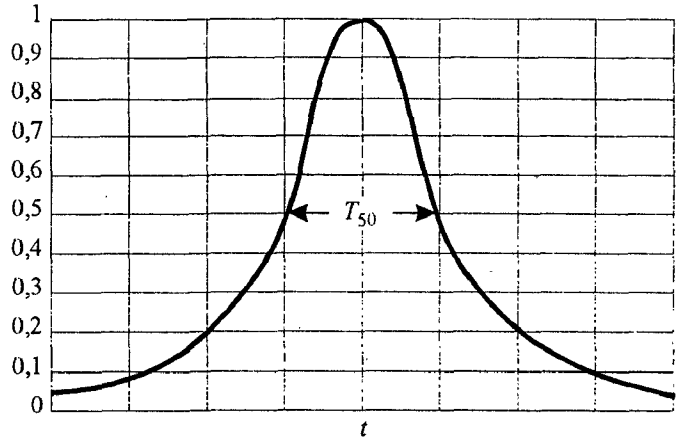


Рис. 9.4.2. Считывающий импульс в магнитной системе записи

Теперь предположим, мы записали положительный переход, следующий после отрицательного перехода. Пусть изменяется интервал времени между двумя переходами, который мы обозначим через T_b (битовый интервал). Рис. 9.4.3 иллюстрирует импульсы считывающего сигнала, полученного суперпозицией $p(t)$ и $-p(t-T_b)$.

Параметр $\Delta = T_{50}/T_b$ определён как *нормированная плотность*. Чем плотнее битовые переходы (T_b меньше), тем больше величина нормированной плотности и, следовательно, тем больше будет плотность упаковки информации на магнитном носителе. Заметим, что если Δ увеличивается, пиковые амплитуды сигнала считывания уменьшаются и сдвигаются во времени от желательных моментов времени.

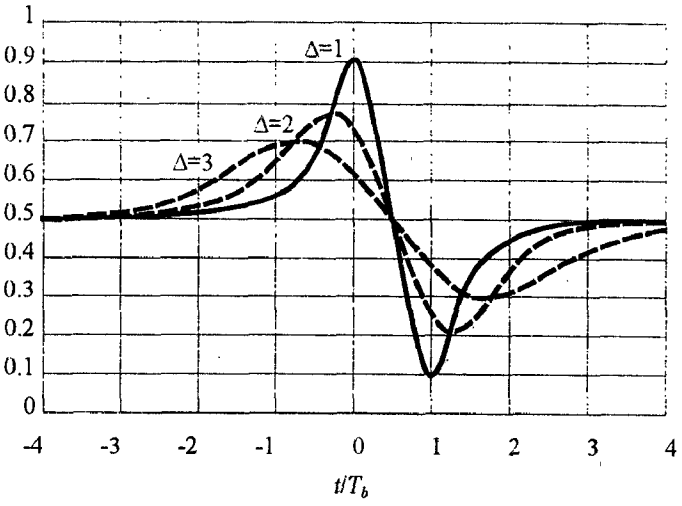


Рис. 9.4.3. Отклик считывающего устройства на смену знака перехода

Другими словами, импульсы интерферируют друг с другом, ограничивая таким образом плотность упаковки.

Эта проблема служит мотивом для синтеза модуляционных кодов, которые берут исходную последовательность данных и преобразуют (кодируют) в другую

последовательность, которая приводит к форме сигнала записи, в котором переходы амплитуд рассеиваются дальше друг от друга. Для примера, если мы используем ДБНП, закодированная последовательность на входе модулятора должна содержать один или больше нулей между единицами.

Вторая проблема, встречающаяся в магнитной записи – это необходимость подавления (или минимизации) постоянной составляющей, содержащейся в модулированном сигнале (записывающем токе) и обусловленной частотными характеристиками считывающей системы и связанной с ней электроники. Это требование также возникает в цифровой связи по кабельным каналам. Эту проблему можно преодолеть путём изменения (кодирования) последовательности данных на входе модулятора. Класс кодов, которые удовлетворяют этим целям (задачам)являются модуляционные коды, описываемые ниже.

Коды с ограниченным разбегом. Коды, которые имеют ограниченное число последовательных единиц или нулей в последовательности обычно называют *кодами с ограниченным разбегом*. Эти коды в общем описываются двумя параметрами, скажем d и κ, где d означает минимальное число нулей между двумя единицами в последовательности, а κ означает максимальное число нулей между единицами в последовательности. Применяя их совместно с ДБНП модуляцией, влияние расположения d нулей между последовательными единицами сводится к рассеянию переходов дальше друг от друга, тем самым сокращается перекрытие канальных откликов, вызванных последовательными переходами и таким образом сокращается МСИ. Установка верхней границы κ для пачки нулей обеспечивает достаточно частое появление переходов. так что информация символьной синхронизации может быть восстановлена из принимаемого модулированного сигнала. Коды с ограниченным разбегом обычно называют (d,κ) кодами[1]. Ограниченную кодовую последовательность (d,κ) кода можно представить конечным автоматом – последовательной системой с ограниченным числом состояний, именно $\kappa+1$ состояниями, обозначаемые как S_i, $1 \le i \le \kappa+1$, как показано на рис. 9.4.4.

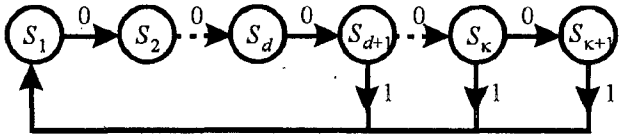


Рис. 9.4.4. Последовательный конечный автомат для (d,κ)-кодовой последовательности

Видно, что выходной символ 0 последовательность принимает переходя от состояния S_i до S_{i+1}, $i \le \kappa$. Выходной символ кодера может быть 1 только тогда, когда последовательность находится в состоянии S_i, $d+1 \le i \le \kappa+1$. Когда последовательность находится в состоянии $S_{\kappa+1}$, выходной символ всегда 1.

Конечный автомат можно также представить через *матрицу переходов состояний*, обозначаемую **D**, которая является квадратной $(\kappa+1)\times(\kappa+1)$ с элементами d_{ij}, где

$$d_{i1} = 1 \quad (i \ge d+1)$$
$$d_{ij} = \begin{cases} 1 & (j = i+1) \\ 0 & (\text{для других } i,\ j). \end{cases} \qquad (9.4.2)$$

[1] В действительности их обычно называют (d,k) кодами, где k – максимальная длина последовательных нулей. Мы использовали греческую букву κ вместо k, чтобы избежать путаницу с ранее использованным символом k.

Пример 9.4.1. Определим матрицу переходов состояний для $(d,\kappa)=(1,3)$ кода. Код (1,3) имеет четыре состояния. Из рис. 9.4.4 получаем его матрицу переходов состояний, которая равна

$$D=\begin{bmatrix}0&1&0&0\\1&0&1&0\\1&0&0&1\\1&0&0&0\end{bmatrix}. \qquad (9.4.3)$$

Важным параметром любого (d,κ) кода является число последовательностей определённой длины, скажем n, которые удовлетворяют ограничениям (d,κ) кода. Если n позволено расти, то число последовательностей $N(n)$, удовлетворяющих ограничениям (d,κ) кода, также растёт. Число информационных символов, которые можно однозначно представить посредством $N(n)$ кодовых последовательностей, равно

$$k=\lfloor \log_2 N(n) \rfloor,$$

где $\lfloor x \rfloor$ означает наибольшее целое, содержащееся в x. Тогда максимальная скорость кода $R_c = k/n$.

Пропускная способность (d,κ) кода определяется так

$$C(d,\kappa)=\lim_{n\to\infty}\frac{1}{n}\log_2 N(n). \qquad (9.4.4)$$

Ясно, что $C(d,k)$ – это максимально возможная скорость, которую можно достичь при ограничениях на (d,κ). Шеннон (1948) показал, что пропускная способность $C(d,k)$ определяется так:

$$C(d,k)=\log_2 \lambda_{\max}, \qquad (9.4.5)$$

где $\lambda_{\max}$ – наибольшее вещественное собственное число матрицы состояний переходов $\mathbf{D}$.

Пример 9.4.2. Определим пропускную способность $(d,\kappa)=(1,3)$ кода. Используя матрицу состояний переходов из примера (9.4.1) для (1,3) кода имеем

$$\det(\mathbf{D}-\lambda\mathbf{I})=\det\begin{bmatrix}-\lambda&1&0&0\\1&-\lambda&1&0\\1&0&-\lambda&1\\1&0&0&-\lambda\end{bmatrix}=\lambda^4-\lambda^2-\lambda-1=0. \qquad (9.4.6)$$

Максимальный вещественный корень этого полинома найден и равен $\lambda_{\max}=1{,}4656$. Следовательно, пропускная способность кода $C(1,3)=\log_2\lambda_{\max}=0{,}5515$.

Пропускная способность (d,κ) кодов для $0\le d\le 6$ и $2\le\kappa\le 15$ даны в табл. 9.4.1. Видим, что $C(d,\kappa)<\frac{1}{2}$ для $d\ge 3$ при любом κ. Наиболее часто используемые коды для магнитной записи имеют $d\le 2$, следовательно, их скорость R_c по крайней мере $\frac{1}{2}$.

Теперь обратим наше внимание на конструирование некоторых кодов с ограниченным разбегом. В общем (d,κ) коды можно конструировать или как коды фиксированной длины, или как коды переменной длины. В коде фиксированной длины каждый символ или блок из κ символов кодируется в блок из $n>\kappa$ символов.

В принципе конструкция кода с фиксированной длиной стандартна. Для данной длины блока n мы можем выбрать подмножество кодовых слов, которые удовлетворяют специфическим ограничениям для кодов с ограниченным разбегом. Из этого

подмножества мы исключаем кодовые слова, которые не удовлетворяют ограничениям при их каскадном соединении. Так мы получаем набор кодовых слов, которые удовлетворяют ограничениям, и могут использоваться в отображениях входных символов данных кодера. Операции кодирования и декодирования можно выполнить табличным методом.

Табл. 9.4.1. Пропускная способность $C(d,k)$ как функция от параметров d и к

k	d=0	d=1	d=2	d=3	d=4	d=5	d=6
2	.8791	.4057					
3	.9468	.5515	.2878				
4	.9752	.6174	.4057	.2232			
5	.9881	.6509	.4650	.3218	.1823		
6	.9942	.6690	.4979	.3746	.2269	.1542	
7	.9971	.6793	.5174	.4057	.3142	.2281	.1335
8	.9986	.6853	.5293	.4251	.3432	.2709	.1993
9	.9993	.6888	.5369	.4376	.3620	.2979	.2382
10	.9996	.6909	.5418	.4460	.3746	.3158	.2633
11	.9998	.6922	.5450	.4516	.3833	.3285	.2804
12	.9999	.6930	.5471	.4555	.3894	.3369	.2924
13	.9999	.6935	.5485	.4583	.3937	.3432	.3011
14	.9999	.6938	.5495	.4602	.3968	.3478	.3074
15	.9999	.6939	.5501	.4615	.3991	.3513	.3122
∞	1,000	.6942	.5515	.4650	.4057	.3620	.3282

Пример 9.4.3. Синтезируем код с $d=0$, $\kappa=2$ и длиной $n=3$ и определим его эффективность. Из списка кодовых слов мы находим пять кодовых слов, удовлетворяющих ограничениям (0, 2) кода: (0 1 0), (0 1 1), (1 0 1), (1 1 0), (1 1 1). Мы можем выбрать любые четыре из этих кодовых слов и использовать их для кодирования пар кодовых символов (0 0), (0 1), (1 0), (1 1). Так мы получаем код со скоростью $k/n=2/3$, который удовлетворяет условиям (0, 2) кода.

Код фиксированной длины в этом примере не очень эффективен. Его пропускная способность $C(0,\ 2)=0{,}8791$, так что этот код имеет *эффективность*

$$\text{эффективность}=\frac{R_c}{C(d,k)}=\frac{2/3}{0{,}8791}=0{,}76.$$

Конечно, лучшие (0, 2) коды можно сконструировать увеличением длины блока n.

В следующем примере, мы не накладываем ограничение на максимальное простирание нулей.

Пример 9.4.4. Сконструируем код длиной $n=5$ с параметрами $d=1$, $\kappa=\infty$. В этом случае, мы не накладываем ограничения на число последовательных нулей. Для конструирования кода мы выберем из 32 возможных кодовых слов те, которые удовлетворяют ограничению $d=1$. Имеется восемь таких кодовых слов, что подразумевает, что мы можем кодировать три информационных символа с каждым кодовым словом. Код дан в таблице 9.4.2. Заметим, что первый символа каждого кодового слова есть 0, в то время как последний символ может быть или 0, или 1. Следовательно, когда эти кодовые слова каскадно соединяются, ограничение $d=1$ удовлетворяется. Этот код имеет скорость $R_c=3/5$. Если её сопоставить с пропускной способностью

$C(1, \infty) = 0{,}6942$, полученной из таблицы 9.4.1, то эффективность кода равна 0,864, что вполне приёмлемо.

Таблица 9.4.2. Код с фиксированной длиной, $d = 1$, $\kappa = \infty$

Входные символы данных	Выходные кодированные последовательности
0 0 0	0 0 0 0 0
0 0 1	0 0 0 0 1
0 1 0	0 0 0 1 0
0 1 1	0 0 1 0 0
1 0 0	0 0 1 0 1
1 0 1	0 1 0 0 0
1 1 0	0 1 0 0 1
1 1 1	0 1 0 1 0

Метод конструирования кодов, описанный выше в двух примерах, приводит к кодам (d, κ) фиксированной длины, которые *независимы по состояниям*. Под независимостью состояний мы имеем в виду, что кодовые слова кода фиксированной длины могут быть каскадно объединены без нарушения ограничений на (d, κ) коды. В общем, когда d велико, коды (d, κ) фиксированной длины с независимостью по состояниям требуют большую длину блока символов. Возможны более простые (короткой длины) коды, в которых допустима зависимость состояний и переменная длина кодовых слов. Ниже мы рассмотрим коды, для которых как длина входных блоков на кодер, так и выходные блоки могут иметь переменную длину. Для однозначного декодирования кодовых слов в приёмнике коды переменной длины должны удовлетворять префиксному условию, описанному в главе 3.

Пример 9.4.5. Очень простым уникальным декодируемым кодом переменной длины с параметрами $d = 0, \kappa = 2$ является код

$$0 \rightarrow 01$$
$$10 \rightarrow 10$$
$$11 \rightarrow 11.$$

Код в этом примере имеет фиксированный размер выходного блока, но переменный размер входного блока. В общем, входные и выходные блоки могут иметь переменные размеры. Следующий пример иллюстрирует последний случай.

Пример 9.4.6. Сконструируем код (2, 7) с переменными размерами блока. Это несложный код. Мы выбрали этот пример потому, что код (2, 7) широко известен в технике ПЭВМ для многих дисковых систем хранения информации. Код иллюстрируется в таблице 9.4.3. Мы видим, что входные блоки данных из 2, 3, и 4 символов отображаются в выходные блоки данных из 4,6 и 8 символов, соответственно. Следовательно скорость кода $R_c = 1/2$. Поскольку это скорость кода для всех кодовых слов, код называется кодом *с фиксированной скоростью*. Это код имеет эффективность $0{,}5/0{,}5174 = 0{,}966$. Заметим, что этот код удовлетворяет префиксному условию.

Другой код, который широко используется в магнитной записи – это код $(d, \kappa) = (1, 3)$ со скоростью $1/2$, иллюстрированный табл. 9.4.4. Видим, что когда информационный входной символ 0, то первый выходной символ 1, если предшествующий входной символ был 0, или первый выходной символ 0, если предшествующий входной символ был 1, а второй выходной символ 0.

Таблица 9.4.3. Кодовая таблица для кода переменной длины (2,7).

Входные символы данных	Кодированные выходные последовательности
1 0	1 0 0 0
1 1	0 1 0 0
0 1 1	0 0 0 1 0 0
0 1 0	0 0 1 0 0 0
0 0 0	1 0 0 1 0 0
0 0 1 1	0 0 1 0 0 1 0 0
0 0 1 0	0 0 0 0 1 0 0 0

Таблица 9.4.4. Кодер для кода Миллера (1,3)

Входные символы данных	Кодированные выходные последовательности
0	x 0
1	0 1

x=0, если предыдущий входной символ 1,
x=1, если предыдущий входной символ 0.

Если входной информационный символ 1, то выход кодера 01. Декодирование этого кода простое. Первый символ двухсимвольного блока избыточный и может быть отброшен. Второй символ – это информационный символ. Этот код обычно называется *кодом Миллера*. Видим, что это код с зависимыми состояниями, которые описываются диаграммой состояний, показанный на рисунке 9.4.5. Имеются два состояния, обозначаемые S_1 и S_2, с переходами, указанными на рисунке. Когда кодер в состоянии S_1 то входной символ 1 приводит к тому, что кодер остается в состоянии S_1, а выход кодера 01. Это обозначено 1/01. Если входной символ 0, кодер переходит в состояние S_2, а выход кодера 00. Это обозначено 0/00. Аналогично, если кодер в состоянии S_2, то входной символ 0 не вызывает переход, а выход кодера 10. С другой стороны, если входной символ 1, кодер переходит в состояние S_1, а выход кодера 01. Рис. 9.4.6 показывает решётку для кода Миллера.

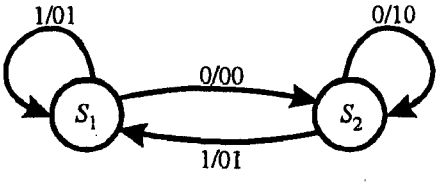


Рис. 9.4.5. Диаграмма состояний для кода Миллера (d=1, к=3)

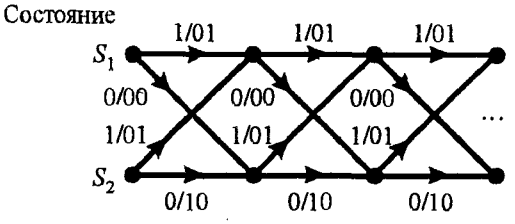


Рис. 9.4.6. Решётка для кода Миллера (d=1, к=3)

Отображение кодовых символов в сигналы. Выходная последовательность $(d, к)$ кодера отображается модулятором в сигнал для передачи по каналу. Если двоичный символ 1 отображается прямоугольным импульсом амплитуды A, а двоичный символ 0 отображается прямоугольным импульсом с амплитудой $-A$ результатом является $(d, к)$ кодированный сигнал ДБН (NRZ). Заметим, что длительность прямоугольных кодовых импульсов равна $T_c = R_c/R_b = R_c T_b$ где R_b – информационная скорость на входе кодера, T_b – битовый интервал (не кодированных символов), а R_c – скорость кода для $(d, к)$ кода.

Когда $(d,\text{к})$ код с фиксированной длинной и независимыми состояниями имеет скорость $R_c = k/n$, мы можем рассматривать образование каждого n-символьного блока, как генерирование одного сигнала длительностью nT_c. Таким образом мы имеем M=$2^{\text{к}}$ сигналов, один для каждого из 2^k возможных n символьных блоков. Эти кодированные сигналы имеют общую форму, определённую (4.3.6) и (4.3.38). В этом случае нет зависимости между передаваемыми последовательными сигналами.

В противоположность ситуации, рассмотренной выше, модулированный сигнал имеет память, если используют сигнал ДБНП и (или) код $(d,\text{к})$ имеет зависимые состояния. Рассмотрим влияние отображения кодовых символов в ДБНП сигнал.

Напомни, что зависимость состояний ДБНП сигнала обусловлена дифференциальным кодированием информационной последовательности. Дифференциальное кодирование является формой предварительного кодирования, которое описывается математически так

$$p_k = d_k \oplus p_{k-1},$$

где $\{d_k\}$ является двоичной последовательностью на входе предварительного кодера предкодера, $\{p_k\}$ – выходная последовательность символов предкодера, а $\oplus$ означает суммирование по mod 2. Это кодирование характеризуется диаграммой состояний, показанной на рис. 9.4.7(a).

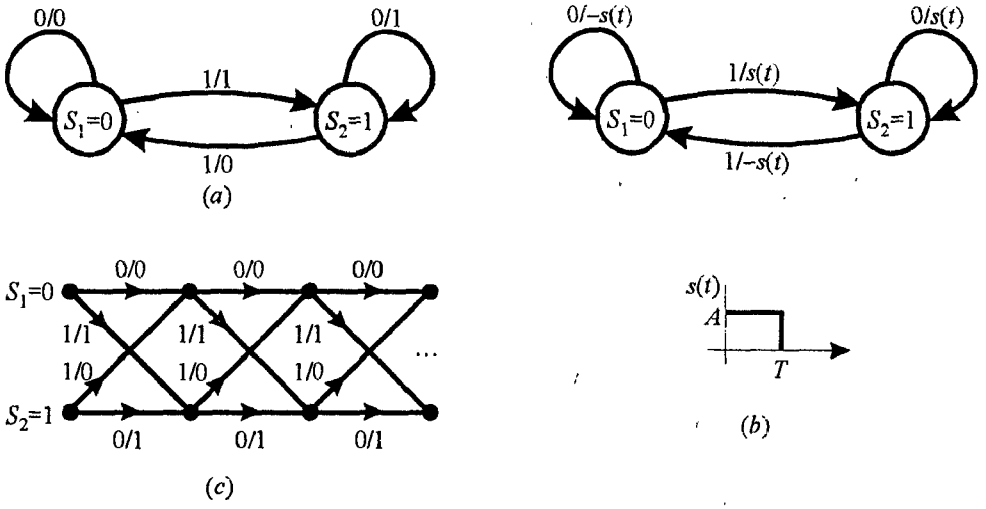


Рис. 9.4.7. Диаграмма состояний и решётка состояний для сигнала ДБНП (NRZI)

Последовательность $\{p_k\}$ передаётся посредством сигнала ДБН. Таким образом, когда $p_k = 1$, то выход модулятора является прямоугольным импульсом амплитуды A, а когда $p_k = 0$ – выходом модулятора является прямоугольный импульс амплитуды $-A$. Когда кодирование отображается диаграммой состояний рис. 9.4.7(a) мы получаем соответствующую диаграмму состояний, показанную на рис. 9.4.7(b). Соответствующая решётка показана на рис. 9.4.7(c).

Когда выход кодера $(d,\text{к})$ с зависимыми состояниями подаётся на ДБНП модулятор, мы можем просто сочетать две диаграммы состояний в единую диаграмму состояний для $(d,\text{к})$ кода с предварительным кодированием. Аналогичное сочетание можно выполнить с соответствующими решётками. Следующий пример иллюстрирует подход для кода Миллера (1, 3) обслуживающий ДБНП модулятор.

Пример 9.4.7. Определим диаграмму состояний комбинации кода Миллера (1, 3), за которым следует предварительное кодирование, свойственное ДБНП модуляции. Поскольку код Миллера (1, 3) имеет два состояния и предкодер имеет два состояния, диаграмма состояний комбинированного кодера имеет четыре состояния, которые мы обозначим так $(S_M, S_N) = (\sigma_1, s_1)$, (σ_1, s_2), (σ_2, s_1), (σ_2, s_2), где $S_M = \{\sigma_1, \sigma_2\}$ представляют два состояния кода Миллера, а $S_N = \{s_1, s_2\}$ представляют два состояния предкодера для ДБНП. Для каждого символа входных данных кодера Миллера мы получаем два выходных символа, которые затем идут на предкодер, превращаясь в два выходных символа предкодера. Результирующая диаграмма состояний показана на рис. 9.4.8, где первый символ означает информационный символ на входе кодера Миллера, а следующие два символа представляют соответствующий выход предкодера. Решётчатую диаграмму состояний для подвергающейся предварительному кодированию последовательности Миллера можно получить непосредственно из комбинированной диаграммы состояний или из комбинаций решёток двух кодов. Результатом такого комбинирования является решетка с четырьмя состояниями, одна ступень которой показана на рис. 9.4.9.

Читатель в порядке упражнения может попытаться показать, что четыре сигнала, полученные отображением каждой пары символов предварительно кодированной последовательности Миллера в ДБН сигнал, биортогональны и что результирующий модулированный сигнал идентичен сигналу, полученному модуляцией с задержкой, описанной в разделе 4.3.2.

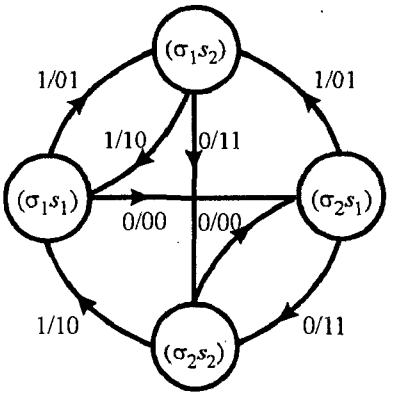


Рис. 9.4.8. Диаграмма состояний каскадного соединения кода Миллера и ДБНП

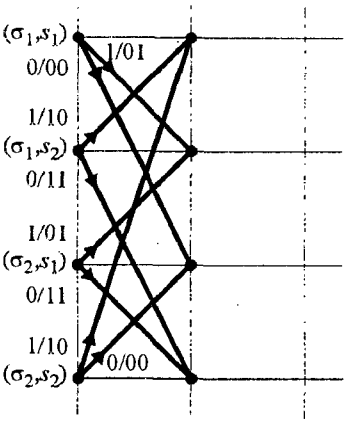


Рис. 9.4.9. Один шаг решётчатой диаграммы каскадного соединения кода Миллера и ДБНП

Из диаграмм состояний с зависимыми состояниями можно получить матрицу вероятностей переходов, как описано в разделе 4.3.2. Затем можно определить спектральную плотность мощности кодированных сигналов, как показано в разделе 4.4.3.

9.5. БИБЛИОГРАФИЧЕСКИЕ ЗАМЕЧАНИЯ И ССЫЛКИ

Основополагающая работа по синтезу сигналов для ограниченных по полосе каналов выполнена Найквистом (1928). Использование двоичных сигналов с парциальным откликом было первоначально предложено Лендером (1963) и было позднее обобщено Кретчмером (1966). Другие ранние работы по проблемам, имеющим отношение к межсимвольной интерференции (МСИ) и оптимизации передатчика и приёмника для ограничений МСИ даны Герстом и Даймондом (1961), Тафтсом (1965) Смитом (1965) и Бергером и Тафтсом (1967). Передача «быстрее Найквиста» изучалась Мазо (1975) и Фошини (1984).

Модуляционные коды были также впервые введены Шенноном (1948). Некоторые ранние работы по конструированию кодов с ограниченным разбегом можно найти в статьях Фреймана и Вайнера (1964), Габора (1967), Франашека (1968, 1969, 1970), Танга и Баля (1970) и Якоби (1977). Более свежие работы можно найти в статьях Адлера Коппершмитта и Хаснера (1983) и Карабеда и Зигеля (1991). Мотивация большинства работ по кодам с ограниченным простиранием по длине было обеспечено их приложением к магнитной и оптической записи. Хорошо написанные учебные статьи по кодам с ограниченным простиранием по длине опубликованы Иминком (1990).

ЗАДАЧИ

9.1. Говорят, что канал является неискажающим, если отклик канала $y(t)$ на входное воздействие $x(t)$ равен $Kx(t-t_0)$, где K и t_0 константы. Покажите, что если частотная характеристика канала равна $A(f)e^{j\theta(f)}$, где $A(f)$ и $\theta(f)$ вещественные, достаточными и необходимыми условиями для неискажающей передачи являются условия $A(f)=K$ и $\theta(f)=-2\pi f t_0 + 2\pi n$, $n=0, 1, \ldots$

9.2. Спектральная характеристика приподнятого косинуса даётся (9.2.26).

а) Покажите, что соответствующая импульсная характеристика равна

$$x(t)=\frac{\sin(\pi t/T)}{\pi t/T}\,\frac{\cos(\beta\pi t/T)}{1-4\beta^2t^2/T^2}.$$

b) Найдите преобразование Гильберта $\hat{x}(t)$ от $x(t)$, когда $\beta=1$.

c) Обладает ли $\hat{x}(t)$ желательными свойствами сигнала $x(t)$, которые делают его подходящим для передачи данных? Объясните.

d) Определите огибающую ОБП сигнала с подавленной несущей, если модулирующим сигналом является $x(t)$.

9.3 а) Покажите что (формула суммирования Пуассона)

$$x(t)=\sum_{k=-\infty}^{\infty} g(t)h(t-kT) \Rightarrow X(f)=\frac{1}{T}\sum_{n=-\infty}^{\infty} H\left(\frac{n}{T}\right)G\left(f-\frac{n}{T}\right).$$

Подсказка: Используйте разложение в ряд Фурье периодического множителя

$$\sum_{k=-\infty}^{\infty} h(t-kT).$$

b) Используя результат а), подтвердите следующие версии суммы Пуассона:

$$\sum_{k=-\infty}^{\infty} h(kT)=\frac{1}{T}\sum_{n=-\infty}^{\infty} H\left(\frac{n}{T}\right) \qquad \text{(i)}$$

$$\sum_{k=-\infty}^{\infty} h(t-kT)=\frac{1}{T}\sum_{n=-\infty}^{\infty} H\left(\frac{n}{T}\right)\exp\left(\frac{j2\pi nt}{T}\right) \qquad \text{(ii)}$$

$$\sum_{k=-\infty}^{\infty} h(kT)\exp(-j2\pi kTf)=\frac{1}{T}\sum_{n=-\infty}^{\infty} H\left(f-\frac{n}{T}\right) \qquad \text{(iii)}$$

c) Получите условия отсутствия МСИ (критерий Найквиста), используя формулу суммирования Пуассона.

9.4. Предположим, что цифровая система связи использует гауссовскую форму огибающей сигнала вида

$$x(t)=\exp(-\pi a^2t^2).$$

Чтобы снизить уровень МСИ до относительно малой величины, мы навяжем условие, что $x(T)=0{,}01$, где T – символьный интервал. Полоса W для импульса $x(t)$ определяется как величина, при которой $X(W)/X(0)=0{,}01$, где $X(f)$ – преобразование Фурье от $x(t)$. Определите величину W и сравните эту величину с полосой сигнала, принятого со спектром приподнятого косинуса при $\beta=1$.

9.5. Ограниченный по спектру сигнал с полосой W можно представить в виде

$$x(t)=\sum_{n=-\infty}^{\infty} x_n\,\frac{\sin[2\pi W(t-n/2W)]}{2\pi W(t-n/2W)}.$$

а) Определите спектр $X(f)$ и постройте $|X(f)|$ для следующих случаев:

$$x_0 = 2,\ x_1 = 1,\ x_2 = -1,\ x_n = 0,\ n \neq 0,\ 1,\ 2, \qquad \text{(i)}$$

$$x_{-1} = -1,\ x_0 = 2,\ x_1 = -1,\ x_n = 0,\ n \neq -1,\ 0,\ 1. \qquad \text{(ii)}$$

b) Нарисуйте $x(t)$ для этих двух случаев.

c) Если эти сигналы используются для передачи двоичных сообщений в канале с АБГШ, определите число возможных принимаемых уровней в моментах стробирования $t = nT = n/2W$ и вероятность того, что принятые уровни правильные. Предположите, что двоичные символы на передаче равновероятны.

9.6. Полосовой канал с полосой 4 кГц используется для передачи данных со скоростью 9600 бит/с. Если спектральная плотность мощности аддитивного гауссовского шума с нулевым средним равна $\frac{1}{2}N_0 = 10^{-10}$ Вт/Гц, синтезируйте сигнал КАМ и определите среднюю мощность, при которой достигается вероятность ошибки на бит 10^{-6}. Используйте импульс сигнала со спектром приподнятого косинуса, если коэффициент ската по крайней мере 50% ($\beta = 0{,}5$).

9.7. Определите битовую скорость, которую можно передать через речевой телефонный канал с полосой 4 кГц, если используются следующие методы модуляции:

(a) двоичная АМ; (b) четырехфазная ФМ; (c) восьмиточечная КАМ; (d) двоичная ортогональная ЧМ с некогерентным детектированием; (e) ортогональная четырехпозиционная ЧМ с некогерентным детектированием, (f) ортогональная 8-позиционная ЧМ с некогерентным детектированнем. Для (a) – (c) предположите, что огибающая передаваемого импульса имеет спектр приподнятого косинуса с 50 % коэффициентом ската.

9.8. Идеальный речевой полосовой телефонный линейный канал имеет частотную характеристику, охватывающую частотную область 600–3000 Гц.

а) Синтезируйте $M = 4$ ФМ систему (квадратурную или КФМ) для передачи данных со скоростью 2400 бит/с при несущей $f_c = 1800$ Гц. Из спектральных соображений используйте частотную характеристику приподнятого косинуса. Нарисуйте блок-схему системы и опишите функциональные операции каждого блока.

b) Повторите (a) для битовой скорости $R = 4800$ бит/с.

9.9. Телефонный канал для передачи речи пропускает частоты в области от 300 до 3300 Гц. Желательно синтезировать модем, который передает 2400 символов/с с целью достижения битовой скорости 9600 бит/с. Выберите подходящее сигнальное созвездие КАМ, частоту несущей и коэффициент ската импульса со спектром приподнятого косинуса, который использует всю полосу частот. Нарисуйте спектр передаваемого сигнального импульса и укажите основной диапазон частот.

9.10. Система связи для канала с речевой полосой (3 кГц) синтезируется для случая, когда ОСШ на приёме у детектора равно 30 дБ в то время как мощность передатчика $P_s = -3$ дБВт. Определите величину P_s, если желательно расширить полосу системы до 10 кГц при сохранении того же ОСШ у детектора.

9.11. Покажите, что импульс, имеющий спектр приподнятого косинуса, определяемый (9.2.26), удовлетворяет критерию Найквиста (9.2.13) при любой величине коэффициента ската β.

9.12. Покажите, что при любой величине β, спектр приподнятого косинуса, определенный (9.2.26), удовлетворяет условию

$$\int_{-\infty}^{\infty} X_{rc}(f)\,df = 1.$$

Подсказка: используйте факт, что $X_{rc}(f)$ удовлетворяет критерию Найквиста (9.2.13).

9.13. Критерий Найквиста дает необходимые и достаточные условия для спектра $X(f)$ импульса $x(t)$, который обеспечивает нулевое МСИ. Докажите, что для любого импульса, ограниченного по полосе $|f| < 1/T$, условие нулевой МСИ удовлетворяется, если $\mathrm{Re}[X(f)]$ для $f > 0$ состоит из прямоугольной функции плюс произвольной дополнительной функции около $f = 1/2T$, а $\mathrm{Im}[X(f)]$ – произвольная чётная функция около $f = 1/2T$.

9.14. Телефонный канал с речевой полосой пропускает частоты в полосе 300 Гц $< f <$ 3000 Гц.

а) Выберите скорость передачи символов и эффективный по мощности размер созвездия для достижения информационной скорости передачи 9600 бит/с.

b) Если используется для передаваемого импульса спектр, равный квадратному корню из спектра приподнятого косинуса, выберите коэффициент ската. Предположите, что канал имеет идеальную частотную характеристику.

9.15. Синтезируйте M -ичную систему АМ, которая передает цифровую информацию по идеальному каналу с полосой $W = 2400$ Гц. Битовая скорость равна 14400 бит/с. Определите число передаваемых сигнальных точек, число принимаемых сигнальных точек, используя дуобинарный сигнальный импульс, а также требуемую энергию $\mathscr{E}_b$ для достижения вероятности ошибки 10^{-6}. Аддитивный шум гауссовский, с нулевым средним и спектральной плотностью мощности 10^{-4} Вт/Гц.

9.16. Двоичный АМ сигнал генерируется путем возбуждения фильтра с частотной характеристикой приподнятого косинуса с коэффициентом ската 50 %, а затем образуется двухполосный АМ сигнал без несущей, как показано на рис. Р9.16. Битовая скорость 2400 бит/с.

a) Определите спектр модулированного двоичного АМ сигнала и нарисуйте его.

b) Начертите блок-схему оптимального демодулятора-детектора для принимаемого сигнала, который равен передаваемому сигналу плюс АБГШ.

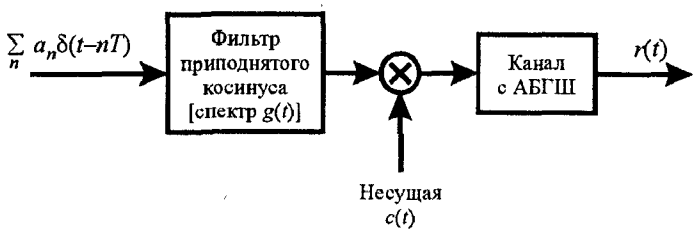


Рис. Р9.16

9.17. Элементы последовательности $\{a_n\}_{n=-\infty}^{\infty}$ являются независимыми двоичными, случайными величинами, принимающие значения ± 1 с равной вероятностью. Эта последовательность данных используется для модуляции базового импульса $g(t)$, показанного на рис. Р9.17(a). Модулированный сигнал

$$X(t) = \sum_{n=-\infty}^{\infty} a_n g(t - nT).$$

a) Найдите спектральную плотность $X(t)$.

b) Если вместо $g(t)$ используется импульс $g_1(t)$ (см. рис. 9.17 b), как меняется спектр мощности из (a)?

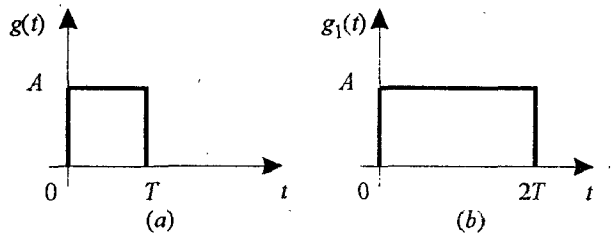


Рис. Р9.17

c) Предположим, что в (b) мы желаем иметь нуль в спектре на частоте $f = 1/3T$. Это делается посредством предкодера в виде $b_n = a_n + \alpha a_{n-3}$. Найдите α, которое обеспечит желательный нуль.

d) Возможно ли использовать предварительное кодирование в виде $b_n = a_n + \sum_{i=1}^{N} \alpha_i a_{n-i}$ для некоторых ограниченных N так, что окончательный спектр мощности равен нулю для $1/3T \le |f| \le 1/2T$? Если да, то как? Если нет, почему? (Подсказка: используйте свойства аналитических функций).

9.18. Рассмотрите передачу данных посредством АМ по речевому телефонному каналу, который имеет полосу 3000 Гц. Покажите, как меняется скорость передачи символов, как функция от излишка полосы. В частности, определите скорость передачи символов при излишке полосы на 25 %, 33 %, 50 %, 67 %, 75 % и 100 %.

9.19. Двоичная последовательность 10010110010 является входом предкодера, выход которого используется для модуляции дуобинарного передающего фильтра. Постройте таблицу, наподобие таблицы 9.2.1, показывающую последовательность на выходе предкодера, уровни передаваемых амплитуд, уровни принимаемых сигналов и декодированную последовательность.

9.20. Повторите задачу 9.19 для модифицированного дуобинарного сигнального импульса.

9.21. Предкодер для сигнала с парциальным откликом не в состоянии работать, если желательный

парциальный отклик при $n=0$ равен нулю по модулю M. Например, рассмотрите желательный отклик для $M=2$:

$$x(nT)=\begin{cases} 2 & (n=0),\\ 1 & (n=1),\\ -1 & (n=2),\\ 0 & (\text{для других } n),\end{cases}$$

Покажите, почему такой отклик нельзя подвергнуть предварительному кодированию.

9.22. Рассмотрите RC фильтр низких частот, показанный на рис. Р9.22, где $\tau = RC = 10^{-6}$.

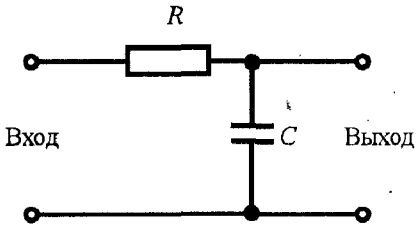


Рис. Р9.22

a) Определите и нарисуйте задержку огибающей (групповую задержку) фильтра, как функцию от частоты.

b) Предположите, что вход фильтра – низкочастотный сигнал с полосой $\Delta f = 1$ кГц. Определите отклик RC фильтра на этот сигнал.

9.23. Микроволновый радиоканал имеет частотную характеристику

$$C(f)=1+0{,}3\cos 2\pi fT.$$

Определите частотные характеристики оптимального передающего и приёмного фильтров, которые дают нулевое МСИ при скорости $1/T$ символов/с при наличии 50 % избыточной полосы. Предположите, что спектр аддитивного шума равномерный.

9.24. Четырехпозиционная АМ используется для передачи с битовой скоростью 9600 бит/с по каналу, имеющему частотную характеристику

$$C(f)=\frac{1}{1+j(f/2400)}$$

для $|f|\le 2400$ и $C(f)=0$ вне этой области частот. Аддитивный шум является гауссовским, белым с нулевым средним и спектральной плотностью мощности $\frac{1}{2}N_0$ Вт/Гц. Определите (амплитудные) частотные характеристики оптимального фильтра на передаче и приёме.

9.25. Определите пропускную способность кода $(0, 1)$ с ограниченным разбегом. Сравните эту пропускную способность с той же характеристикой для кода $(1, \infty)$ и объясните соотношения.

9.26. Троичный формат сигнала синтезируется для канала, который не пропускает постоянную составляющую. Входная двоичная информационная последовательность передается путём отображения 1 или в положительный, или отрицательный импульс, а 0 передается отсутствием импульса. Следовательно, для передачи 1 полярность импульса чередуется. Это называют AMI (alternate mark inversion) кодом. Определите пропускную способность кода.

9.27. Дайте другое описание AMI кода, описанного в задаче 9.26, используя набегающую сумму цифр (RDS) с ограничением, что RDS может принимать только значения 0 и +1.

9.28. (Коды kBnT, B – binary, T – ternary). В задаче 9.26 обратите внимание, что AMI является «псевдотроичным» кодом, в котором для передачи одного бита информации используется троичный символ, имеющий информационную ёмкость $\log_2 3 = 1{,}58$ бит. Этот код не обеспечивает необходимую форму спектра. Лучшую форму спектра даёт код, обозначаемый как kBnT, где k означает количество информационных бит, а n – число троичных символов на блок. При выборе наибольшего возможного k для каждого n получается следующая таблица:

k	n	код
1	1	1B1T
3	2	3B2T
4	3	4B3T
6	4	6B4T

Определите эффективность этих кодов путём вычисления отношения $k/(n\log_2 3)$. Заметьте, что 1B1T – это код AMI.

9.29. Эта задача имеет дело с пропускной способностью двух (d,κ) кодов.

a) Определите пропускную способность (d,κ) кода, который имеет следующую матрицу переходов состояний

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

b) Повторите (a) для

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

c) Прокомментируйте разницу (a) и (b).

9.30. Простейшая модель телеграфного кода состоит из двух символов (Блэйхут, 1990). Точка содержит элемент сигнала «1» и «0». Тире состоит из трёх следующих без перерыва элементов «1» и одного элемента «0».

Рассмотрите этот код как код, составленный из символов равной длительности.

Определите матрицу переходов состояний.

Определите информационную ёмкость кода.

9.31. Определите матрицу переходов состояний кода с ограниченным разбегом, который описывается диаграммой состояний, показанной на рис. P9.31. Нарисуйте соответствующую решетку.

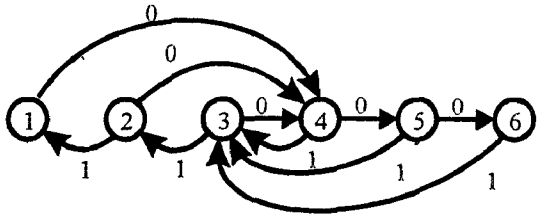


Рис. P9.31

9.32. Определите матрицу переходов состояний для кода (2,7) с ограниченным разбегом, удовлетворяющего диаграмме состояний, показанной на рис. P9.32.

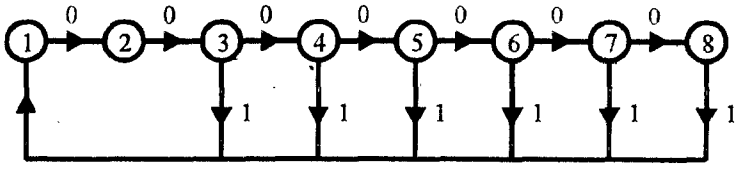


Рис. P9.32

10

СВЯЗЬ В ОГРАНИЧЕННЫХ ПО ПОЛОСЕ ЛИНЕЙНЫХ ФИЛЬТРОВЫХ КАНАЛАХ

В гл. 9 мы сосредоточили внимание на синтезе фильтров модулятора и демодулятора для ограниченных по полосе каналов. Процедура синтеза основывалась на предположении, что частотная характеристика (идеального или неидеального) канала $C(f)$ известна априори. Однако в практике цифровых систем связи, рассчитанных на высокоскоростную передачу через ограниченные по полосе каналам, частотная характеристика $C(f)$ не известна с достаточной точностью для возможности синтеза оптимальных фильтров модулятора и демодулятора. Например, при цифровой связи по коммутируемым телефонным сетям канал связи различен каждый раз при наборе нового номера, поскольку маршрут по каналу каждый раз различен. Это пример канала, характеристики которого не известны априори. Имеются другие типы каналов, например беспроводные, такие как радиоканалы и подводные акустические каналы, чьи частотные характеристики переменны во времени. Для таких каналов нет возможности синтезировать оптимальные фиксированные фильтры для демодулятора.

В этой главе мы рассмотрим проблему синтеза приёмника в условиях искажений в канале, не известных априори, и наличия АБГШ. Искажения в канале приводят к МСИ, которая, оставаясь некомпенсированной, часто вызывает ошибки. Решение проблемы МСИ можно свести к синтезу приёмника, который использует способ компенсации или сокращения МСИ в принимаемом сигнале. Компенсатор МСИ назван *эквалайзером или выравнивателем*.

Здесь рассматриваются три вида методов выравнивания. Один из них основывается на правиле максимально-правдоподобного (МП) детектирования последовательностей, и является оптимальным с точки зрения вероятности ошибки. Второй метод выравнивания основывается на использовании линейного фильтра с регулируемыми коэффициентами. Третий метод выравнивания, который использует символы, продетектированные раньше, для подавления МСИ при детектировании текущего символа назван выравниванием с обратной связью по решению. Мы начнём с рассмотрения оптимального детектора для каналов с МСИ.

10.1. ОПТИМАЛЬНЫЙ ПРИЕМНИК ДЛЯ КАНАЛА С МСИ И АБГШ

В этом разделе мы рассмотрим структуру оптимального демодулятора и детектора для цифровой передачи по неидеальному, ограниченному по полосе, каналу с аддитивным гауссовским шумом. Мы начнем с передаваемого (эквивалентного низкочастотного) сигнала, определяемого (9.2.1). Принимаемый сигнал (эквивалентный низкочастотный) выражается так

$$r_l(t) = \sum_n I_n h(t - nT) + z(t), \qquad (10.1.1)$$

где $h(t)$ представляет отклик эквивалентного низкочастотного канала на входной сигнальный импульс $g(t)$, а $z(t)$ представляет АБГШ.

Сначала мы покажем, что оптимальный демодулятор можно реализовать как фильтр, согласованный с $h(t)$, за ним идет операция стробирования со скоростью передачи символов $1/T$ и последующий алгоритм обработки для оценивания информационной последовательности $\{I_n\}$ по принятым отсчётным значениям. Следовательно, отсчеты на выходе согласованного фильтра достаточны для оценивания последовательности $\{I_n\}$.

10.1.1. Оптимальный приёмник максимального правдоподобия

Представим принимаемый сигнал $r_l(t)$ в виде разложения

$$r_l(t) = \lim_{N\to\infty} \sum_{k=1}^{N} r_k f_k(t), \qquad (10.1.2)$$

где $\{f_k(t)\}$ – полная система ортонормированных функций, а $\{r_k\}$ – наблюдаемые (на фоне шума) случайные величины, полученные проектированием $r_l(t)$ на ансамбль $\{f_k(t)\}$. Легко показать, что

$$r_k = \sum_n I_n h_{kn} + z_k, \quad k = 1, 2, \ldots, \qquad (10.1.3)$$

где h_{kn} – величина, полученная от проектирования $h(t-nT)$ на $f_k(t)$, а z_k – величина, полученная от проектирования $z(t)$ на $f_k(t)$. Последовательность $\{z_k\}$ является гауссовской с нулевым средним и ковариацией

$$\frac{1}{2}E(z_k^* z_m) = N_0 \delta_{km}. \qquad (10.1.4)$$

Совместная ФПВ случайных величин $\mathbf{r}_N \equiv [r_1 r_2 \ldots r_N]$ при условии передачи последовательности $\mathbf{I}_p \equiv [I_1\ I_2 \ldots I_p]$, где $p \le N$, равна

$$p\left(\mathbf{r}_N \middle| \mathbf{I}_p\right) = \left(\frac{1}{2\pi N_0}\right)^N \exp\left(-\frac{1}{2N_0}\sum_{k=1}^{N}\left|r_k - \sum_n I_n h_{kn}\right|^2\right). \qquad (10.1.5)$$

В пределе, когда число N наблюдаемых случайных величин становится неограниченным, логарифм $p\left(\mathbf{r}_N \middle| \mathbf{I}_p\right)$ пропорционален метрикам $PM(\mathbf{I}_p)$, определяемые так

$$\begin{aligned} PM(\mathbf{I}_p) &= -\int_{-\infty}^{\infty}\left|r_l(t) - \sum_n I_n h(t-nT)\right|^2 dt \\ &= -\int_{-\infty}^{\infty}|r_l(t)|^2\, dt + 2\,\mathrm{Re}\sum_n\left[I_n^*\int_{-\infty}^{\infty} r_l(t)h^*(t-nT)dt\right] \\ &\quad -\sum_n\sum_m I_n^* I_m \int_{-\infty}^{\infty} h^*(t-nt)h(t-mT)dt. \end{aligned} \qquad (10.1.6)$$

Максимально-правдоподобные оценки символов $I_1, I_2, \ldots, I_p$ – это такие, которые максимизируют (10.1.6). Заметим, однако, что интеграл от $|r_l(t)|^2$ общий для всех метрик и, следовательно, его можно исключить. Второй интеграл, включающий $r_l(t)$, зависит от величины

$$y_n \equiv y(nT) = \int_{-\infty}^{\infty} r_l(t)h^*(t-nT)dt. \qquad (10.1.7)$$

Эти величины можно генерировать, пропуская $r_l(t)$ через фильтр, согласованный с

$h(t)$ и стробируя выход со скоростью $1/T$. Отсчеты $\{y_n\}$ образуют ряд достаточных статистик для вычисления $PM(\mathbf{I}_p)$ или, что эквивалентно, корреляционных метрик

$$CM(\mathbf{I}_p) = 2\mathrm{Re}\left(\sum_n I_n^* y_n\right) - \sum_n \sum_m I_n^* I_m x_{n-m}, \qquad (10.1.8)$$

где, по определению, $x(t)$ – отклик согласованного фильтра на $h(t)$, а

$$x_n \equiv x(nT) = \int_{-\infty}^{\infty} h^*(t) h(t+nT)dt. \qquad (10.1.9)$$

Таким образом, $x(t)$ представляет выход фильтра, имеющего импульсную характеристику $h^*(-t)$ и вход $h(t)$. Другими словами, $x(t)$ представляет автокорреляционную функцию $h(t)$. Соответственно $\{x_n\}$ представляют отсчёты автокорреляционной функции $h(t)$, взятые периодически через $1/T$. Мы не интересуемся тонкостями построения фильтра, согласованного с $h(t)$, поскольку на практике мы можем ввести необходимую задержку, чтобы обеспечить реализуемость согласованного фильтра.

Если мы подставим для $r_l(t)$ в (10.1.2) соотношение (10.1.1) мы получим

$$y_k = \sum_n I_n x_{k-n} + \nu_k, \qquad (10.1.10)$$

где ν_k означает последовательность отсчётов аддитивного шума на выходе фильтра, т.е.

$$\nu_k = \int_{-\infty}^{\infty} z(t) h^*(t-kT)dt. \qquad (10.1.11)$$

Выходы демодулятора (согласованного фильтра) в моменты отсчёта искажены МСИ, как указано в (10.1.10). В любой практической системе разумно предположить, что МСИ влияет на ограниченное число символов. Таким образом, мы можем предположить, что $x_n = 0$ для $|n| > L$. Следовательно, МСИ, наблюдаемую на выходе демодулятора, можно рассматривать как выход машины с конечным числом состояний. Это позволяет выход канала с МСИ представить диаграммой решётки, а оценки максимального правдоподобия информационной последовательности $(I_1, I_2, \ldots, I_p)$ определяются наиболее вероятным путём по решётке, при данной принятой на выходе демодулятора последовательности $\{y_n\}$. Ясно, что алгоритм Витерби обеспечивает эффективный метод выполнения поиска по такой решётке.

Метрики, которые вычисляются при максимально правдоподобном последовательном оценивании (МППО, MLSE) последовательности $\{I_k\}$, определены (10.1.8). Можно видеть, что эти метрики можно вычислить рекуррентно посредствам алгоритма Витерби согласно соотношению

$$CM_n(\mathbf{I}_n) = CM_{n-1}(\mathbf{I}_{n-1}) + \mathrm{Re}\left[I_n^*\left(2y_n - x_0 I_n - 2\sum_{m=1}^{L} x_m I_{n-m}\right)\right]. \qquad (10.1.12)$$

На рис.10.1.1 дана блок-схема оптимального приёмника для канала с МСИ и АБГШ.

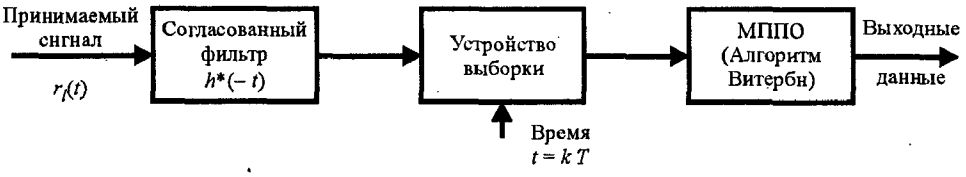


Рис.10.1.1. Оптимальный приёмник для канала с АБГШ и МСИ

10.1.2. Модель канала с МСИ с дискретным временем

При рассмотрении ограниченных по полосе каналов с МСИ удобно разработать эквивалентную модель с дискретным временем для аналоговой (с непрерывным временем) системы. Поскольку передатчик посылает символы в дискретные моменты времени со скоростью $1/T$ символов в секунду, а стробированный выход согласованного фильтра приемника также является сигналом дискретного времени с отчетами, возникающими со скоростью $1/T$, то следует, что каскадное соединение аналогового фильтра передатчика с импульсной характеристикой $g(t)$, канала с импульсной характеристикой $c(t)$, согласованного фильтра в приемнике с импульсной характеристикой $h^*(-t)$ и стробирующего устройства можно представить эквивалентным трансверсальным фильтром с дискретным временем, имеющий набор коэффициентов усиления $\{x_k\}$. Следовательно, мы имеем эквивалентный трансверсальный фильтр с дискретным временем, который покрывает временной интервал $2LT$ секунд. Его входом является информационная последовательность символов $\{I_k\}$, а его выходом является последовательность с дискретным временем $\{y_k\}$, определяемая (10.1.10). Эквивалентная модель с дискретным временем дана на рис.10.1.2.

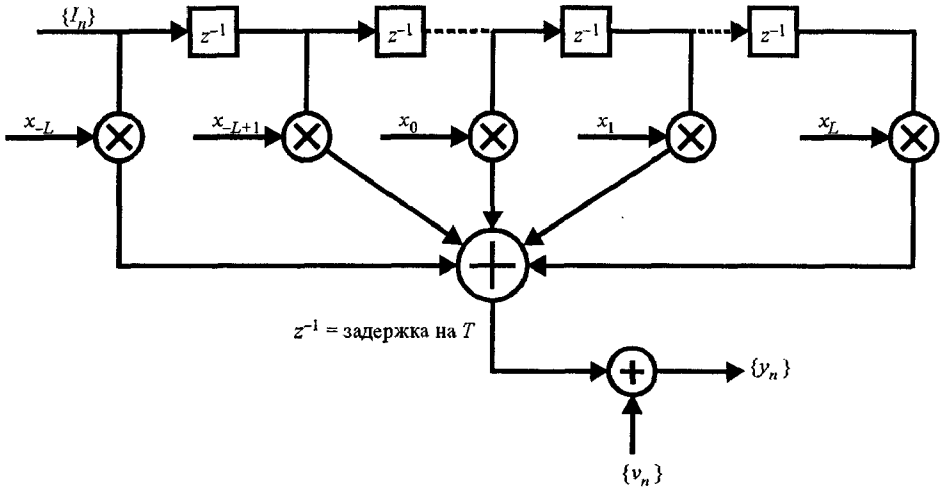


Рис. 10.1.2. Эквивалентная модель дискретного времени для канала с МСИ

Основная трудность при использовании этой модели с дискретным временем возникает при оценивании качества различной техники выравнивания или техники оценивания, что обсуждается в следующих разделах. Трудности обусловлены корреляцией отсчётов шумовой последовательности $\{v_k\}$ на выходе согласованного фильтра. Ряд шумовых величин $\{v_k\}$ образуют последовательность с гауссовским распределением, с нулевым средним и автокорреляционной функцией (смотри задачу 10.5)

$$\frac{1}{2}E(v_k^* v_j) = \begin{cases} N_0 x_{k-j} & (|k-j| \le L) \\ 0 & (\text{при других } k, j) \end{cases} \qquad (10.1.13)$$

Таким образом, шумовая последовательность коррелирована, если не выполняется условие $x_k = 0, \ k \neq 0$. Поскольку более удобно иметь дело при расчёте такой характеристики качества как вероятность ошибки с белой шумовой последовательностью, то желательно обелить шумовую последовательность путём дальнейшей фильтрации последовательности $\{y_k\}$.

Обеляющий фильтр с дискретным временем определяется следующим образом:

Пусть $X(z)$ обозначает (двухстороннее) z-преобразование отсчетов автокорреляционной функции $\{x_k\}$, т.е.

$$X(z)=\sum_{k=-L}^{L} x_k z^{-k}. \qquad (10.1.14)$$

Поскольку $x_k = x_{-k}^*$, следует $X(z)=X^*(z^{-1})$ и $2L$ корней $X(z)$ имеют симметрию, так что, если ρ корень, то $1/\rho^*$ тоже корень. Следовательно, $X(z)$ можно факторизовать и выразить так

$$X(z)=F(z)F^*(z^{-1}), \qquad (10.1.15)$$

где $F(z)$ – полином степени L, имеющий корни $\rho_1, \rho_2, \ldots, \rho_L$, а $F^*(z^{-1})$ – полином степени L, имеющий корни $1/\rho_1^*, 1/\rho_2^*, \ldots, 1/\rho_L^*$. Подходящий обеляющий фильтр имеет z-преобразование $1/F^*(z^{-1})$. Поскольку имеется 2^L возможных способов выбора корней $F^*(z^{-1})$, а каждый выбор ведет к фильтру, который одинаков по амплитудной характеристике и различен по фазе по сравнению с другими выборами, то мы предлагаем выбрать уникальное $F^*(z^{-1})$, имеющее минимальную фазу, т.е. полином, имеющий все свои корни внутри единичного круга. Тогда все корни $F^*(z^{-1})$ лежат внутри единичной окружности (с центром в начале координат), а $1/F^*(z^{-1})$ – физически реализуемый, устойчивый фильтр с дискретным временем. Следовательно, пропуская последовательность $\{y_k\}$ через цифровой фильтр $1/F^*(z^{-1})$ получаем выходную последовательность $\{\upsilon_k\}$, которую можно представить так

$$\upsilon_k=\sum_{n=0}^{L} f_n I_{k-n}+\eta_k, \qquad (10.1.16)$$

где $\{\eta_k\}$ – последовательность отсчетов гауссовского белого шума с нулевым средним, а $\{f_k\}$ – набор взвешивающих коэффициентов в эквивалентном трансверсальном фильтре с дискретным временем, имеющий передаточную функцию $F(z)$ (причём не $F^*(z^{-1})$). В общем последовательность $\{\upsilon_k\}$ комплексная .

В совокупности каскадное соединение фильтра передатчика $g(t)$, канала $c(t)$, согласованного фильтра $h^*(-t)$, стробирующего устройства и фильтра для взвешивания шума с дискретным временем $1/F^*(z^{-1})$ можно представить в виде эквивалентного трансверсального фильтра с дискретным временем, имеющего набор взвешивающих коэффициентов $\{f_k\}$. Аддитивная шумовая последовательность $\{\eta_k\}$, искажающая сигнал на выходе трансверсального фильтра с дискретным временем, является белой гауссовской шумовой последовательностью с нулевым средним и дисперсией N_0. Рис.10.1.3 иллюстрирует модель эквивалентной дискретной системы с белым шумом.

Мы будем ссылаться на эту модель, как на *эквивалентную модель с дискретным временем и белым шумом.*

Пример 10.1.1. Допустим, что сигнальный импульс передатчика $g(t)$ имеет длительность T и единичную энергию, а принимаемый сигнальный импульс равен $h(t)=g(t)+ag(t-T)$.

Определим эквивалентную модель с дискретным временем и белым шумом. Отсчеты автокорреляционной функции определены так

$$x_k = \begin{cases} a^* & (k=-1) \\ 1+|a|^2 & (k=0) \\ a & (k=1). \end{cases} \qquad (10.1.17)$$

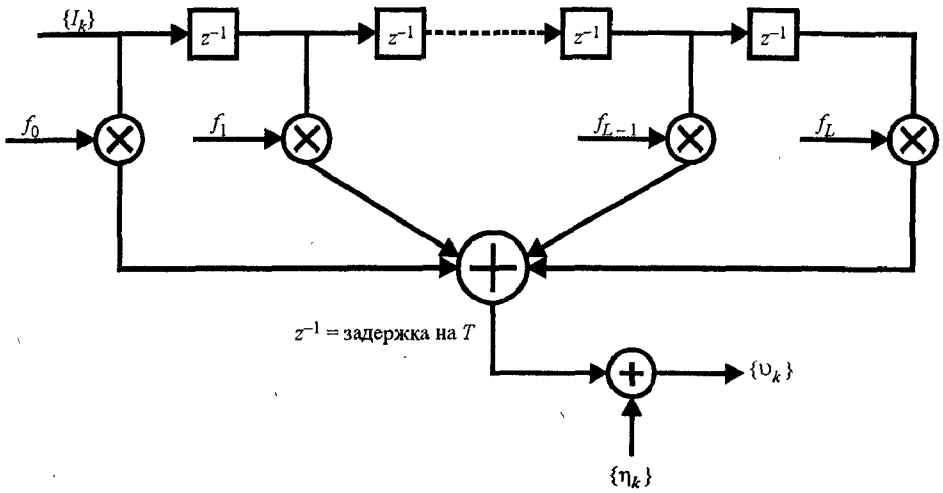


Рис. 10.1.3. Эквивалентная модель дискретного времени для каиала с МСИ и АБГШ

Затем, z-преобразование x_k даёт

$$X(z) = \sum_{k=-1}^{1} x_k z^{-k} = a^* z + (1+|a|^2) + az^{-1} = (az^{-1}+1)(a^* z+1). \qquad (10.1.18)$$

Предполагая, что $|a|>1$, выберем $F(z) = az^{-1}+1$ так, чтобы эквивалентный трансверсальный фильтр состоял из двух ячеек, имеющих коэффициенты усиления ячеек $f_0 = 1, f_1 = a$. Заметим, что корреляционную последовательность $\{x_k\}$ можно выразить через $\{f_n\}$ так

$$x_k = \sum_{n=0}^{L-k} f_n^* f_{n+k}, \qquad k = 0,\ 1,\ 2,\ldots,L. \qquad (10.1.19)$$

Если канальный отклик меняется медленно со временем, согласованный фильтр приемника становится меняющимся во времени фильтром (с переменными параметрами). В этом случае изменение во времени пары канал – согласованный фильтр приводит к фильтру с дискретным временем с переменными во времени коэффициентами. Как следствие, мы имеем эффект переменной во времени МСИ, которую можно моделировать фильтром, показанным на рис.10.1.3, у которого коэффициенты медленно меняются во времени.

Линейная фильтровая модель с дискретным временем и белым шумом для МСИ отражает то, что происходит при высокоскоростной передаче по идеальному ограниченному по полосе каналу. Она будет использоваться на протяжении всей этой главы при обсуждении техники компенсации МСИ. В общем, методы компенсации называют *техникой выравнивания или алгоритмом выравнивания.*

10.1.3. Алгоритм Витерби для модели фильтра с дискретным временем и белым шумом

Алгоритм МПП0 для оценки информационной последовательности $\{I_k\}$ наиболее легко описывается через принимаемую последовательность $\{\upsilon_k\}$ на выходе обеляющего фильтра. В присутствии МСИ, которое покрывает L+1 символа (L интерферирующих

компонент), реализация правила МППО эквивалентно оцениванию состояния конечного автомата с дискретным временем. Конечный автомат в этом случае является эквивалентом канала с дискретным временем с коэффициентами $\{f_k\}$, а его состояние в любой момент времени определяется L новыми (последними) входами, т.е. состояние в точке k определяется так:

$$S_k = (I_{k-1}, I_{k-2}, \ldots, I_{k-L}), \qquad (10.1.20)$$

где $I_k = 0$ для $k \le 0$. Таким образом если информационные символы являются M-ичными, канальный фильтр имеет M^L состояний. Следовательно, канал описывается M^L состояниями решётки и алгоритм Витерби можно использовать для определения наиболее вероятного пути на решетке.

Метрики, используемые в поиске по решётке, подобны метрикам, используемым при декодировании мягких решений свёрточных кодов. Вкратце, мы начинаем с отсчётов $\upsilon_1, \upsilon_2, \ldots, \upsilon_{L+1}$, по которым вычисляем M^{L+1} метрик

$$\sum_{k=1}^{L+1} \ln p(\upsilon_k | I_k, I_{k-1}, \ldots, I_{k-L}). \qquad (10.1.21)$$

M^{L+1} возможных последовательностей $I_{L+1}, I_L, \ldots, I_2, I_1$ подразделяется на M^L групп, соответствующих M^L состояниям $I_{L+1}, I_L, \ldots, I_2$. Заметим, что M последовательностей в каждой группе (состояний) отличается в символе I_1 и соответствуют путям по решетке, которые сходятся в одном узле. Из M последовательностей в каждом из M^L состояний мы выберем последовательность с наибольшей вероятностью (по отношению к I_1) и определяем для выживших последовательностей метрики

$$PM_1(\mathbf{I}_{L+1}) = PM_1(I_{L+1}, I_L, \ldots, I_2) = \max_{I_1} \sum_{k=1}^{L+1} \ln p(v_k | I_k, I_{k-1}, \ldots, I_{k-L}). \quad (10.1.22)$$

M-L оставшихся последовательностей из каждой из M^L групп исключаются. Таким образом мы оставляем M^L выживших последовательностей и их метрик.

При приеме υ_{L+2} M^L выживших последовательностей расширяются на один шаг и вычисляются соответствующие M^{L+1} вероятностей для расширенных последовательностей, используя предыдущие метрики и новое приращение, которое равно $\ln p(\upsilon_{L+2} | I_{L+2}, I_{L+1}, \ldots, I_2)$. Снова M^{L+1} последовательностей делятся на M^L групп, соответствующих M^L возможным состояниям $(I_{L+2}, \ldots, I_3)$ и из каждой группы выбирается наиболее вероятная последовательность, в то время как другие M-1 последовательностей отбрасываются.

Описанная процедура продолжается с приемом последовательных сигнальных отсчетов. В общем при приёме υ_{L+k} вычисляются метрики[1]

$$PM_k(\mathbf{I}_{L+k}) = \max_{I_k} \left[\ln p(v_{L+k} | I_{L+k}, \ldots, I_k) + PM_{k-1}(\mathbf{I}_{L+k-1}) \right] \qquad (10.1.23)$$

и определяются вероятности M^L выживших последовательностей. Таким образом по мере приема каждого отсчета сигнала, алгоритм Витерби включает в себя сначала вычисление M^{L+1} вероятностей

$$\ln p(\upsilon_{L+k} | I_{L+k}, \ldots, I_k) + PM_{k-1}(\mathbf{I}_{L+k-1}), \qquad (10.1.24)$$

---

[1] Мы видели, что метрики $PM_K(I)$ связаны с евклидовым расстоянием $DM_K(I)$, когда аддитивный шум гауссовский.

соответствующих M^{L+1} последовательностям, которые формируют продолжение M^L выживших последовательностей на предыдущих шагах процесса. Затем M^{L+1} последовательностей подразделяются на M^L групп. Каждая группа содержит M последовательностей, которые заканчиваются тем же набором символом $I_{L+k},\ldots,I_{k+1}$ и отличается в символе I_k. Из каждой группы из M последовательностей мы выбираем одну, имеющую наибольшую вероятность, как отмечено (10.2.23), в то время как оставшиеся M-1 последовательностей исключаются. Таким образом мы оставляем снова M^L последовательностей, имеющие метрики $PM_k(\mathbf{I}_{L+k})$.

Как отмечено ранее, задержка в детектировании каждого информационного символа по Витерби, вообще говоря, меняется. На практике изменение задержки устраняют путём удержания выживших последовательностей с q последними символами, где $q >> L$. Тем самым достигается фиксированная задержка. В случае, когда M^L выживших последовательностей на k-м шаге не совпадают в символе I_{k-q}, можно выбрать символ в наиболее вероятной последовательности Потеря в качестве, возникающая из-за этой субоптимальной процедуры оценивания, пренебрежимо мала, если $q \geq 5L$.

Пример 10.1.2. Для иллюстративных целей предположим, что для передачи четырехуровневой (M=4) АМ используется дуобинарный сигнальный импульс. Таким образом, каждый символ – это число, выбираемое из ряда (–3, –1, 1, 3). Контролируемая МСИ в этом сигнале с парциальным откликом представлена эквивалентной моделью канала с дискретным временем, показанной на рис.10.1.4. Предположим, мы приняли отсчеты υ_1 и υ_2, где

$$\begin{aligned} \upsilon_1 &= I_1 + \eta_1 \\ \upsilon_2 &= I_2 + I_1 + \eta_2 \end{aligned} \qquad (10.1.25)$$

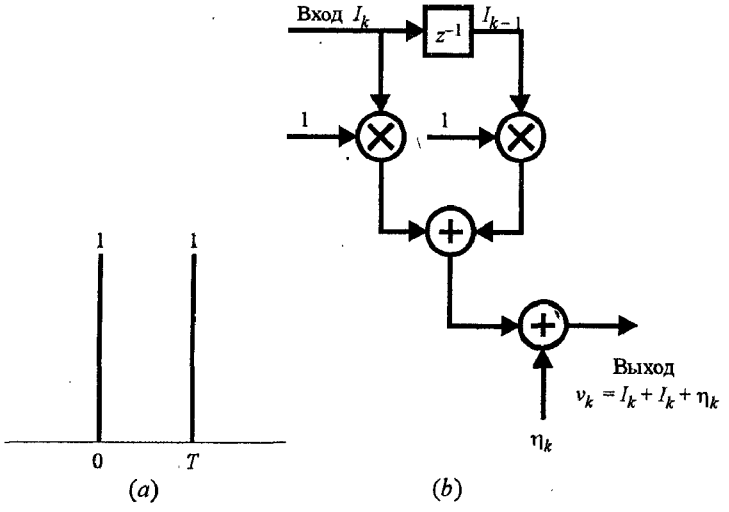


Рис. 10.1.4. Эквивалентная модель дискретного времени для межсимвольной интерференции, образованной дуобинарным импульсом

а $\{\eta_i\}$ является последовательностью статистически независимых гауссовских случайных величин с нулевым средним. Мы можем теперь вычислить 16 метрик

$$PM_1(I_2, I_i) = -\sum_{k=1}^{2}\left(v_k - \sum_{j=o}^{1} I_{k-j}\right)^2, \quad I_1, I_2 = \pm 1, \pm 3, \qquad (10.1.26)$$

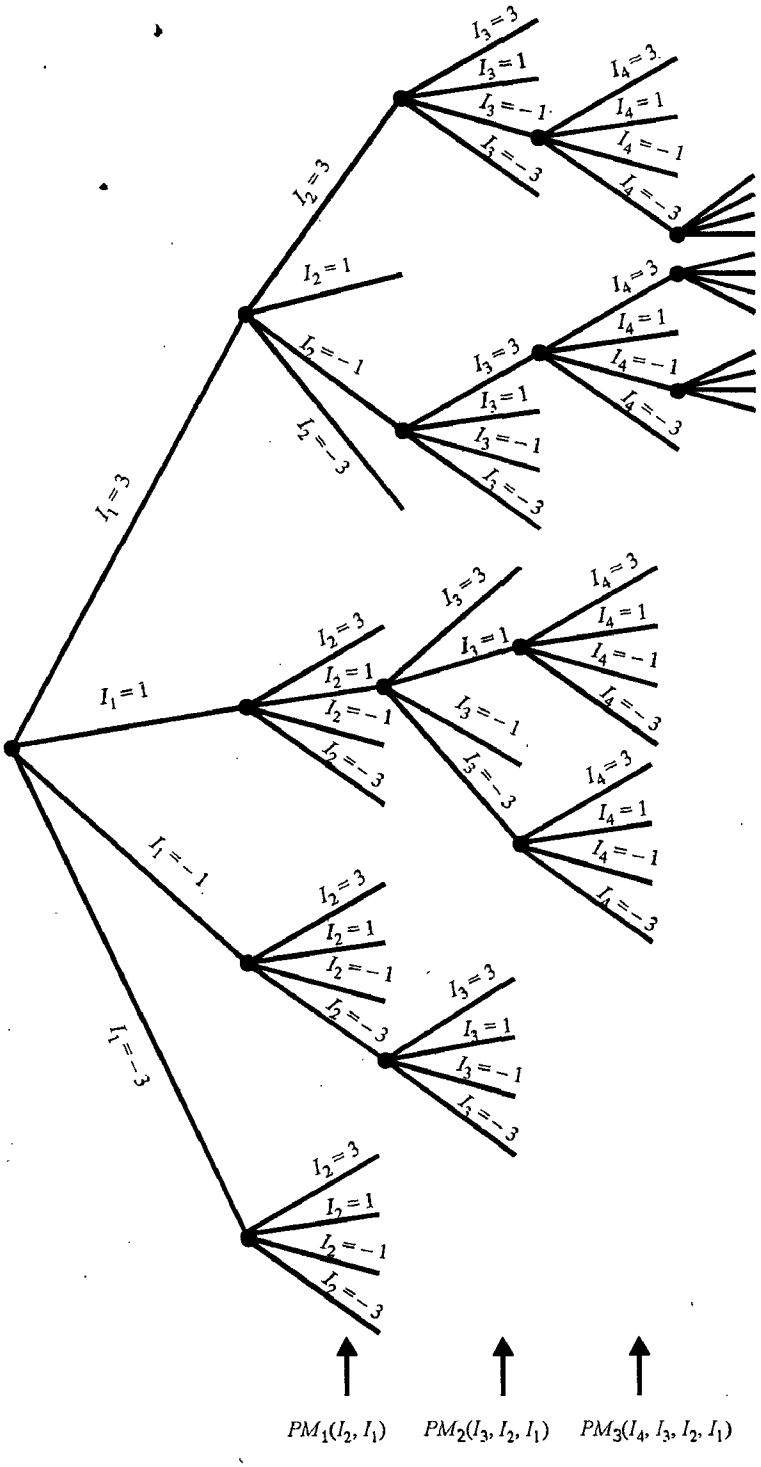


Рис. 10.1.5. Древовидная диаграмма для декодирования по Витерби дуобинарного импульса

где $I_k = 0$ для $k \le 0$.

Заметим, что не все последовательно принимаемые сигналы $\{\upsilon_i\}$ включают в себя I_1. Таким образом, на этом шаге мы можем исключить 12 из 16 возможных пар $\{I_1, I_2\}$. Этот шаг иллюстрирует древовидная диаграмма, показанная на рис.10.1.5.

Другими словами, после вычисления 16 метрик, соответствующих 16 путям древовидной диаграммы, мы исключаем три из четырех возможных путей, которые кончаются на $I_2 = 3$ и накапливаем наиболее правдоподобные из этих четырёх. Таким образом, метрики для выживших путей равны

$$PM_1(I_2 = 3,\, I_1) = \max_{I_1}\left[-\sum_{k=1}^{2}\left(\upsilon_k - \sum_{j=0}^{1} I_{k-j}\right)^2\right]$$

Процесс повторяется для каждого набора четырех путей, заканчивающихся на $I_2 = 1$, $I_2 = -1$ и $I_2 = -3$. Таким образом, четыре пути и их соответствующие метрики выживают после того, как приняты υ_1 и υ_2.

Когда принято υ_3, четверо путей расширяется так, как показано на рис.10.1.5, чтобы производить 16 путей и 16 соответствующих метрик, определяемых так

$$PM_2(I_3, I_2, I_1) = PM_1(I_2, I_1) - \left(v_3 - \sum_{j=0}^{1} I_{3-j}\right)^2. \qquad (10.1.27)$$

Из четырех путей, заканчивающихся I_3=3 мы сохраняем наиболее правдоподобные. Это процедура снова повторяется для $I_3 = 1$, $I_3 = -1$ и $I_3 = -3$. Следовательно, только четверо путей выживают на этом шаге. Процедура затем повторяется для каждого последовательного принимаемого сигнала υ_k для $k > 3$.

10.1.4. Качество алгоритма МППО для каналов с МСИ

Теперь определим вероятность ошибки при использовании алгоритма МППО (MLSE) для принимаемой информационной последовательности, если информация предается посредством АМ, а аддитивный шум в канале гауссовский. Похожесть между сверточным кодом и МСИ конечной длительности в канале подразумевает, что метод вычисления вероятности ошибки последней вытекает из первой. В частности, метод вычисления качества декодирования мягких решений сверточного кода посредствам алгоритма Витерби, описанный в разделе 8.2.3, применим здесь с некоторой модификацией.

При использовании в канале с аддитивным гауссовским шумом и МСИ сигналов АМ, метрики, используемые в алгоритме Витерби, можно выразить как в (10-1-23) или, что эквивалентно, так

$$PM_{k-L}(\mathbf{I}_k) = PM_{k-L-1}(\mathbf{I}_{k-1}) - \left(v_k - \sum_{j=o}^{L} f_j I_{k-j} \right)^2, \qquad (10.1.28)$$

где символы $\{I_n\}$ могут принять значение $\pm d, \pm 3d, ..., \pm(M-1)d$, а $2d$ – это расстояние между соседними уровнями. Решетка имеет M^L состояний и определяется в момент k так

$$S_k(I_{k-1}, I_{k-2}, \ldots, I_{k-L}). \qquad (10.1.29)$$

Обозначим оцененные символы посредством алгоритма Витерби через $\{\tilde{I}_n\}$, а соответствующие оцененные состояние в момент k так

$$\tilde{S}_k = (\tilde{I}_{k-1}, \tilde{I}_{k-2}, \ldots, \tilde{I}_{k-L}). \qquad (10.1.30)$$

Теперь предположим, что оцениваемый путь по решётке ответвляется от правильного пути в момент k и сливается с правильным путём в момент k+1. Таким образом, $\tilde{S}_k = S_k$ и $\tilde{S}_{k+l} = S_{k+l}$, но $\tilde{S}_m \neq S_m$ для $k < m < k+L$. Как и в свёрточном коде, мы назовём это *ошибочным событием*. Поскольку МСИ канала простирается на $L+1$ символов, то следует, что $l \geq L+1$

Для такого ошибочного события мы имеем $\tilde{I}_k \neq I_k$ и $\tilde{I}_{k+l-L-1} \neq I_{k+l-L-1}$, но $\tilde{I}_m = I_m$ для $k-L \leq m \leq k-1$ и $k+l-L \leq m \leq k+l-1$. Удобно определить вектор ошибки $\tilde{\varepsilon}$, соответствующий этим ошибочным событием, так

$$\boldsymbol{\varepsilon} = [\varepsilon_k \;\; \varepsilon_{k+1} \;\; \ldots \;\; \varepsilon_{k+l-L-1}], \qquad (10.1.31)$$

где компоненты $\tilde{\varepsilon}$ определяются так

$$\varepsilon_j = \frac{1}{2d}(I_j - \tilde{I}_j), \quad j = k, \; k+1..., \; k+l-L-1. \qquad (10.1.32)$$

Нормирующий множитель $1/2d$ в (10.1.32) приводит к тому, что элемент ε_j принимает значения $\pm 1, \pm 2, \pm 3, \ldots, \pm(M-1)$. Более того, вектор ошибок характеризуется свойствами, что $\varepsilon_k \neq 0, \varepsilon_{k+l-L-1} \neq 0$ и что нет последовательности из L соседних элементов, которые равны нулю. С вектором ошибок в (10.1.31) связан полином степени $l-L-1$.

$$\varepsilon(z) = \varepsilon_k + \varepsilon_{k+1} z^{-1} + \varepsilon_{k+2} z^{-2} + \ldots + \varepsilon_{k+l-L-1} z^{-(l-L-1)}. \qquad (10.1.33)$$

Мы хотим определить вероятность появления ошибочного события, которое начинается в момент k и характеризуется вектором ошибок $\tilde{\varepsilon}$, определяемым (10.1.31) или, что эквивалентно, полиномом (10.1.33). Чтобы найти его, будем следовать процедуре, разработанной Форни (1972). Конкретнее, чтобы произошло ошибочное событие $\tilde{\varepsilon}$, должны произойти следующие три подсобытия E_1, E_2 *и* E_3:

E_1: в момент k, $\tilde{S}_k = S_k$;

E_2: если суммировать информационную последовательность $I_k, I_{k+1}, \ldots, I_{k+l-L-1}$ с масштабируемой последовательностью ошибок $2d(\varepsilon_k, \varepsilon_{k+1}, \ldots, \varepsilon_{k+l-L-1})$ должна получиться разрешённая последовательность, т.е. последовательность $\widetilde{I}_k, \widetilde{I}_{k+1}, \ldots, \widetilde{I}_{k+l-L-1}$ должна иметь значения, выбираемые из ряда $\pm d, \pm 3d, \pm \ldots \pm (M-1)d$;

E_3: для $k \le m < k+l$ сумма метрик ветвей оцениваемого пути превышает сумму метрик ветвей правильного пути.

Вероятность появления E_3 равна

$$P(E_3) = P\left[\sum_{i=k}^{k+l-1}\left(\upsilon_i - \sum_{j=0}^{L} f_i \widetilde{I}_{i-j}\right)^2 < \sum_{i=k}^{k+l-1}\left(\upsilon_i - \sum_{j=0}^{L} f_i I_{i-j}\right)^2\right] \qquad (10.1.34)$$

Но

$$\upsilon_i = \sum_{j=0}^{L} f_i I_{i-j} + \eta_i, \qquad (10.1.35)$$

где $\{\eta_i\}$ – вещественная белая гауссовская шумовая последовательность. Подстановка (10.1.35) в (10.1.34) даёт

$$\begin{aligned} P(E_3) &= P\left[\sum_{i=k}^{k+l-1}\left(\eta_i + 2d\sum_{j=0}^{L} f_i \varepsilon_{i-j}\right)^2 < \sum_{i=k}^{k+l-1}\eta_i^2\right] \\ &= P\left[4d\sum_{i=k}^{k+l-1}\eta_i\left(\sum_{j=0}^{L} f_i \varepsilon_{i-j}\right) < -4d^2\sum_{i=k}^{k+l-1}\left(\sum_{j=0}^{L} f_i \varepsilon_{i-j}\right)^2\right], \end{aligned} \qquad (10.1.36)$$

где $\varepsilon_j = 0$ для $j < k$ и $j > k+l-L-1$. Если определим

$$\alpha_i = \sum_{j=0}^{L} f_i \varepsilon_{i-j}, \qquad (10.1.37)$$

тогда (10.1.36) можно выразить так

$$P(E_3) = P\left(\sum_{i=k}^{k+l-1}\alpha_i\eta_i < -d\sum_{i=k}^{k+l-1}\alpha_i^2\right), \qquad (10.1.38)$$

где множитель *4d*, общий для обоих слагаемых исключен. Теперь (10.1.38) как раз определяет вероятность того, что линейная комбинация статистически независимых гауссовских случайных величин с нулевым средним меньше некоторого отрицательного числа. Т.е.

$$P(E_3) = Q\left(\sqrt{\frac{2d^2}{N_0}\sum_{i=k}^{k+l-1}\alpha_i^2}\right). \qquad (10.1.39)$$

Для удобства определим

$$\delta^2(\varepsilon) = \sum_{i=k}^{k+l-1}\alpha_i^2 = \sum_{i=k}^{k+l-1}\left(\sum_{j=0}^{L} f_i \varepsilon_{i-j}\right)^2, \qquad (10.1.40)$$

где $\varepsilon_j = 0$ для $j < k$ и $j > k+l-L-1$. Заметим, что $\{\alpha_i\}$, определяемые свёрткой $\{f_i\}$ и $\{\varepsilon_j\}$, являются коэффициентами полинома

$$\alpha(z) = F(z)\varepsilon(z) = \alpha_k + \alpha_{k+1}z^{-1} + \ldots + \alpha_{k+l-1}z^{-(l-1)}. \qquad (10.1.41)$$

Далее $\delta^2(\varepsilon)$ просто равно коэффициенту при z^0 в полиноме

$$\alpha(z)\alpha(z^{-1}) = F(z)F(z^{-1})\varepsilon(z)\varepsilon(z^{-1}) = X(z)\varepsilon(z)\varepsilon(z^{-1}). \qquad (10.1.42)$$

Мы назовем $\delta^2(\varepsilon)$ *евклидовым весом* ошибочного события $\tilde{\varepsilon}$

Альтернативный метод для представление результата свертки $\{f_j\}$ и $\{q_j\}$ – это матричная форма

$$\boldsymbol{\alpha} = \mathbf{ef},$$

где $\boldsymbol{\alpha}$ – L_4-мерный вектор, $\mathbf{f}$ – $(L+1)$-мерный вектор, а $\mathbf{e} - l \times (L+1)$ матрица:

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_k \\ \alpha_{k+1} \\ \vdots \\ \alpha_{k+l-1} \end{bmatrix} \qquad \mathbf{f} = \begin{bmatrix} f_0 \\ f_1 \\ \vdots \\ f_L \end{bmatrix}$$

$$\mathbf{e} = \begin{bmatrix} \varepsilon_k & 0 & 0 & \dots & 0 & \dots & 0 \\ \varepsilon_{k+1} & \varepsilon_k & 0 & \dots & 0 & \dots & 0 \\ \varepsilon_{k+2} & \varepsilon_{k+1} & \varepsilon_k & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots & & \vdots \\ \varepsilon_{k+l-1} & \dots & \dots & \dots & \dots & \dots & \varepsilon_{k+l-L-1} \end{bmatrix} \qquad (10.1.43)$$

Тогда

$$\delta^2(\varepsilon) = \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\alpha} = \mathbf{f}^{\mathrm{T}}\mathbf{e}^{\mathrm{T}}\mathbf{ef} = \mathbf{f}^{\mathrm{T}}\mathbf{Af}, \qquad (10.1.44)$$

где $\mathbf{A}$ $(L+1)\times(L+1)$ – матрица вида

$$\mathbf{A} = \mathbf{e}^{\mathrm{T}}\mathbf{e} = \begin{bmatrix} \beta_0 & \beta_1 & \beta_2 & \dots & \beta_L \\ \beta_1 & \beta_0 & \beta_1 & \dots & \beta_{L-1} \\ \beta_2 & \beta_1 & \beta_0 & \beta_1 & \beta_{L-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \beta_L & \dots & \dots & \dots & \beta_0 \end{bmatrix} \qquad (10.1.45)$$

и

$$\beta_m = \sum_{i=k}^{k+l-1-m} \varepsilon_i \varepsilon_{i+m}. \qquad (10.1.46)$$

Мы можем использовать или (10.1.40) и (10.1.41) или (10.1.45) и (10.1.46) для расчёта вероятности ошибки. Мы обсудим эти вычисления позже. Теперь же мы сделаем вывод, что вероятность подсобытия E_3, определяемого (10.1.39), можно выразить так

$$P(E_3) = Q\left(\sqrt{\frac{2d^2}{N_0}\delta^2(\varepsilon)}\right) = Q\left(\sqrt{\frac{6}{M^2-1}\gamma_{\text{ср}}\delta^2(\varepsilon)}\right), \qquad (10.1.47)$$

где мы использовали отношение

$$d^2 = \frac{3}{M^2-1} TP_{\text{ср}}. \qquad (10.1.48)$$

Для исключения d^2, а $\gamma_{\text{ср}} = TP_{\text{ср}} / N_0$. Заметим, что в отсутствии МСИ $\delta^2(\varepsilon) = 1$ и $P(E_3)$ пропорционально вероятности ошибки на символ в М-позиционной АМ.

Вероятность подсобытия E_2 зависит только от статистических свойств входной последовательности. Мы предположим, что информационные символы равновероятны и что символы в передаваемой последовательности статистически независимы. Тогда для ошибки вида $|\varepsilon_i| = j$, $j = 1, 2, \dots, M-1$ имеется М-j возможных значений I_i, таких что

$I_i = \tilde{I}_i + 2d\varepsilon_i$, следовательно

$$P(E_2)=\prod_{i=o}^{l-L-1}\frac{M-|i|}{M} \tag{10.1.49}$$

Вероятность подсобытия E_1 значительно более трудно вычислить точно из-за ее зависимости от подсобытия E_3. Это значит, что мы должны вычислять $P(E_1|E_3)$. Однако $P(E_1|E_3)=1-P_M$, где P_M – вероятность ошибки на символ. Следовательно $P(E_1|E_3)$ хорошо аппроксимируется (и ограничено сверху) единицей для разумных низких значений вероятности ошибки. Таким образом, вероятность ошибочного события $\tilde{\varepsilon}$ хорошо аппроксимируется и ограничена сверху так:

$$P(\varepsilon)\le Q\left(\sqrt{\frac{6}{M^2-1}\gamma_{\text{ср}}\delta^2(\varepsilon)}\right)\prod_{i=0}^{l-L-1}\frac{M-|i|}{M} \tag{10.1.50}$$

Пусть E является набором всех ошибочных событий ε, начавшихся в момент k и пусть $w(\varepsilon)$ является соответствующим числом ненулевых компонент (весом Хемминга или числом ошибочных символов) в каждом ошибочном событий ε. Тогда вероятность ошибки на символ ограничена сверху (объединённая граница) так

$$P_M\le\sum_{\varepsilon\in E}w(\varepsilon)P(\varepsilon)\le\sum_{\varepsilon\in E}w(\varepsilon)Q\left(\sqrt{\frac{6}{M^2-1}\gamma_{\text{ср}}\delta^2(\varepsilon)}\right)\prod_{i=0}^{l-L-1}\frac{M-|i|}{M}. \tag{10.1.51}$$

Теперь пусть D является множеством всех $\delta(\varepsilon)$. Для каждого $\delta\in D$, пусть E_δ являются подмножеством ошибочных событий для которых $\delta(\varepsilon)=\delta$. Тогда (10.1.51) можно выразить так

$$P_M\le\sum_{\delta\in D}Q\left(\sqrt{\frac{6}{M^2-1}\gamma_{\text{ср}}\delta^2}\right)\left[\sum_{\varepsilon\in E_\delta}w(\varepsilon)\prod_{i=0}^{l-L-1}\frac{M-|i|}{M}\right]\le\sum_{\delta\in D}K_\delta Q\left(\sqrt{\frac{6}{M^2-1}\gamma_{\text{ср}}\delta^2}\right), \tag{10.1.52}$$

где

$$K_\delta=\sum_{\varepsilon\in E_\delta}w(\varepsilon)\prod_{i=0}^{l-L-1}\frac{M-|i|}{M}. \tag{10.1.53}$$

Выражение для вероятности ошибки в (10.1.52) похоже по форме на вероятность ошибки для свёрточного кода при детектировании мягких решений, определяемая (8.2.26). Взвешивающие множители $\{K_\delta\}$ можно определить посредством диаграммы состояний ошибок, которая схожа диаграмме состояний сверточного кодера. Этот подход был показан Форни (1972) и Витерби и Омура (1979).

В общем, однако, использование диаграмм состояний ошибок для вычисления P_M утомительно. Вместо этого мы можем упростить вычисление P_M, сосредоточившись на основной член суммы (10.1.52). Из-за экспоненциальной зависимости каждого слагаемого суммы, выражение P_M в основном определяется слагаемым, соответствующим минимальному значению δ, которое обозначим $\delta_{\min}$. Тогда вероятность ошибки на символ можно аппроксимировать так

$$P_M\approx K_{\delta_{\min}}Q\left(\sqrt{\frac{6}{M^2-1}\gamma_{\text{ср}}\delta_{\min}^2}\right), \tag{10.1.54}$$

где

$$K_{\delta_{\min}}=\sum_{\varepsilon\in E_{\delta_{\min}}}w(\varepsilon)\prod_{i=0}^{l-L-1}\frac{M-|i|}{M} \tag{10.1.55}$$

В общем $\delta^2_{\min} \le 1$. Таким образом, $10\lg\delta^2_{\min}$ представляет потери в ОСШ, обусловленные МСИ.

Минимальное значение δ можно определить или из (10.1.40) или из оценки квадратичной формы (10.1.44) для различных последовательностей ошибок. В следующих двух примерах мы используем (10.1.40).

Пример 10.1.3. Рассмотрим двухпутевой канал $(L=1)$ с произвольными коэффициентами f_0 и f_1, удовлетворяющих условию $f_0^2 + f_1^2 = 1$. Характеристика канала равна

$$F(z) = f_0 + f_1 z^{-1}. \tag{10.1.56}$$

Для ошибочного события длиной n

$$\varepsilon(z) = \varepsilon_0 + \varepsilon_1 z^{-1} + \ldots + \varepsilon_{n-1} z^{-(n-1)}, \quad n \ge 1. \tag{10.1.57}$$

Произведение $\alpha(z) = F(z)\varepsilon(z)$ можно выразить так

$$\alpha(z) = \alpha_0 + \alpha_1 z^{-1} + \ldots + \alpha_n z^{-n}, \tag{10.1.58}$$

где $\alpha_0 = \varepsilon_0 f_0$, а $\alpha_n = f_1 \varepsilon_{n-1}$. Поскольку $\varepsilon_0 \ne 0$, $\varepsilon_{n-1} \ne 0$ и

$$\delta^2(\varepsilon) = \sum_{k=0}^{n} \alpha_k^2, \tag{10.1.59}$$

то следует что $\delta^2_{\min} \ge f_0^2 + f_1^2 = 1$.

Действительно $\delta^2_{\min} = 1$, когда возникает одна ошибка, т.е. $\varepsilon(z) = \varepsilon_0$. Таким образом заключаем, что в этом случае нет потерь в ОСШ при максимально правдоподобной оценке информационного символа, когда длина дисперсии канала (канального расстояния) равна 2.

Пример 10.1.4. Контролируемое МСИ при сигнале с парциальным откликом можно рассматривать как результат генерации канала с временным рассеянием. Таким образом, МСИ от дуобинарного импульса можно представить через (нормированную) канальную характеристику.

$$F(z) = \sqrt{\tfrac{1}{2}} + \sqrt{\tfrac{1}{2}} z^{-1} \tag{10.1.60}$$

Аналогично представление для модифицированного дуобинарного импульса равна

$$F(z) = \sqrt{\tfrac{1}{2}} - \sqrt{\tfrac{1}{2}} z^{-2} \tag{10.1.61}$$

Минимальное расстояние $\delta^2_{\min} = 1$ для любого ошибочного события в виде

$$\varepsilon(z) = \pm(1 - z^{-1} - z^{-2} - \ldots - z^{-(n-1)}), \quad n \ge 1 \tag{10.1.62}$$

для канала, определяемого (10.1.60), поскольку

$$\alpha(z) = \pm\sqrt{\tfrac{1}{2}} \mp \sqrt{\tfrac{1}{2}} z^{-n}$$

Аналогично, когда

$$\varepsilon(z) = \pm(1 + z^{-2} + z^{-4} + \ldots + z^{-2(n-1)}), \qquad n \ge 1, \tag{10.1.63}$$

$\delta^2_{\min} = 1$ для канала, определяемого (10.1.61), поскольку

$$\alpha(z) = \pm\sqrt{\tfrac{1}{2}} \mp \sqrt{\tfrac{1}{2}} z^{-2n}.$$

Таким образом, использование МППО в случае этих двух сигнальных откликов не ведет к потере в ОСШ. В противоположность этому, субоптимальное посимвольное детектирование, описанное выше, ведет к потере в 2,1 дБ.

Константу $K_{\delta_{\min}}$ для этих двух сигналов легко рассчитать. С предкодированием число выходных ошибочных символов (вес Хемминга), связанное с ошибочными событиями (10.1.62) и (10.1.63) равно 2. Таким образом

$$K_{\delta_{\min}} = 2\sum_{n=1}^{\infty} n\left(\frac{M-1}{M}\right)^n = 2M(M-1) \qquad (10.1.65)$$

Как заключительное упражнение мы рассмотрим оценивание $\delta_{\min}$ из квадратичной формы (10.1.44). Матрица **A** в (10.1.44) положительно определённая, следовательно, все её собственные значения положительные. Если $\{\mu_k(\varepsilon)\}$ являются собственными значениями, а $\{\mathbf{v}_k(\varepsilon)\}$ являются соответствующими ортогональными собственными векторами **A**, тогда для ошибочного события ε, квадратичную форма (10.1.44) можно представить так:

$$\delta^2(\varepsilon) = \sum_{k=1}^{L+1} \mu_k(\varepsilon)\left[\mathbf{f}^{\mathrm{T}}\mathbf{v}_k(\varepsilon)\right]^2. \qquad (10.1.66)$$

Другими словами, δ^2 выражается как линейная комбинация квадратов проекций канального вектора f на собственные векторы A. Каждый квадрат проекции в сумме взвешивается соответствующим собственным значением $\mu_k(\varepsilon)$, $k = 1, 2, \ldots L+1$. Тогда

$$\delta^2_{\min} = \min_{\varepsilon} \delta^2(\varepsilon) \qquad (10.1.67)$$

Интересно отметить, что наихудшую характеристику канала с рассеянием заданной длины $L+1$ можно получить, найдя собственный вектор, соответствующий минимальному собственному значению. Так если $\mu_{\min}(\varepsilon)$ – минимальное собственное значение для заданного ошибочного события e, а $\mathbf{v}_{\min}(\varepsilon)$, является соответствующим собственным вектором, тогда

$$\mu_{\min} = \min_{\varepsilon} \mu_{\min}(\varepsilon), \quad \mathbf{f} = \min_{\varepsilon} \mathbf{v}_{\min}(\varepsilon) \quad \text{и} \quad \delta^2_{\min} = \mu_{\min}.$$

Пример 10.1.5. Определим наихудший канал с рассеянием во времени длиной 3 (L=2), найдя минимальное собственное значение A для различных ошибочных событий. Итак,

$$F(z) = f_0 - f_1 z^{-1} + f_2 z^{-2},$$

где f_0, f_1 и f_2 – компоненты собственного вектора **A**, соответствующие минимальному собственному значению. Ошибочное событие вида

$$\varepsilon(z) = 1 - z^{-1}$$

ведёт к матрице

$$\mathbf{A} = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix},$$

которая имеет собственные значения $\mu_1 = 2$, $\mu_2 = 2+\sqrt{2}$, $\mu_3 = 2-\sqrt{2}$. Собственный вектор, соответствующий μ_3, равен

$$\mathbf{v}_3^{\mathrm{T}} = \left[\tfrac{1}{2} \quad \sqrt{\tfrac{1}{2}} \quad \tfrac{1}{2}\right]. \qquad (10.1.68)$$

Мы также хотим рассмотреть дуальное ошибочное событие

$$\varepsilon(z) = 1 - z^{-1},$$

которое ведет к матрице

$$\mathbf{A} = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix}.$$

Эта матрица имеет те же собственные значения, как для $\varepsilon(z) = 1 - z^{-1}$. Соответствующий собственный вектор для $\mu_3 = 2 - \sqrt{2}$ равен

$$\mathbf{v}_3^{\mathrm{T}} = \left[-\tfrac{1}{2} \quad \sqrt{\tfrac{1}{2}} \quad -\tfrac{1}{2}\right]. \qquad (10.1.69)$$

Другое ошибочное событие ведет к большим значениям $\mu_{\min}$. Таким образом, $\mu_{\min} = 2 - \sqrt{2}$ и наихудший канал характеризуется собственным вектором

$$\left[\tfrac{1}{2} \quad \sqrt{\tfrac{1}{2}} \quad \tfrac{1}{2}\right] \text{ или } \left[-\tfrac{1}{2} \quad \sqrt{\tfrac{1}{2}} \quad -\tfrac{1}{2}\right].$$

Потеря в ОСШ для канала равна

$$-10\log\delta^2_{\min} = -10\log\mu_{\min} = 2{,}3 \text{ дБ}.$$

Повторение приведённых выше вычислений для каналов с L=3,4 и 5 дает результаты, данные в табл.10.1.1.

Таблица 10.1.1. Максимальные потери качества и соответствующие характеристики канала

Рассеяние канала L+1	Потеря качества $-10\log\delta^2_{\min}$	Импульсная характеристика наихудшего канала
3	2,3	0,50; 0,71; 0,50
4	4,2	0,38; 0,60; 0,60; 0,38
5	5,7	0,29; 0,50; 0,58; 0,50; 0,29
6	7,0	0,23; 0,42; 0,52; 0,52; 0,42; 0,23

10.2. ЛИНЕЙНОЕ ВЫРАВНИВАНИЕ

Алгоритм МППО для канала с МСИ имеет вычислительную сложность, которая возрастает экспоненциально с длиной временного рассеяния в канале. Если объем алфавита символов равно M, а число интерферирующих символов, обуславливающих МСИ, равно L, алгоритм Витерби вычисляют M^{L+1} метрик для каждого нового принимаемого символа. Для большинства каналов, представляющих практический интерес, такая большая вычислительная сложность чрезмерно высока для её реализации.

В этом и последующих разделах мы опишем два подхода к субоптимальному канальному выравниванию для компенсации МСИ. Один подход использует линейный трансверсальный фильтр, который описывается в этом разделе. Структура этого фильтра имеет вычислительную сложность, являющуюся линейной функцией от величины канального рассеяния L.

Линейный фильтр, наиболее часто используемый для выравнивания, это трансверсальный фильтр, показанный на рис.10.2.1.

Его входом является последовательность $\{\upsilon_k\}$, определяемая (10.1.16), а его выходом являются оценки информационной последовательности $\{I_k\}$. Оценка k-го символа можно выразить так

$$\hat{I}_k = \sum_{j=-K}^{K} c_j \upsilon_{k-j}, \qquad (10.2.1)$$

где $\{c_j\}$ является $2K+1$ комплексно-значных взвешивающих коэффициентов для ячеек фильтра. Оценка $\hat{I}_k$ квантуется до ближайшего (по расстоянию) информационного

символа для формирования решения $\widetilde{I}_k$. Если $\widetilde{I}_k$ не идентично передаваемому символу I_k имеет место ошибка.

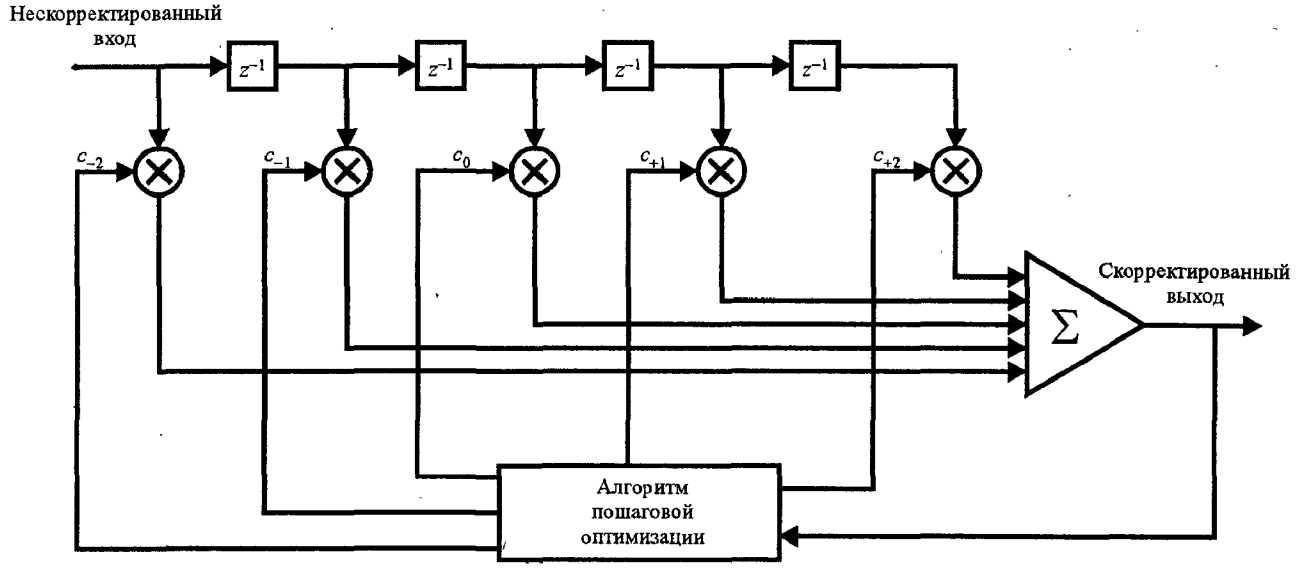


Рис. 10.2.1. Линейный трансверсальный фильтр

Значительные исследования были выполнены по нахождению критерия оптимизации коэффициентов фильтра $\{c_k\}$. Поскольку наиболее употребительная мера качества для цифровой системы связи – это средняя вероятность ошибки, желательно выбрать коэффициенты так, чтобы минимизировать этот показатель качества. Однако вероятность ошибки существенно нелинейная функция $\{c_j\}$. Следовательно, вероятность ошибки как показатель качества для оптимизации взвешивающих коэффициентов ячеек эквалайзера не практичен.

Два критерия нашли широкое распространение при оптимизации коэффициентов $\{c_j\}$ эквалайзера. Один – это критерий пикового искажения, а второй – критерий среднеквадратичной ошибки.

10.2.1. Критерий пикового искажения

Пиковое искажение просто определяется как наиболее плохой случай МСИ на выходе эквалайзера. Минимизацию этого показателя качества называют *критерием пикового искажения*. Сначала мы рассмотрим минимизацию пикового искажения, предполагая, что эквалайзер имеет неограниченное число ячеек. Затем мы обсудим случай, когда трансверсальный эквалайзер имеет конечное число ячеек.

Мы видели, что каскадное объединение модели линейного фильтра дискретного времени с импульсной характеристикой $\{f_n\}$ и эквалайзера, имеющего импульсную характеристику $\{c_n\}$, можно представить одним эквивалентным фильтром с импульсной характеристикой

$$q_n = \sum_{j=-\infty}^{\infty} c_j f_{n-j} . \qquad (10.2.2)$$

Это значит, что $\{q_n\}$ – это просто свертка $\{c_n\}$ и $\{f_n\}$. Считается, что эквалайзер имеет неограниченное число ячеек. Его выход в k-й отсчетный момент можно выразить в виде

$$\hat{I}_k = q_0 I_k + \sum_{n \neq k} I_n q_{k-n} + \sum_{j=-\infty}^{\infty} c_j \eta_{k-j} . \qquad (10.2.3)$$

Первое слагаемое в (10.2.3) представляет взвешенная версия желательного символа. Для удобства, мы нормируем q_0 к единице. Второе слагаемое является МСИ. Пиковое значение этой интерференции, которое называется *пиковым искажением*, равно

$$\mathcal{D}(\mathbf{c}) = \sum_{\substack{n=-\infty \\ n \neq 0}}^{\infty} |q_n| = \sum_{\substack{n=-\infty \\ n \neq 0}}^{\infty} \left| \sum_{j=-\infty}^{\infty} c_j f_{n-j} \right| . \qquad (10.2.4)$$

Таким образом, $\mathcal{D}(\mathbf{c})$ является функцией взвешивающих коэффициентов ячеек эквалайзера.

При помощи эквалайзера с неограниченным числом ячеек возможно выбрать веса ячеек так, что $\mathcal{D}(\mathbf{c}) = 0$, т.е. $q_n = 0$ для всех n, исключая $n = 0$. Это значит, что МСИ может быть полностью исключено. Величины весов ячеек для выполнения этой цели определяются из условия

$$q_n = \sum_{j=-\infty}^{\infty} c_j f_{n-j} = \begin{cases} 1 & (n = 0) \\ 0 & (n \neq 0) \end{cases} \qquad (10.2.5)$$

Взяв z-преобразование от (10.2.5), получим

$$Q(z) = C(z)F(z) = 1 \qquad (10.2.6)$$

или просто

$$C(z) = \frac{1}{F(z)}, \qquad (10.2.7)$$

где $C(z)$ означает z-преобразование $\{c_j\}$. Заметим, что эквалайзер с передаточной функцией $C(z)$ это просто обратный фильтр по отношению к линейной модели канального фильтра $F(z)$. Другими словами, полное исключение МСИ требует использования фильтра, обратного $F(z)$. Мы называем такой фильтр *фильтром с нулевыми взаимными помехами* («нуль-форсирующим» фильтром). Рис.10.2.2 иллюстрирует блок-схему эквивалентного канала с дискретным временем и эквалайзера.

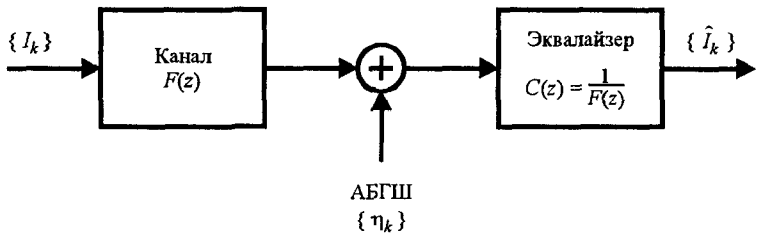


Рис. 10.2.2. Блок-схема канала с обнуляющим эквалайзером

Каскадное объединение обеляющего фильтра с передаточной функцией $1/F^*(z^{-1})$ и эквалайзера с нулевыми взаимными помехами (ЭНВП) с передаточной функцией $1/F(z)$ приводит к эквивалентному ЭНВП с передаточной функцией

$$C'(z) = \frac{1}{F(z)F^*(z^{-1})} = \frac{1}{X(z)}, \qquad (10.2.8)$$

как показано на рис.10.2.3.

Этот комбинированный фильтр имеет на входе последовательность $\{y_k\}$ отсчётов согласованного фильтра, определённую (10.1.10). Его выход состоит из желательных символов, искажённых только аддитивным гауссовским белым шумом с нулевым средним. Импульсная характеристика комбинированного фильтра равна

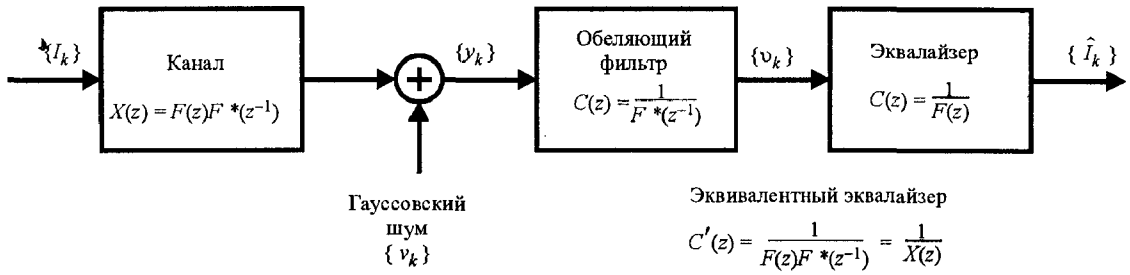


Рис. 10.2.3. Блок-схема канала с эквивалентным обнуляющим эквалайзером

$$c'_k = \frac{1}{2\pi j}\oint C'(z)z^{k-1}dz = \frac{1}{2\pi j}\oint \frac{z^{k-1}}{X(z)}dz, \qquad (10.2.9)$$

где интегрирование выполняется по замкнутому контуру, который содержит внутри себя область сходимости $C'(z)$. Поскольку $X(z)$ – это полином с $2L$ корнями $\left(\rho_1, \rho_2, \ldots \rho_L, 1/\rho_1^*, 1/\rho_2^*, \ldots 1/\rho_L^*\right)$, то следует, что $C'(z)$ должен сходиться в плоскости, внутри единичной окружности $\left(z = e^{j\phi}\right)$. Следовательно, контуром интегрирования может быть единичная окружность.

Качество эквалайзера с неограниченным числом ячеек, который полностью устраняет МСИ, легко выразить через отношение сигнал-шум (ОСШ) на его выходе. Для математического удобства мы нормируем энергию принимаемого сигнала к единице[1]. Это предполагает, что $q_0 = 1$ и что ожидаемая величина $|I_k|^2$ также равна единице. Тогда ОСШ равно обратной величине дисперсии шума σ_n^2 на выходе эквалайзера.

Величину σ_n^2 можно просто определить, если заметить, что шумовая последовательность $\{\nu_k\}$ на входе эквивалентного ЭНВП с характеристикой $C'(z)$ имеет нулевое среднее и спектральную плотность мощности

$$\Phi_{\nu\nu}(\omega) = N_0 X(e^{j\omega T}), \quad |\omega| \le \frac{\pi}{T}, \qquad (10.2.10)$$

где $X(e^{j\omega T})$ получено из $X(z)$ подстановкой $z = e^{j\omega T}$. Поскольку $C'(z) = 1/X(z)$, следует, что выходная шумовая последовательность эквалайзера имеет спектральную плотность мощности

$$\Phi_{nn}(\omega) = \frac{N_0}{X(e^{J\omega T})}, \quad |\omega| \le \frac{\pi}{T}. \qquad (10.2.11)$$

Следовательно, дисперсия шума на выходе эквалайзера

$$\sigma_n^2 = \frac{T}{2\pi}\int_{-\pi/T}^{\pi/T}\Phi_{nn}(\omega)d\omega = \frac{TN_0}{2\pi}\int_{-\pi/T}^{\pi/T}\frac{d\omega}{X(e^{j\omega T})}, \qquad (10.2.12)$$

а ОСШ на выходе ЭНВП равна

$$\gamma_\infty = 1/\sigma_n^2 = \left[\frac{TN_0}{2\pi}\int_{-\pi/T}^{\pi/T}\frac{d\omega}{X(e^{j\omega T})}\right]^{-1}, \qquad (10.2.13)$$

где индекс у γ указывает на то, что эквалайзер имеет неограниченное число ячеек.

Спектральная характеристика $X(e^{j\omega T})$, соответствующая преобразованию Фурье последовательности отсчётов $\{x_k\}$, имеет интересную связь с характеристикой аналогового фильтра $H(\omega)$, используемого в приёмнике. Поскольку

---

[1] Эта нормировка используется во всей главе для математического удобства.

$$x_k = \int_{-\infty}^{\infty} h^*(t)h(t+kT)dt,$$

то используя теорему Парсеваля имеем

$$x_k = \frac{1}{2\pi}\int_{-\infty}^{\infty} |H(\omega)|^2 e^{j\omega kT} d\omega \qquad (10.2.14)$$

где $H(\omega)$ – преобразование Фурье от $h(t)$. Но интеграл в 10.2.14 можно выразить в форме

$$x_k = \frac{1}{2\pi}\int_{-\pi/T}^{\pi/T}\left[\sum_{n=-\infty}^{\infty}\left|H\left(\omega+\frac{2\pi n}{T}\right)\right|^2\right] e^{j\omega kT} d\omega \qquad (10.2.15)$$

Теперь преобразование Фурье (дискретное) $\{x_k\}$ равно

$$X(e^{j\omega T}) = \sum_{k=-\infty}^{\infty} x_k e^{-j\omega kT}, \qquad (10.2.16)$$

а обратное преобразование Фурье легко выразить так

$$x_k = \frac{T}{2\pi}\int_{-\pi/T}^{\pi/T} X(e^{j\omega T}) e^{j\omega kT} d\omega. \qquad (10.2.17)$$

Из сравнения (10.2.15) и (10.2.17) мы получаем желательное соотношение между $X(e^{j\omega T})$ и $H(\omega)$. Оно таково

$$X(e^{j\omega T}) = \frac{1}{T}\sum_{n=-\infty}^{\infty}\left|H\left(\omega+\frac{2\pi n}{T}\right)\right|^2, \quad |\omega| \le \frac{\pi}{T}, \qquad (10.2.18)$$

где правая часть (10.2.18) называется *сложенным спектром* $|H(\omega)|^2$. Мы также видим, что $|H(\omega)|^2 = X(\omega)$, где $X(\omega)$ – преобразование Фурье от сигнала $x(t)$, а $x(t)$ – отклик согласованного фильтра на входное воздействие $h(t)$. Следовательно, правую часть (10.2.18) можно также выразить через $X(\omega)$.

Подставив $X(e^{j\omega T})$ согласно (10.2.18) в (10.2.13), получаем желательное выражение для ОСШ в виде

$$\gamma_\infty = \left[\frac{TN_0}{2\pi}\int_{-\pi/T}^{\pi/T}\frac{d\omega}{\sum_{n=-\infty}^{\infty}|H(\omega+2\pi n/T)|^2}\right]^{-1}. \qquad (10.2.19)$$

Мы видим, что если сложенная спектральная характеристика $|H(\omega)|^2$ имеет нули, интеграл оказывается неограниченным, а ОСШ становится равным нулю. Другими словами, качество эквалайзера плохое всякий раз, когда сложенная спектральная характеристика проходит через нуль или имеет малое значение. Такое поведение возникает прежде всего потому, что эквалайзер, устраняя МСИ, увеличивает аддитивный шум. Например, если канал имеет нуль в своей частотной характеристике, линейный ЭНВП пытается это компенсировать введением неограниченного усиления на этой частоте. Но такая компенсация искажений в канале достигается ценой увеличения аддитивного шума. С другой стороны, идеальный канал, связанный с подходящим синтезом, который ведёт к отсутствию МСИ, будет иметь сложенный спектр, который удовлетворяет условию:

$$\sum_{n=-\infty}^{\infty}\left|H\left(\omega+\frac{2\pi n}{T}\right)\right|^2 = T, \quad |\omega| \le \frac{\pi}{T}. \qquad (10.2.20)$$

В этом случае ОСШ достигает максимального значения, а именно

$$\gamma_\infty = \frac{1}{N_0}. \qquad (10.2.21)$$

Эквалайзер ограниченной длины. Теперь обратим наше внимание на эквалайзер, имеющий $2K+1$ ячеек. Поскольку $c_j = 0$ для $|j| > K$ свёртка от $\{f_n\}$ и $\{c_n\}$ равно нулю вне области $-K \le n \le K+L-1$. Это значит, что $q_n = 0$ для $n < -K$ и $n > K+L-1$. При нормировке q_0 к единице, пиковое искажение равно

$$\mathscr{D}(\mathbf{c}) = \sum_{\substack{n=-K \\ n\neq 0}}^{K+L-1} |q_n| = \sum_{\substack{n=-K \\ n\neq 0}}^{K+L-1} \left| \sum_j c_j f_{n-j} \right|. \qquad (10.2.22)$$

Хотя эквалайзер имеет $2K+1$ регулируемых параметров, имеется $2K+L$ ненулевых значений откликов $\{q_n\}$. Следовательно, в общем невозможно полностью исключить МСИ на выходе эквалайзера. Здесь всегда имеется остаточная интерференция даже при использовании оптимальных коэффициентов. Проблема заключается в минимизации $\mathscr{D}(\mathbf{c})$ по коэффициентам $\{c_j\}$.

Лакки (1965) показал, что пиковое искажение, определяемое (10.2.22), является выпуклой функцией коэффициентов $\{c_j\}$. Это значит, что она обладает глобальным минимумом, а не относительным минимумом. Её минимизацию можно выполнить численно, например, методом скорейшего спуска. Немного больше можно сказать об общем решении этой проблемы минимизации. Однако, для одного частного, но важного случая, решение по минимизации $\mathscr{D}(\mathbf{c})$ известно. Это случай, когда искажение на выходе эквалайзера, определяемое как

$$D_0 = \frac{1}{|f_0|} \sum_{n=1}^{L} |f_n|, \qquad (10.2.23)$$

меньше единицы. Это условие эквивалентно наличия открытого глазка априори до выравнивания. Это значит, что МСИ не настолько тяжёлая, чтобы закрыть глазок. При этом условии пиковое искажение $\mathscr{D}(\mathbf{c})$ минимизируется выбором коэффициентов эквалайзера для обеспечения $q_n = 0$ для $1 \le |n| \le K$ и $q_0 = 1$. Другими словами, общее решение по минимизации $\mathscr{D}(\mathbf{c})$, когда $D_0 < 1$, является «нуль-форсированное» решение для $\{q_n\}$ в области $1 \le |n| \le K$. Однако величины $\{q_n\}$ для $K+1 \le n \le K+L-1$ в общем ненулевые. Эти ненулевые величины образуют остаточную МСИ на выходе эквалайзера.

10.2.2. Критерий минимума среднеквадратичной ошибки (СКО)

При использовании критерия минимума СКО, взвешивающие коэффициенты ячеек $\{c_j\}$ эквалайзера подстраиваются так, чтобы минимизировать средний квадрат ошибки

$$\varepsilon_k = I_k - \hat{I}_k, \qquad (10.2.24)$$

где I_k – информационный символ, переданный на k-ом сигнальном интервале, а $\hat{I}_k$ – оценка этого символа на выходе эквалайзера, определяемая (10.2.1). Если информационные символы $\{I_k\}$ комплексные, то показатель качества при СКО критерия, обозначаемый J, определяется так

$$J = E|\varepsilon_k|^2 = E|I_k - \hat{I}_k|^2. \qquad (10.2.25)$$

С другой стороны, когда информационные символы вещественные, показатель качества просто равен квадрату вещественной величины ε_k. В любом случае, J является

квадратичной функцией коэффициентов эквалайзера $\{c_j\}$. При дальнейшем обсуждении мы рассмотрим минимизацию комплексной формы, даваемой (10.2.25).

Эквалайзер неограниченной длины. Сначала определим взвешивающие коэффициенты ячеек, которые минимизируют J, когда эквалайзер имеет неограниченное число ячеек. В этом случае, оценка I_k определяется так

$$\hat{I}_k = \sum_{j=-\infty}^{\infty} c_j \upsilon_{k-j} . \qquad (10.2.26)$$

Подстановка (10.2.26) в выражение для J, определяемая (10.2.25), и расширение результата приводит к квадратичной функции от коэффициентов $\{c_j\}$. Эту функцию можно легко минимизировать по $\{c_j\}$ посредством решения системы (неограниченной) линейных уравнений для $\{c_j\}$. Альтернативно, систему линейных уравнений можно получить путём использования принципа ортогональности при среднеквадратичном оценивании. Это значит, мы выбираем коэффициенты $\{c_j\}$ такие, что ошибка ε_k ортогональна сигнальной последовательности υ^*_{k-l} для $-\infty < l < \infty$. То есть

$$E(\varepsilon_k \upsilon^*_{k-l}) = 0, \quad -\infty < l < \infty \qquad (10.2.27)$$

Подстановка ε_k в (10.2.27) даёт

$$E\left[\left(I_k - \sum_{j=-\infty}^{\infty} c_j \upsilon_{k-j}\right)\upsilon^*_{k-l}\right] = 0$$

или, что эквивалентно,

$$\sum_{j=-\infty}^{\infty} c_j E(\upsilon_{k-j}\upsilon^*_{k-l}) = E(I_k \upsilon^*_{k-l}), \quad -\infty < l < \infty . \qquad (10.2.28)$$

Чтобы вычислить моменты в (10.2.28), мы используем выражение для υ_k, даваемое (10.1.16). Таким образом, получим

$$E(\upsilon_{k-j}\upsilon^*_{k-l}) = \sum_{n=0}^{L} f^*_n f_{n+l-j} - N_0\delta_{lj} = \begin{cases} x_{l-j} + N_0\delta_{lj} & (|l-j| \le L) \\ 0 & (\text{при других } l, j) \end{cases} \qquad (10.2.29)$$

и

$$E(I_k \upsilon^*_{k-l}) = \begin{cases} f^*_{-l} & (-L \le l \le 0) \\ 0 & (\text{при других } l) \end{cases} \qquad (10.2.30)$$

Теперь, если подставим (10.2.29) и (10.2.30) в (10.2.28) и возьмём z-преобразование от обеих частей результирующего уравнения, мы находим

$$C(z)\left[F(z)F^*(z^{-1}) + N_0\right] = F^*(z^{-1}) . \qquad (10.2.31)$$

Следовательно, передаточная функция эквалайзера, основанного на критерии минимума СКО, равна

$$C(z) = \frac{F^*(z^{-1})}{F(z)F^*(z^{-1}) + N_0} . \qquad (10.2.32)$$

Если обеляющий фильтр включён в $C(z)$, мы получаем эквивалентный эквалайзер с передаточной функцией

$$C'(z) = \frac{1}{F(z)F^*(z^{-1}) + N_0} = \frac{1}{X(z) + N_0} . \qquad (10.2.33)$$

Видим, что единственная разница между этим выражением для $C'(z)$ и тем, которое базируется на критерии пикового искажения – это спектральная плотность шума N_0, которая появилась в (10.2.33). Если N_0 очень мало по сравнению с сигналом, коэффициенты, которые минимизируют пиковые искажения $\mathcal{D}(\mathbf{c})$ приближённо равны коэффициентам, которые минимизируют по СКО показатель качества J. Это значит, что в пределе, когда $N_0 \to 0$, два критерия дают одинаковое решение для взвешивающих коэффициентов. Следовательно, когда $N_0 = 0$, минимизация СКО ведёт к полному исключению МСИ. С другой стороны, это не так, когда $N_0 \neq 0$. В общем, когда $N_0 \neq 0$, оба критерия дают остаточное МСИ и аддитивный шум на выходе эквалайзера.

Меру остаточного МСИ и аддитивного шума на выходе эквалайзера можно получить расчётом минимальной величины J, обозначаемую $J_{\min}$, когда передаточная функция $C(z)$ эквалайзера определена (10.2.32). Поскольку $J = E|\varepsilon_k|^2 = E\left(\varepsilon_k I_k^*\right) - E\left(\varepsilon_k \hat{I}_k^*\right)$ и поскольку $E\left(\varepsilon_k \hat{I}_k^*\right) = 0$ с учётом условия ортогональности (10.2.27), следует

$$J_{\min} = E(\varepsilon_k I_k^*) = E|I_k|^2 - \sum_{j=-\infty}^{\infty} c_j E(\upsilon_{k-j} I_k^*) = 1 - \sum_{j=-\infty}^{\infty} c_j f_{-j} . \qquad (10.2.34)$$

Эта частная форма для $J_{\min}$ не очень информативна. Больше понимания зависимости качества эквалайзера от канальных характеристик можно получить, если суммы в (10.2.34) преобразовать в частотную область. Это можно выполнить, заметив, что сумма в (10.2.34) является свёрткой $\{c_j\}$ и $\{f_j\}$, вычисленной при нулевом сдвиге. Так, если через $\{b_k\}$ обозначить свёртку этих последовательностей, то сумма в (10.2.34) просто равна b_0. Поскольку z-преобразование последовательности $\{b_k\}$ равно

$$B(z) = C(z)F(z) = \frac{F(z)F^*(z^{-1})}{F(z)F^*(z^{-1}) + N_0} = \frac{X(z)}{X(z) + N_0}, \qquad (10.2.35)$$

то слагаемое b_0 равно

$$b_0 = \frac{1}{2\pi j}\oint \frac{B(z)}{z} dz = \frac{1}{2\pi j}\oint \frac{X(z)}{z[X(z) + N_0]} dz . \qquad (10.2.36)$$

Контурный интеграл в (10.2.36) можно преобразовать в эквивалентный линейный интеграл путём замены переменной $z = e^{j\omega T}$. В результате этой замены получаем

$$b_0 = \frac{T}{2\pi}\int_{-\pi/T}^{\pi/T} \frac{X(e^{j\omega T})}{X(e^{j\omega T}) + N_0} d\omega . \qquad (10.2.37)$$

Наконец, подставив (10.2.37) в сумму (10.2.34), получаем желательное выражение для минимума СКО в виде

$$\begin{aligned} J_{\min} &= 1 - \frac{T}{2\pi}\int_{-\pi/T}^{\pi/T} \frac{X(e^{j\omega T})}{X(e^{j\omega T}) + N_0} d\omega = \frac{T}{2\pi}\int_{-\pi/T}^{\pi/T} \frac{N_0}{X(e^{j\omega T}) + N_0} d\omega = \\ &= \frac{T}{2\pi}\int_{-\pi/T}^{\pi/T} \frac{N_0}{T^{-1}\sum_{n=-\infty}^{\infty} |H(\omega + 2\pi n/T)|^2 + N_0} d\omega . \end{aligned} \qquad (10.2.38)$$

В отсутствие МСИ $X(e^{j\omega T}) = 1$ и, следовательно,

$$J_{\min} = \frac{N_0}{1 + N_0} . \qquad (10.2.39)$$

Видим, что $0 \le J_{\min} \le 1$. Далее, соотношение между выходным (нормированного по энергии сигнала) ОСШ γ_∞ и $J_{\min}$ выглядит так

$$\gamma_\infty = \frac{1 - J_{mn}}{J_{\min}}. \qquad (10.2.40)$$

Более существенно то, что соотношение между γ_∞ и $J_{\min}$ также имеет силу, когда имеется остаточная МСИ в дополнении к шуму на выходе эквалайзера.

Эквалайзер ограниченной длины. Теперь вернём наше внимание к случаю, когда длительность импульсной характеристики трансверсального эквалайзера простирается на ограниченном временном интервале, т.е. эквалайзер имеет конечную память или ограниченную длину. Выход эквалайзера на k-м сигнальном интервале равен

$$\hat{I}_k = \sum_{j=-K}^{K} c_j \upsilon_{k-j}. \qquad (10.2.41)$$

СКО эквалайзера с $2K+1$ ячейками, обозначаемый $J(K)$, равен

$$J(K) = E\left|I_k - \hat{I}_k\right|^2 = E\left|I_k - \sum_{j=-K}^{K} c_j \upsilon_{k-j}\right|^2. \qquad (10.2.42)$$

Минимизация $J(K)$ по взвешивающим коэффициентам ячеек $\{c_j\}$ или, что эквивалентно, требуя, чтобы ошибка $\varepsilon_k = I_k - \hat{I}_k$ была бы ортогональна сигнальным отсчётам υ^*_{j-l}, $|l| \le K$, приводит к следующей системе уравнений:

$$\sum_{j=-K}^{K} c_j \Gamma_{lj} = \xi_l, \quad l = -K, \ldots, -1, 0, 1, \ldots, K, \qquad (10.2.43)$$

где

$$\Gamma_{lj} = \begin{cases} x_{l-j} + N_0 \delta_{lj} & (|l-j| \le L), \\ 0 & (\text{при других } l, j), \end{cases} \qquad (10.2.44)$$

$$\xi_l = \begin{cases} f^*_{-l} & (-L \le l \le 0), \\ 0 & (\text{при других } l). \end{cases} \qquad (10.2.45)$$

Удобно выразить систему линейных уравнений в матричной форме, т.е.

$$\mathbf{\Gamma C} = \boldsymbol{\xi}, \qquad (10.2.46)$$

где $\mathbf{C}$ означает вектор столбец $2K+1$ взвешивающих значений кодовых ячеек, $\mathbf{\Gamma}$ означает $(2K+1)\times(2K+1)$ матрицу ковариаций Эрмита с элементами Γ_{ij}; а $\boldsymbol{\xi}$ – $(2K+1)$ мерный вектор столбец с элементами ξ_l. Решение (10.2.46) можно записать в виде

$$\mathbf{C}_{\text{опт}} = \mathbf{\Gamma}^{-1}\boldsymbol{\xi}. \qquad (10.2.47)$$

Таким образом, решение для $\mathbf{C}_{\text{опт}}$ включает в себя обращение матрицы $\mathbf{\Gamma}$. Оптимальные взвешивающие коэффициенты ячеек, даваемые (10.2.47), минимизируют показатель качества $J(K)$, что приводит к минимальной величине $J(K)$

$$J_{\min}(K) = 1 - \sum_{j=-K}^{0} c_j f_{-j} = 1 - \boldsymbol{\xi}^{T*}\mathbf{\Gamma}^{-1}\boldsymbol{\xi}, \qquad (10.2.48)$$

где $\boldsymbol{\xi}^T$ определяет транспонированный вектор столбец $\boldsymbol{\xi}$. $J_{\min}(K)$ можно использовать в (10.2.40) для вычисления ОСШ линейного эквалайзера с $2K+1$ коэффициентами ячеек.

10.2.3. Характеристики качества эквалайзера по минимуму СКО

В этом разделе мы рассмотрим характеристики качества линейного эквалайзера, который оптимизирован при использовании критерия минимума СКО. Как минимум СКО,

так и вероятность ошибки рассматриваются как меры качества для некоторых специфических каналов. Мы начнем с вычисления минимума СКО $J_{\min}$ и выходного ОСШ γ_∞ для двух специфических каналов. Затем мы рассмотрим оценку для вероятности ошибки.

Пример 10.2.1. Сначала мы рассмотрим эквивалентную модель канала с дискретным временем, который состоит из двух компонент f_0 и $f_1$0, которые нормированы так $|f_0|^2+|f_1|^2=1$. Имеем

$$F(z)=f_0+f_1z^{-1} \tag{10.2.49}$$

и

$$X(z)=f_0f_1^*z+1+f_0^*f_1z^{-1} \tag{10.2.50}$$

Соответствующая частотная характеристика равна

$$X(\mathrm{e}^{j\omega T})=f_0f_1\,\mathrm{e}^{j\omega T}+1+f_0^*f_1\mathrm{e}^{-j\omega T}=1+2|f_0||f_1|\cos(\omega T+\theta), \tag{10.2.51}$$

где θ – угол $f_0f_1^*$. Заметим, что эта канальная характеристика содержит нуль на частоте $\omega=\pi/T$, когда $f_0=f_1=\sqrt{\frac{1}{2}}$.

Линейный эквалайзер с неограниченным числом ячеек, построенный на основе критерия минимума СКО, будет иметь минимум СКО, определенный (10.2.38). Вычисление интеграла (10.2.38) при $X\left(\mathrm{e}^{j\omega T}\right)$, определяемом (10.2.51), приводит к результату

$$J_{\min}=\frac{N_0}{\sqrt{N_0^2+2N_0\left(|f_0|^2+|f_1|^2\right)+\left(|f_0|^2-|f_1|^2\right)^2}}=\frac{N_0}{\sqrt{N_0^2+2N_0+\left(|f_0|^2-|f_1|^2\right)^2}}. \tag{10.2.52}$$

Рассмотрим частный случай, когда $f_0=f_1=\sqrt{\frac{1}{2}}$. Тогда минимум СКО $J_{\min}=N_0\Big/\sqrt{N_0^2+2N_0}$ а соответствующее выходное ОСШ равно

$$\gamma_\infty=\sqrt{1+\frac{2}{N_0}}-1\approx\left(\frac{2}{N_0}\right)^{1/2},\quad N_0\ll1. \tag{10.2.53}$$

Этот результат можно сравнить с выходным ОСШ $1/N_0$, полученным для случая отсутствия МСИ. В этом канале возникает незначительная потеря в ОСШ.

Пример 10.2.2. В качестве второго примера рассмотрим показательную затухающую характеристику канала в виде $f_k=\sqrt{1-a^2}a^k$, $k=0,1,\ldots$ где $a<1$. Преобразование Фурье этой последовательности

$$X(\mathrm{e}^{j\omega T})=\frac{1-a^2}{1+a^2-2a\cos\omega T} \tag{10.2.54}$$

является функция с минимумом при $\omega=\pi/T$.

Выходное ОСШ для этого канала

$$\gamma_\infty=\left(\sqrt{1+2N_0\frac{1+a^2}{1-a^2}+N_0^2}-1\right)^{-1}\approx\frac{1-a^2}{(1+a^2)N_0},\quad N_0\ll1. \tag{10.2.55}$$

Следовательно, потеря в ОСШ из-за интерференции равна $10\lg\left((1-a^2)/(1+a^2)\right)$

Вероятность ошибки в линейном эквалайзере по минимуму СКО. Выше мы обсуждали качество линейного эквалайзера через минимально достижимое СКО $J_{\min}$ и выходное ОСШ γ, связанное с $J_{\min}$ формулой (10.2.40). К сожалению нет простого отношения между этими характеристиками и вероятностью ошибки. Причина в том, что линейный эквалайзер по минимуму СКО содержит некоторую остаточную МСИ на своем выходе. Эта ситуация не похожа на ту, которая имеет место в ЭНВП неограниченной длины, в котором нет остаточной интерференции, но только гауссовский шум. Остаточная интерференция на выходе эквалайзера по минимуму СКО не удовлетворительно характеризуется, как аддитивный гаусовский шум, и, следовательно, выходное ОСШ нельзя легко преобразовать в эквивалентную вероятность ошибки.

Одним из подходов к вычислению вероятности ошибки является жестко форсированный метод, который дает точный результат. Для иллюстрации этого метода рассмотрим АМ сигнал, в котором информационные символы отображаются набором значений амплитуд $2n-M$, $n=1, 2, ..., M$ с равной вероятностью. Теперь рассмотрим решение о символе I_n. Оценка I_n равна

$$\hat{I}_n = q_0 I_n + \sum_{k\neq n} I_k q_{n-k} + \sum_{j=-K}^{K} c_j \eta_{n-j}, \qquad (10.2.56)$$

где $\{q_n\}$ представляет свертку импульсной характеристики эквалайзера и эквивалентного канала, т.е.

$$q_n = \sum_{k=-K}^{K} c_k f_{n-k}, \qquad (10.2.57)$$

а входной сигнал на эквалайзер равен

$$\upsilon_k = \sum_{j=0}^{L} f_j I_{k-j} + \eta_k. \qquad (10.2.58)$$

Первое слагаемое суммы в правой части (10.2.56) – это желательный символ, остальные слагаемые суммы – это МСИ, а последнее слагаемое является гауссовским шумом. Дисперсия шума

$$\sigma_n^2 = N_0 \sum_{j=-K}^{K} c_j^2. \qquad (10.2.59)$$

Для эквалайзера с $2K+1$ ячейками и канальным откликом, который простирается на $L+1$ символов, число символов, участвующих в МСИ, равно $2K+1$. Определим

$$\mathcal{D} = \sum_{k\neq n} I_k q_{n-k}. \qquad (10.2.60)$$

Для частной последовательности из $2K+1$ информационных символов, скажем последовательности $\mathbf{I}_J$, слагаемое МСИ $\mathcal{D} = D_J$ фиксировано. Вероятность ошибки для фиксированной D_J равна

$$P_M(D_J) = 2\frac{(M-1)}{M} P(N + D_J > q_0) = \frac{2(M-1)}{M} Q\left(\sqrt{\frac{(q_0 - D_J)^2}{\sigma_n^2}}\right), \qquad (10.2.61)$$

где N означает слагаемое аддитивного шума. Средняя вероятность ошибки получается путем усреднения $P_M(D_J)$ по всем возможным последовательностям $\mathbf{I}_J$. Это дает

$$P_M = \sum_{\mathbf{I}_J} P_M(D_J) P(\mathbf{I}_J) = \frac{2(M-1)}{M} \sum_{\mathbf{I}_J} Q\left(\sqrt{\frac{(q_0 - D_J)^2}{\sigma_n^2}}\right) P(\mathbf{I}_J). \qquad (10.2.62)$$

Когда все последовательности равновероятны, то

$$P(\mathbf{I}_J) = \frac{1}{M^{2K+L}}. \qquad (10.2.63)$$

Условные вероятности ошибки $P_M(D_J)$ главным образом определяются последовательностью, которая содержит наибольшее значение D_J, т.е. тогда, когда $I_n = \pm(M-1)$, а знаки информационных символов определяют знаки соответствующих $\{q_n\}$ При этом

$$D_J^* = (M-1)\sum_{k\neq 0}|q_k|$$

и

$$P_M(D_J^*) = \frac{2(M-1)}{M} Q\left(\sqrt{\frac{q_0^2}{\sigma_n^2}\left(1-\frac{M-1}{q_0}\sum_{k\neq 0}|q_k|\right)^2}\right). \qquad (10.2.64)$$

Таким образом, верхняя граница для средней вероятности ошибки для равновероятных последовательностей символов определяется так:

$$P_M \le P_M(D_J^*) \qquad (10.2.65)$$

Если вычисления вероятности ошибки по точной формуле (10.2.62) представляется слишком обремененным и слишком много теряется времени из-за большого числа слагаемых суммы и если верхняя граница слишком свободная, то можно прибегнуть к одному из многих различных приближенных методов, которые были разработаны и которые, как известно, дают плотные границы для P_M. Обсуждение этих различных подходов уведет нас слишком далеко в сторону. Интересующемуся читателю рекомендуются статьи Зальцберга (1968), Луганани (1969), Хо и Е (1970), Шимбо и Целебилера (1971), Главе (1972), Яо (1972) и Яо и Тобина (1976).

В качестве иллюстрации ограничения качества линейного эквалайзера в присутствии существенной МСИ, рассмотрим на рис.10.2.4 вероятность ошибки для двоичных (противоположных) сигналов, полученных моделированием по методу Монте-Карло для трех каналов с дискретным временем, показанным на рис.10.2.5. С целью сравнения на рис.10.2.4 также показана характеристика, полученная для канала без МСИ. Характеристика эквивалентного канала с дискретным временем, показанная на рис.10.2.5(а), типична для телефонного канала хорошего качества.

В противоположность, характеристики эквивалентных каналов с дискретным временем, показанные на рис.10.2.5(*b*) и (*c*), приводят к существенной МСИ. Спектральные характеристики $\left|X\left(e^{j\omega}\right)\right|$ для трёх каналов, иллюстрируемые на рис.10.2.6, ясно показывают, что канал по рис.10.2.5(*c*) имеет наихудшую спектральную характеристику. Следовательно, качество линейного эквалайзера для этого канала наихудшее из трёх рассматриваемых случаев. Следующий по качеству канал показанный на рис.10.2.5(*b*), и наконец, наилучшее качество получается для канала, показанного на рис.10.2.5(*a*). Действительно, он проигрывает каналу без МСИ по ОСШ примерно на 3 дБ.

Одно заключение следует из результатов для выходного ОСШ γ_∞ и ограниченных результатов вероятности ошибки, иллюстрируемых рис.10.2.4; именно, что линейный эквалайзер приводит к хорошему качеству для таких каналов, как телефонные линии, у которых спектральные характеристики хорошо себя ведут и не содержат нулей. С другой стороны, линейный эквалайзер не годится как компенсатор МСИ для каналов со спектральными нулями, которые встречаются в радиосвязи.

Эквалайзер с обратной связью, описываемый в разделе 10.3, представляется эффективным решением проблемы вычислительной сложности.

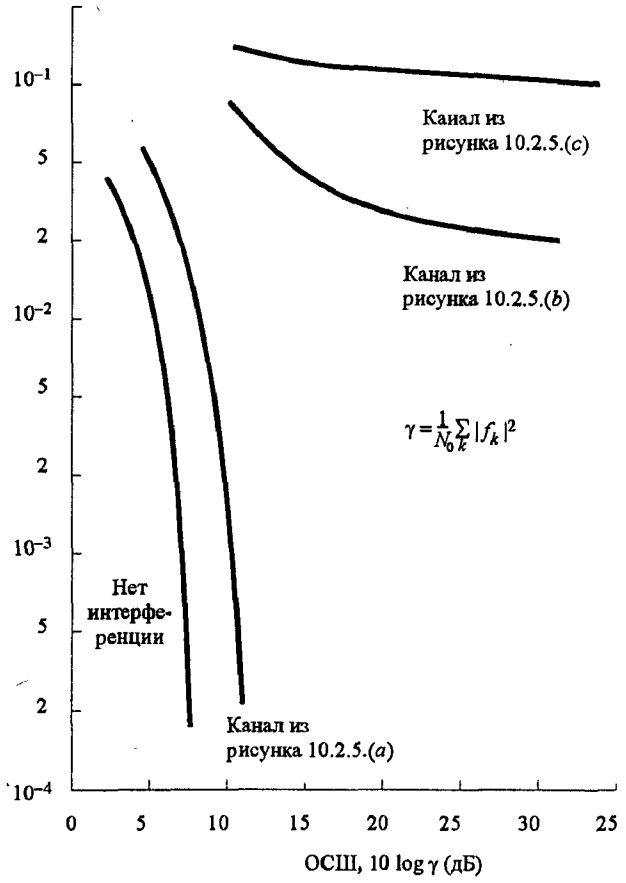


Рис. 10.2.4. Вероятность ошибки с использованием MSE эквалайзера

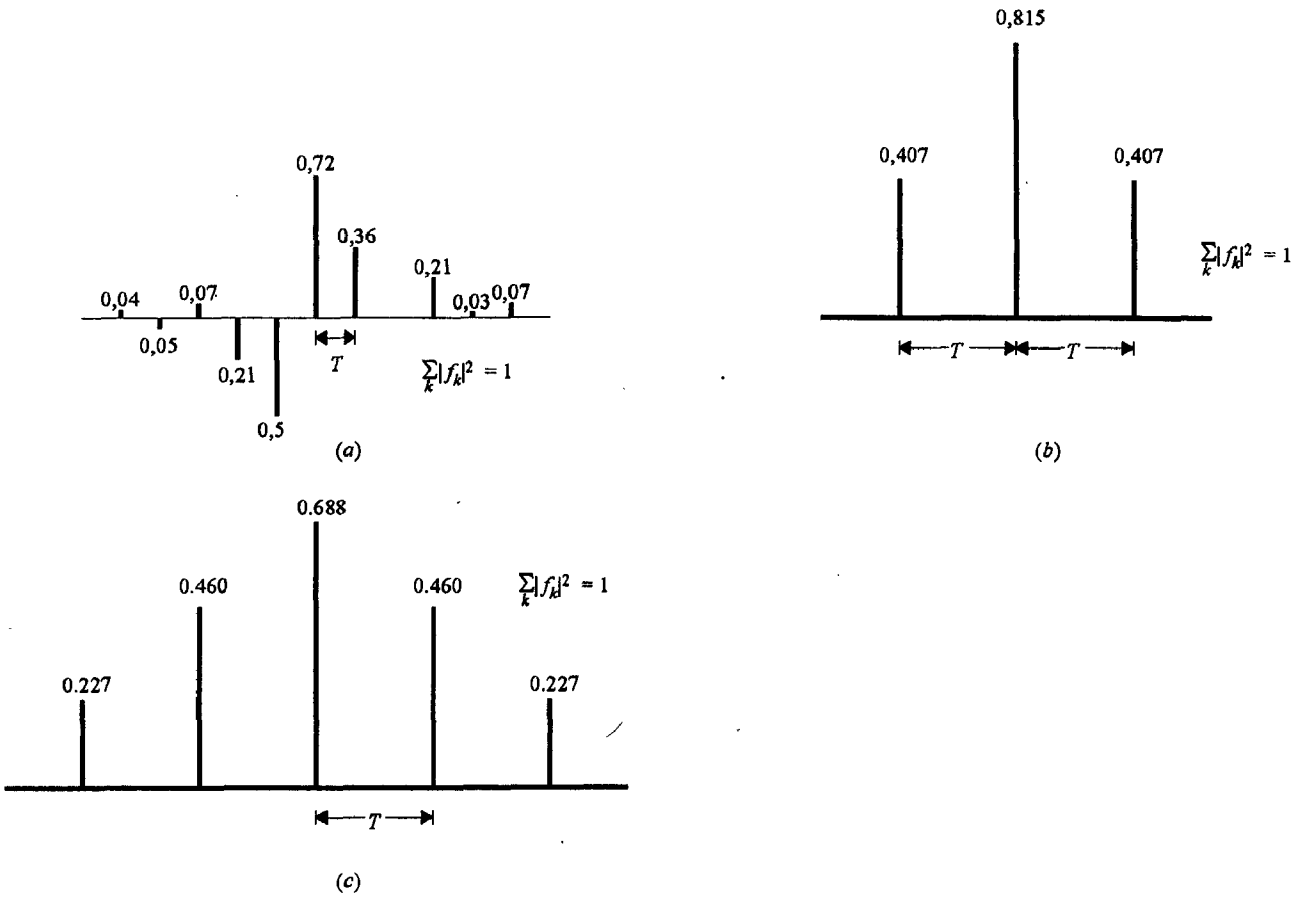


Рис. 10.2.5. Три характеристики дискретного во времени канала

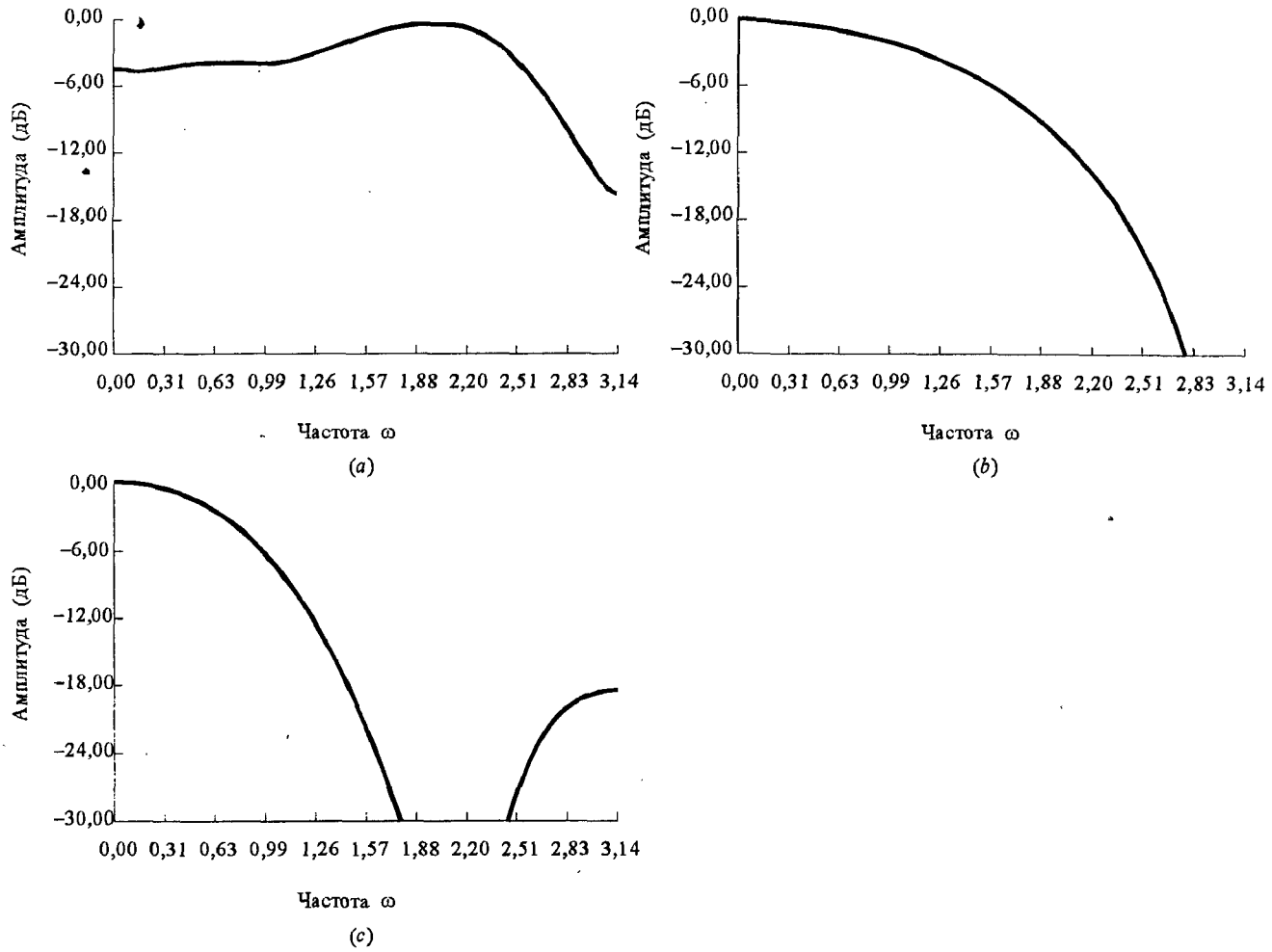


Рис.10.2.6. АЧХ для каналов, показанных на рис. 10.2.5 (*a*), (*b*) и (*c*) соответственно

10.2.4. Дробные эквалайзеры

В структурах линейных эквалайзеров, которые были описаны в предыдущем разделе, ячейки эквалайзера были размещены так, чтобы задержка между ними была обратна скорости подачи символов $1/T$. Такое расположение ячеек оптимально, если до эквалайзера имеется фильтр, согласованный с искажённым каналом переданного импульса. Если канальные характеристики неизвестны, фильтр приёмника обычно согласуется с переданным сигнальным импульсом, а интервал стробирования оптимизируется для этого субоптимального фильтра. В общем, этот подход ведет к качеству эквалайзера, который очень чувствителен к выбору шага временной дискретизации.

Ограничение в скорости передачи символов для эквалайзера более легко просматривается в частотной области. Из (9.2.5) спектр сигнала на входе эквалайзера можно выразить так

$$Y_T(f)=\frac{1}{T}\sum_n X\left(f-\frac{n}{T}\right)e^{j2\pi(f-n/T)\tau_0}, \qquad (10.2.66)$$

где $Y_T(f)$ – сложенный или виртуальный спектр, причём частота сворачивания равна $1/2T$. Заметим, что спектр принимаемого сигнала зависит от выбора задержки при стробировании отсчетов τ_0. Спектр сигнала на выходе эквалайзера равен $C_T(f)Y_T(f)$, где

$$C_T(f) = \sum_{k=-K}^{K} c_k e^{-j2\pi fkT} . \qquad (10.2.67)$$

Из этих соотношений ясно, что эквалайзер, который стробирует со скоростью символов может только компенсировать частотную характеристику принимаемого сигнала. Он не может компенсировать канальное искажение, присущее $X(f)e^{j2\pi ft_0}$.

В противовес эквалайзеру, который использует стробирование со скоростью передачи символов, *дробный эквалайзер* (ДЭ) основывается на стробировании приходящего сигнала по крайней мере со скоростью Найквиста. Например, если передаваемый сигнал состоит из импульсов, имеющих спектр приподнятого косинуса с коэффициентом ската β, его спектр расширяется до $F_{\max} = (1+\beta)/2T$. Этот сигнал можно стробировать в приёмнике со скоростью

$$2F_{\max} = \frac{1+\beta}{T} \qquad (10.2.68)$$

и затем пропустить отсчёты через эквалайзер с шагом размещения ячеек $T/(1+\beta)$. Например, если $\beta = 1$, имеем ДЭ с шагом $\frac{1}{2}T$. Если $\beta = 0{,}5$, имеем ДЭ с шагом $\frac{2}{3}T$ и так далее. В общей реализации цифровой дробно размещённый эквалайзер имеет шаг MT/N, где M и N – целые числа и $N > M$. Обычно ДЭ с шагом $\frac{1}{2}T$ используется во многих приложениях.

Поскольку частотная характеристика ДЭ равна

$$C_{T'}(f) = \sum_{k=-K}^{K} c_k e^{-j2\pi fkT'} , \qquad (10.2.69)$$

где $T' = MT/N$, то следует, что $C_{T'}(f)$ может выравнивать спектр принимаемого сигнала за пределами частоты Найквиста $f = 1/2T$ до $f = (1+\beta)/T = N/MT$. Выровненный спектр равен

$$\begin{aligned} C_{T'}(f)Y_{T'}(f) &= C_{T'}(f)\sum_n X\left(f - \frac{n}{T'}\right)e^{j2\pi(f-n/T')\tau_0} = \\ &= C_{T'}(f)\sum_n X\left(f - \frac{nN}{MT}\right)e^{j2\pi(f-nN/MT)\tau_0}. \end{aligned} \qquad (10.2.70)$$

Поскольку $X(f) = 0$ для $|f| > N/MT$, то (10.2.70) можно выразить так

$$C_{T'}(f)Y_{T'}(f) = C_{T'}(f)X(f)e^{j2\pi ft_0}, \quad |f| \le \frac{1}{2T'}. \qquad (10.2.71)$$

Таким образом, мы видим, что ДЭ компенсировал искажения канала в принимаемом сигнале до влияния, обусловленного скоростью стробирования. Другими словами, $C_{T'}(f)$ может компенсировать искажение при любом фазовом сдвиге.

Выход ДЭ стробируется со скоростью передачи символов $1/T$ и имеет спектр

$$\sum_k C_{T'}\left(f - \frac{k}{T}\right)X\left(f - \frac{k}{T}\right)e^{j2\pi(f-k/T)\tau_0} . \qquad (10.2.72)$$

Фактически оптимальный ДЭ эквивалентен оптимальному линейному приемнику, состоящего из согласованного фильтра, за которым следует эквалайзер со стробированием со скоростью передачи символов.

Теперь рассмотрим регулирование (настройку) коэффициентов ячеек в ДЭ. Вход ДЭ можно выразить так

$$y\left(\frac{kMT}{N}\right)=\sum_n I_n x\left(\frac{kMT}{N}-nT\right)+\nu\left(\frac{kMT}{N}\right). \qquad (10.2.73)$$

На каждом символьном интервале ДЭ выдает выход в виде

$$\hat{I}_k=\sum_{n=-K}^{K} c_n y\left(kT-\frac{nMT}{N}\right), \qquad (10.2.74)$$

где коэффициенты эквалайзера выбираются так, чтобы минимизировать СКО. Эта оптимизация ведет к системе линейных уравнений для коэффициентов эквалайзера, которая имеет решение

$$\mathbf{C}_{\text{опт}}=\mathbf{A}^{-1}\boldsymbol{\alpha}, \qquad (10.2.75)$$

где $\mathbf{A}$ – матрица ковариации входных данных, а $\boldsymbol{\alpha}$ – вектор взаимных корреляций. Эти уравнения идентичны по форме тем, которые получены для эквалайзера с шагом T, но здесь имеется некоторая тонкая разница. Одна заключается в том, что матрица $\mathbf{A}$ является эрмитовой, но не тёплицевой. Дополнительно, $\mathbf{A}$ проявляет периодичность, которая присуща циклостационарному процессу, как показал Куреши (1985). Как результат дробного стробирования, некоторые из собственных значений $\mathbf{A}$ близки к нулю. Лонгом и др. (1988 a, b) были предприняты усилия для использования этого свойства при настройке коэффициентов эквалайзера.

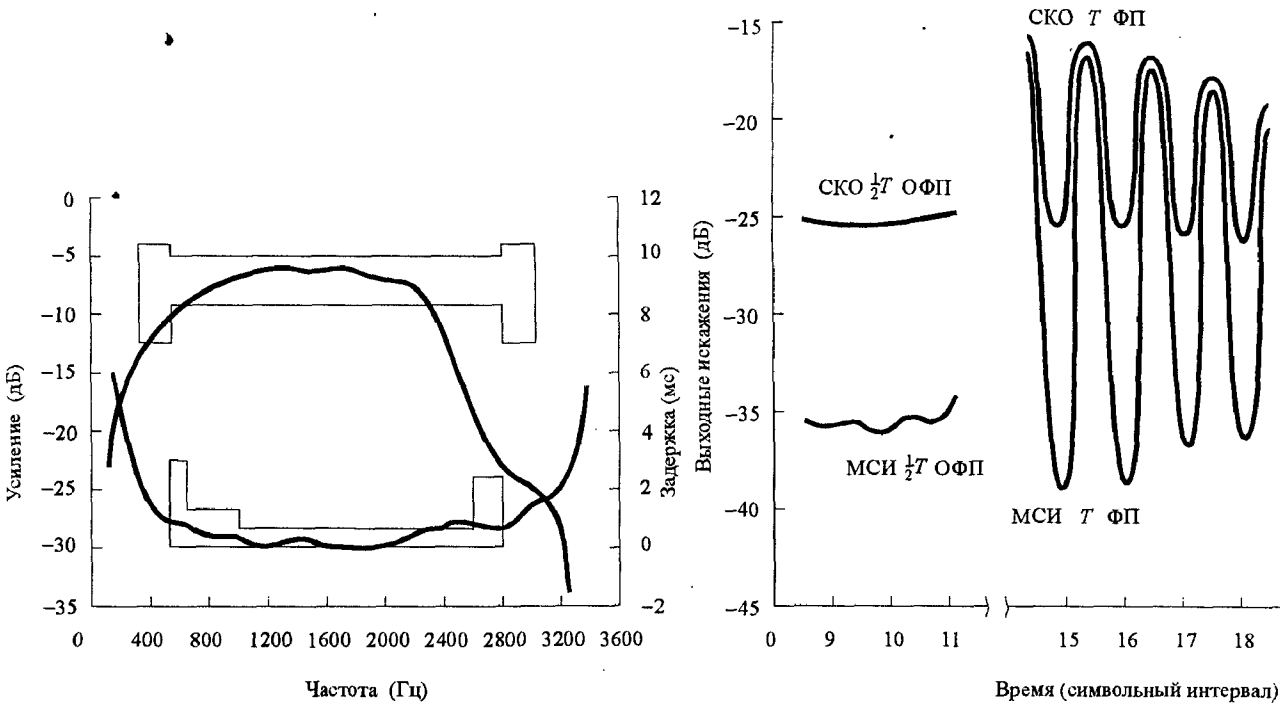
Анализ качества ДЭ включая их свойства сходимости, дан в статье Унгербоека (1976). Результаты моделирования, показывающее эффективность ДЭ относительно эквалайзера, осуществляющего стробирование со скоростью передачи символов, также даны в статьях Куреши и Форни (1977) и Гитлина и Вайнштейна (1981). Мы процитируем два примера из этих статей. Сначала рис.10.2.7 иллюстрирует качество эквалайзера с шагом T и ДЭ с шагом $\frac{1}{2}T$ для канала с большими амплитудными искажениями, а его характеристики также даны на рисунке. Эквалайзеру, стробирующему со скоростью передачи символов, предшествует фильтр, согласованный с переданным импульсом, спектр которого (квадратный корень) имеет вид приподнятого косинуса с 20% коэффициентом ската $\beta=0{,}2$. ДЭ не имеет предшествующей фильтрации. Скорость передачи символов была 2400 символов/с, а при модуляции используется КАМ. ОСШ принятого сигнала было 30 Дб. Оба эквалайзера имели 31 ячейку. Таким образом ДЭ занимал половину временного интервала эквалайзера, стробирующего с тактом передачи символов. Однако ДЭ, в отличие от сравниваемого с ним эквалайзера, не был оптимизирован по наилучшему шагу стробирования. Далее, ДЭ оказался не чувствителен к фазовым отклонениям, как показано на рис.10.2.7.

Аналогичные результаты были получены Гитлиным и Вайнштейном. Для канала с плохой характеристикой задержки огибающей, качество (ОСШ) T эквалайзера и $\frac{1}{2}T$ ДЭ иллюстрируется на рис.10.2.8.

В этом случае, оба эквалайзера имеют одинаковый интервал обработки. T эквалайзер имеет 24 ячейки, в то время как ДЭ имеет 48 ячеек. Скорость передачи символов была 2400 символов/с, а битовая скорость передачи 9600 бит/с с 16 КАМ модуляцией. Сигнальный импульс имеет спектр приподнятого косинуса с $\beta=0{,}12$. Заметим снова, что ДЭ превосходит Т эквалайзер на несколько децибел, даже если последний был настроен для оптимального стробирования. Результаты этих двух статей ясно демонстрируют превосходящее качество, достигнутое посредством ДЭ.

10.3. ВЫРАВНИВАНИЕ С ОБРАТНОЙ СВЯЗЬЮ ПО РЕШЕНИЮ

Эквалайзер с обратной связью по решению (ЭОСР), представленный на рис.10.3.1, состоит из двух фильтров, фильтра прямой и фильтра обратной связи.



(*a*) Канал с большими амплитудными искажениями на краях диапазона (*b*) Качество эквалайзера

Рис. 10.2.7. Характеристики T- и $\frac{1}{2}T$-эквалайзера для скорости 2400 символов в секунду (ОФП означает отсутствие фильтра приёмника ФП) [*Qureshi* и *Forney* (1977), © 1977 *IEEE*]

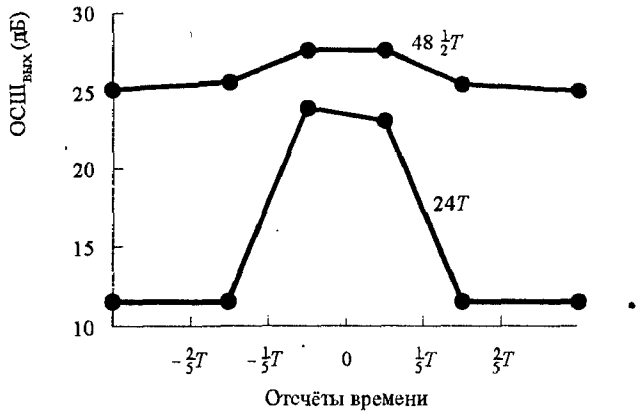


Рис. 10.2.8. Характеристики T- и $\frac{1}{2}T$-эквалайзера для скорости 2400 символов в секунду при модуляции 16-КАМ в канале с плохой характеристикой ГВП [*Gitlin* и *Weinstein* (1981), © 1981 *AT & T*]

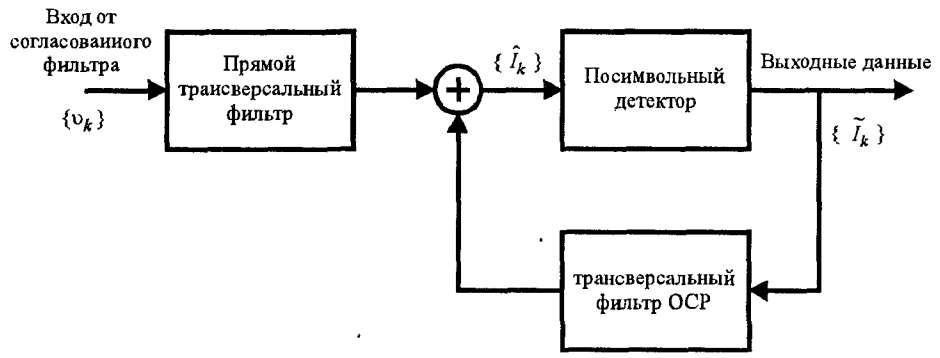


Рис. 10.3.1. Структура эквалайзера с обратной связью по решению (ОСР)

Оба фильтра имеют ячейки, размещённые на символьном интервале T. Входом для прямого фильтра является принимаемая символьная последовательность $\{\upsilon_k\}$. С учётом этого прямой фильтр идентичен линейному трансверсальному эквалайзеру, описанному в разделе 10.2. Фильтр обратной связи имеет на входе последовательность решений по предшествующим продетектированным символам. Функционально фильтр обратной связи используется для устранения в предстоящей оценке части МСИ, вызванной предыдущими продетектированными символами.

10.3.1. Оптимизация коэффициентов

Из описанного выше следует, что выход эквалайзера можно выразить так

$$\hat{I}_k = \sum_{j=-K_1}^{0} c_j v_{k-j} + \sum_{j=1}^{K_2} c_j \tilde{I}_{k-j}, \qquad (10.3.1)$$

где $\hat{I}_k$ – оценка k-го информационного символа, $\{c_j\}$ – коэффициенты ячеек фильтра, а $\tilde{I}_{k-1},\ldots,\tilde{I}_{k-K_2}$ – ранее продетектированные символы. Предполагается, что эквалайзер имеет (K_1+1) ячеек в своем блоке прямой связи и K_2 ячеек в блоке с обратной связью. Заметим, что этот эквалайзер нелинеен, поскольку фильтр обратной связи управляется предшествующими продетектированными символами $\{\tilde{I}_k\}$.

Как критерий минимума пиковых искажений, так и критерий минимума СКО приводит к математически разрешаемой трактуемости оптимизации коэффициентов эквалайзера, как следует из статей Джорджа и других (1971), Прайса (1972), Сальца (1973) и Прокиса (1975). Поскольку критерий минимума СКО более распространен на практике, мы сосредоточим наше внимание на нём. Основываясь на предположении, что предшествующие продетектированные символы в цепи обратной связи приняты правильно, минимизация СКО

$$J(K_1,K_2) = E\left|I_k - \hat{I}_k\right|^2. \qquad (10.3.2)$$

приводит к следующей системе линейных уравнений для коэффициентов в фильтре прямой связи:

$$\sum_{j=-K_1}^{0} \psi_{lj} c_j = f_{-l}^*, \qquad l = -K_1,\ldots,-1,0, \qquad (10.3.3)$$

где

$$\psi_{lj} = \sum_{m=0}^{-l} f_m^* f_{m+l-j} + N_0\delta_{lj}, \qquad l,j = -K_1,\ldots,-1,0. \qquad (10.3.4)$$

Коэффициенты в фильтре обратной связи эквалайзера определяется через коэффициенты фильтра прямой связи следующим выражением:

$$c_k = -\sum_{j=-K_1}^{0} c_j f_{k-j}, \qquad k = 1,2,\ldots,K_2. \qquad (10.3.5)$$

Величины коэффициентов в блоке обратной связи приводят к полному исключению МСИ от предшествующих продетектированных символов в предположении, что предыдущие решения были правильными и что $K_2 \geq L$ (см. задачу 10.9).

10.3.2. Качество эквалайзера с обратной связью по решению ЭОСР

Теперь сосредоточим наше внимание на качестве, достигаемом эквалайзером с обратной связью по решению. Точный расчет качества сложен при существовании

неправильных решений, сделанных детектором, которые затем распространяются по каналу обратной связи. При отсутствии ошибок в решениях получаем для минимума СКО

$$J_{\min}(K_1) = 1 - \sum_{j=-K_1}^{0} c_j f_{-j} \tag{10.3.6}$$

При переходе к пределу $(K_1 \to \infty)$, когда число ячеек в фильтре прямой связи неограниченно, мы получаем наименьшую возможную СКО, обозначаемую $J_{\min}$. При некоторых усилиях $J_{\min}$ можно, как показано Сальцем (1973), выразить через спектральные характеристики канала и аддитивного шума. Эта более приемлемая форма для $J_{\min}$ определяется так:

$$J_{\min} = \exp\left\{\frac{T}{2\pi}\int_{-\pi/T}^{\pi/T} \ln\left[\frac{N_0}{X(e^{j\omega T}) + N_0}\right] d\omega\right\}. \tag{10.3.7}$$

Соответствующее выходное ОСШ

$$\gamma_\infty = \frac{1 - J_{\min}}{J_{\min}} = -1 + \exp\left\{\frac{T}{2\pi}\int_{-\pi/T}^{\pi/T} \ln\left[\frac{N_0 + X(e^{j\omega T})}{N_0}\right] d\omega\right\}. \tag{10.3.8}$$

Снова видим, что при отсутствии МСИ $X(e^{j\omega T}) = 1$ и, следовательно, $J_{\min} = N_0/(1+N_0)$. Соответствующее выходное ОСШ равно $\gamma_\infty = 1/N_0$.

Пример 10.3.1. Интересно сравнить величину $J_{\min}$ для эквалайзера с обратной связью по решению с величиной $J_{\min}$, получаемой линейным эквалайзером с минимальной СКО. Например, рассмотрим эквивалент канала с дискретным временем, состоящий из двух ячеек с коэффициентами f_0 и f_1. Минимальная СКО для этого канала равна

$$\begin{aligned} J_{\min} &= \exp\left\{\frac{T}{2\pi}\int_{-\pi/T}^{\pi} \ln\left[\frac{N_0}{1 + N_0 + 2|f_0||f_1|\cos(\omega T + \theta)}\right] d\omega\right\} \\ &= N_0 \exp\left[-\frac{1}{2\pi}\int_{-\pi/T}^{\pi} \ln(1 + N_0 + 2|f_0||f_1|\cos\omega) d\omega\right] = \frac{2N_0}{1 + N_0 + \sqrt{(1+N_0)^2 - 4|f_0 f_1|^2}}. \end{aligned} \tag{10.3.9}$$

Заметим, что $J_{\min}$ максимизируется, когда $|f_0| = |f_1| = \sqrt{\frac{1}{2}}$. Тогда

$$J_{\min} = \frac{2N_0}{1 + N_0 + \sqrt{(1+N_0)^2 - 1}} \approx 2N_0, \qquad N_0 \ll 1. \tag{10.3.10}$$

Соответствующее выходное ОСШ

$$\gamma_\infty \approx \frac{1}{2N_0}, \qquad N_0 \ll 1. \tag{10.3.11}$$

Следовательно, имеется потеря в ОСШ на 3 дБ, обусловленная наличием МСИ. По сравнению с этим, потеря качества для линейного эквалайзера очень существенная. Его выходное ОСШ, определяемое (10.2.53), равно $\gamma_\infty \approx (2/N_0)^{1/2}$ для $N_0 \ll 1$.

Пример 10.3.2. Рассмотрим канал с экспоненциально затухающей импульсной характеристикой вида

$$f_k = (1 - a^2)^{1/2} a^k, \qquad k = 0, 1, 2, \ldots, \tag{10.3.12}$$

где $a < 1$. Выходное ОСШ для эквалайзера с обратной связью по решению равно

$$\gamma_\infty = -1 + \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln\left[\frac{1+a^2+(1-a^2)/N_0 - 2a\cos\omega}{1+a^2-2a\cos\omega}\right]d\omega\right\} =$$

$$= -1 + \frac{1}{2N_0}\left\{1-a^2+N_0(1+a^2)+\sqrt{\left[1-a^2+N_0(1+a^2)\right]^2 - 4a^2N_0}\right\} \approx \quad (10.3.13)$$

$$\approx \frac{(1-a^2)\left[1+N_0(1+a^2)/(1-a^2)\right]-N_0}{N_0} \approx \frac{1-a^2}{N_0}, \qquad N_0 << 1$$

Таким образом, потеря в ОСШ равна $10\lg(1-a^2)$дБ. Для сравнения, линейный эквалайзер имеет потерю $10\lg[(1-a^2)/(1+a^2)]$ дБ.

Эти результаты иллюстрируют превосходство эквалайзера с обратной связью по решению, когда пренебрегают влиянием на качество ошибок в цепи обратной связи, относительно линейного эквалайзера. Очевидно, что можно достичь значительный выигрыш в качестве относительно линейного эквалайзера включением блока обратной связи по решению, что ограничивает МСИ от предшествующих продетектированных символов.

Оценить влияние ошибок в канале обратной связи на вероятность ошибки эквалайзера с обратной связью по решению возможно путём моделирования методом Монте-Карло. С целью иллюстрации приведём нижеследующие результаты для передачи сигналов двоичной АМ по модели эквивалентного канала с дискретным временем, показанной на рис.10.2.5 b и c.

Результаты моделирования показаны на рис.10.3.2.

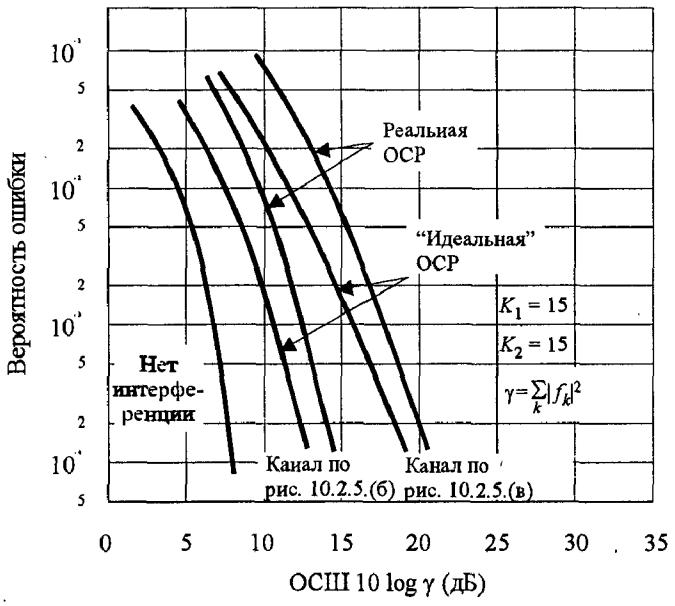


Рис. 10.3.2. Характеристики эквалайзера с обратной связью по решению при наличии и отсутствии ошибок в цепи обратной связи

Прежде всего, сравнение этих результатов с теми, которые представлены на рис.10.2.4, приводят нас к заключению, что эквалайзер с обратной связью по решению даёт существенное улучшение качества относительного линейного эквалайзера, имеющего то же число ячеек. Во-вторых, эти результаты указывают на то, что всё еще имеется существенное ухудшение качества эквалайзера с обратной связью по решению, обусловленное остаточной МСИ, особенно в каналах с большими искажениями, таким, как показано на рис.10.2.5 c. В заключение, потеря качества, обусловленная ошибочными

решениями в канале обратной связи, равна примерно 2 дБ для каналов с рассматриваемыми характеристиками. Дополнительные результаты для вероятности ошибки эквалайзера с обратной связью по решению с ошибками в канале обратной связи можно найти в статьях Дутвайлера и др. (1974) и Боли (1992).

Структура ЭОСР, которая анализировалась выше, имеет T-стробирующий фильтр в прямом направлении. Оптимальность такой структуры основывается на предположении, что аналоговый фильтр, предшествующий ЭОСР, согласован с сигнальным импульсом, искажённым каналом, а его выход стробируется в оптимальные моменты времени. На практике отклик канала априори неизвестен, так что невозможно синтезировать идеальный согласованный фильтр. С учётом этой трудности привычно в практических приложениях использовать дробно-стробирующий прямой фильтр. Конечно, стробирование в фильтре обратной связи остаётся с интервалом T. Использование ДЭ в качестве фильтра прямой связи в ЭОСР уменьшает чувствительность системы к ошибкам синхронизации.

Сравнение качества с МППО. Мы включили этот подраздел, чтобы сравнить качество ЭОСР с качеством эквалайзера МППО. Для двухпутевого канала с $f_0 = f_1 = \sqrt{\frac{1}{2}}$ мы показали, что эквалайзер с МППО не даёт потери в ОСШ по отношению к случаю отсутствия МСИ, в то время как эквалайзер с обратной связью по решению даёт потери около 3 дБ. Для каналов с ещё большими искажениями преимущество в ОСШ эквалайзера МППО относительно ЭОСР даже больше. Рис.10.3.3 иллюстрирует сравнение вероятности ошибок для этих двух алгоритмов выравнивания, полученных моделированием методом Монте-Карло для сигналов двоичной АМ и для характеристик канала, показанных на рис.10.2.5 b и c.

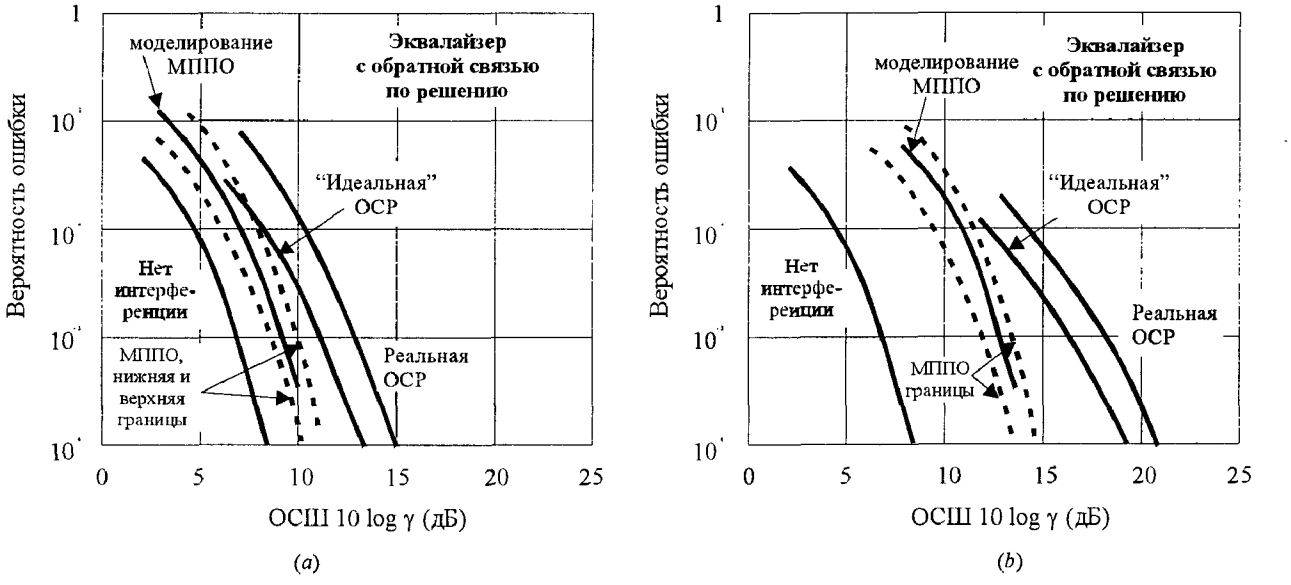


Рис.10.3.3. Сравнение характеристик качества МППО и эквалайзера с ОСР для канальных характеристик показанных (*a*) на рис. 10.2.5(*b*) и (*b*) на рис 10.2.5(*c*)

Кривые вероятности ошибок для двух методов имеют различные наклоны, поэтому разница в ОСШ увеличивается по мере уменьшения вероятности ошибки. Для сравнения на рис.10.3.3 дана также вероятность ошибки в канале с АБГШ без МСИ.

10.3.3. Эквалайзер с обратной связью по решению и предсказанием

Бельфиоре и Парк (1979) предложили другую структуру ЭОСР, которая эквивалентна той, которая показана на рис.10.3.1 при условии, что прямой фильтр имеет неограниченное число ячеек. Эта структура состоит из ДЭ как фильтра прямой связи и линейного предсказателя как фильтра обратной связи, как показано на рис.10.3.4.

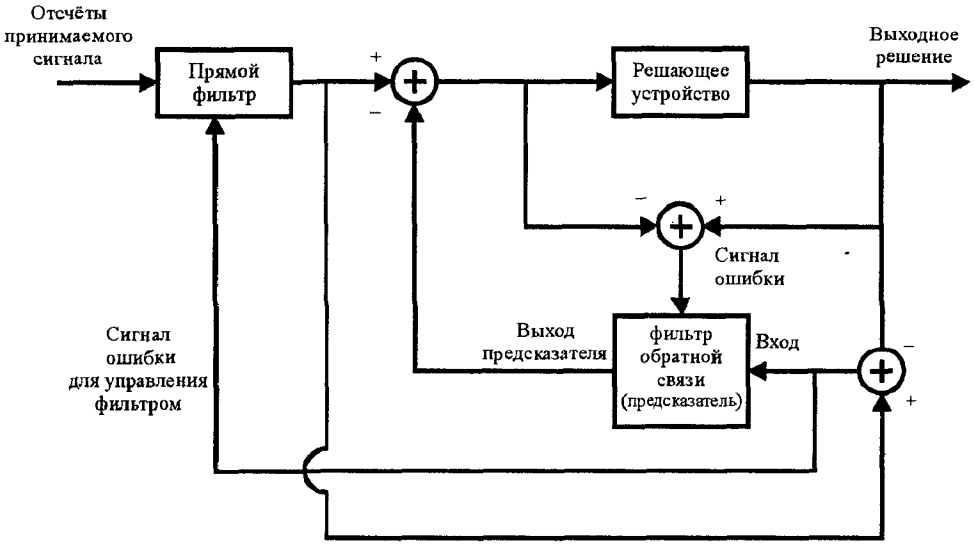


Рис. 10.3.4. Блок-схема ЭОСР с предсказанием

Рассмотрим вкратце характеристики качества этого эквалайзера.

Прежде всего, шум на выходе фильтра прямой связи неограниченной длины имеет спектральную плотность мощности

$$\frac{N_0 X(\mathrm{e}^{j\omega T})}{\left|N_0 + X(\mathrm{e}^{j\omega T})\right|^2}, \quad |\omega| \le \frac{\pi}{T}. \tag{10.3.14}$$

Остаточная МСИ имеет спектральную плотность мощности

$$\left|1 - \frac{X(\mathrm{e}^{j\omega T})}{N_0 + X(\mathrm{e}^{j\omega T})}\right|^2 = \frac{N_0^2}{\left|N_0 + X(\mathrm{e}^{j\omega T})\right|^2}, \quad |\omega| \le \frac{\pi}{T}. \tag{10.3.15}$$

Сумма двух этих спектров представляет спектральную плотность мощности суммарного шума и межсимвольной интерференции на выходе фильтра прямой связи. Таким образом, суммируя (10.3.14) и (10.3.15), мы получаем

$$E(\omega) = \frac{N_0}{N_0 + X(\mathrm{e}^{j\omega T})}, \quad |\omega| \le \frac{\pi}{T}. \tag{10.3.16}$$

Как мы видели раньше, если $X(\mathrm{e}^{j\omega T}) = 1$, канал идеален и, следовательно, невозможно дальше уменьшать СКО. С другой стороны, если имеются искажения в канале, мощность шумовой последовательности на выходе фильтра прямой связи можно уменьшить посредством линейного предсказания, основанного на последних величинах последовательности ошибок.

Если $\mathcal{B}(\omega)$ представляет частотную характеристику неограниченного по длине предсказателя обратной связи, т. е.

$$\mathcal{B}(\omega) = \sum_{n=1}^{\infty} b_n \mathrm{e}^{-j\omega nT}, \tag{10.3.17}$$

то ошибка на выходе предсказателя равна

$$E(\omega)-E(\omega)B(\omega)=E(\omega)\left[1-\mathscr{B}(\omega)\right]. \qquad (10.3.18)$$

Минимизация среднего квадрата этой ошибки, т.е. функционала

$$J=\frac{1}{2\pi}\int_{-\pi/T}^{\pi/T}\left|1-\mathscr{B}(\omega)\right|^2\left|E(\omega)\right|^2 d\omega \qquad (10.3.19)$$

по коэффициентам предсказателя $\{b_n\}$ приводит к оптимальному предсказателю в виде

$$\mathscr{B}(\omega)=1-\frac{G(\omega)}{g_0}, \qquad (10.3.20)$$

где $G(\omega)$ является решением уравнения

$$G(\omega)G^*(-\omega)=\frac{1}{\left|E(\omega)\right|^2} \qquad (10.3.21)$$

и

$$G(\omega)=\sum_{n=0}^{\infty}g_n e^{-j\omega nT}. \qquad (10.3.22)$$

Выходом неограниченного по длине линейного предсказателя является белая шумовая последовательность со спектральной плотностью мощности $1/g_0^2$, а соответствующее минимальное значение СКО определяется (10.3.7). Следовательно, СКО для ЭОСР с предсказанием идентична с общепринятым ЭОСР.

Хотя эти две структуры ЭОСР приводят к эквивалентному качеству, если их длины неограниченны, ЭОСР с предсказателем является субоптимальной, если длины двух фильтров ограничены. Причина оптимальности общепринятого ЭОСР относительно проста. Оптимизация его коэффициентов ячеек в фильтрах прямой и обратной связи выполняется совместно. Следовательно, это даёт минимум СКО. С другой стороны, оптимизация фильтра прямой связи и предсказателя обратной связи в ЭОСР с предсказателем выполняется раздельно. Следовательно, его СКО по крайней мере так же велика, как для общепринятого ЭОСР. Вопреки этой субоптимальности ЭОСР с предсказателем, он пригодный как эквалайзер для решётчато-кодированных сигналов, в то время как общепринятый ЭОСР непригоден, как описывается в следующей главе.

10.4. БИБЛИОГРАФИЧЕСКИЕ ЗАМЕЧАНИЯ И ССЫЛКИ

Канальное выравнивание для цифровой связи было разработано Лакки (1965, 1966), который сосредоточился на линейных эквалайзерах, которые были оптимизированы на основе критерия пикового искажения. Критерий минимума СКО для оптимизации коэффициентов эквалайзера был предложен Уидроу (1966)

Выравнивание с обратной связью по решению было предложено и проанализировано Остином (1967)[1]. Анализ качества ЭОСР можно найти в статьях Монсена (1971), Джордже и др. (1971), Прайса (1972), Сальца (1973), Дутвейлера и др. (1974), Альтекера и Боли (1993).

---

[1] Обратная связь по решению для оптимальной демодуляции сигналов цифровой модуляции в каналах с МСИ (многопутевых каналах или каналах с эхо-сигналами) и АБГШ при нулевой задержке решения (D=0) или при анализе на тактовом интервале была впервые предложена Д.Д. Кловским [1960,]. Ещё ранее [1958,] им была предложена система цифровой связи с когерентным приёмом и периодическим зондированием канала (система с испытательным импульсом и предсказанием – СИИП). Алгоритм Кловского был в дальнейшем (1970) развит Кловским и Николаевым (АКН) для демодуляции символов в каналах с МСИ и АБГШ для произвольной величины задержки принятия решения D [ ], т.е. анализе на произвольном временном интервале. АКН обеспечивает в каналах с МСИ помехоустойчивость примерно такую же, как и алгоритм МППО (АВ) при анализе на том же временном интервале, но требует меньших вычислительных затрат (прп).

Использование алгоритма Витерби как оптимального максимально-правдоподобного последовательного оценивателя для символов, искажённых МСИ, было предложено и проанализировано Форни (1972) и Омура (1971). Его использование для модулированных по несущей сигналов было рассмотрено Унгербоеком (1974) и Мак-Кеншни (1973).

ЗАДАЧИ

10.1. В двоичной системе АМ выход детектора

$$y_m = a_m + n_m + i_m,$$

где $a_m = \pm 1$ – желательный сигнал, n_m – гауссовская случайная величина с нулевым средним и дисперсией σ_n^2, а i_m представляет МСИ, обусловленную искажением в канале. Слагаемое МСИ является случайной величиной, которая принимает значения $-\frac{1}{2}$, 0, $\frac{1}{2}$ с вероятностями $\frac{1}{4}$, $\frac{1}{2}$ и $\frac{1}{4}$ соответственно. Определите среднюю вероятность ошибки, как функцию от σ_n^2.

10.2. В двоичной системе АМ импульс синхронизации, который определяет момент стробирования выхода коррелятора, сдвинут от оптимального времени стробирования на 10 %.

a Если импульс сигнала прямоугольный, определите потерю в ОСШ, обусловленную расстройкой во времени.

b Определите величину МСИ, обусловленной расстройкой во времени, и определите её влияние на качество приёма.

10.3. Частотная характеристика низкочастотного канала может быть аппроксимирована выражением

$$H(f) = \begin{cases} 1 + \alpha \cos 2\pi f t_0 & (|\alpha| < 1,\ |f| \le W) \\ 0 & (\text{при других } \alpha,\ f), \end{cases}$$

где W – полоса частот канала. Входной сигнал $s(t)$ с ограниченной полосой частот W Гц проходит через этот канал.

a Покажите, что выход

$$y(t) = s(t) + \tfrac{1}{2}\alpha\left[s(t - t_0) + s(t + t_0)\right],$$

т.е. канал производит пару эхо-сигналов.

b Предположим, что принимаемый сигнал $y(t)$ проходит через фильтр, согласованный с $s(t)$. Определите выход согласованного фильтра при $t = kT$, $k = 0, \pm 1, \pm 2, \ldots$, где T – длительность символа.

c Какова модель МСИ канала, если $t_0 = T$?

10.4. Проводной канал длиной 1000 км используется для передачи данных посредством двоичной АМ. Регенеративные повторители размещены через 50 км вдоль линии. Каждый сегмент канала имеет идеальную (постоянную) частотную характеристику в полосе частот $0 \le f \le 1200$ Гц и ослабление 1 дБ/км. Канальный шум – АБГШ.

a Какую максимальную битовую скорость можно передавать без МСИ?

b Определите требуемое значение $\mathscr{E}_b/N_0$ для достижения вероятности ошибки на бит $P_2 = 10^{-7}$ для каждого регенератора.

c Определите передаваемую мощность в каждом регенераторе для достижения желательной величины $\mathscr{E}_b/N_0$, где $N_0 = 4{,}1 \cdot 10^{-21}$ Вт/Гц.

10.5. Докажите соотношение (10.1.13) для автокорреляционной функции шума на выходе согласованного фильтра.

10.6. Для случая АМ с коррелированным шумом, корреляционные метрики в алгоритме Витерби можно выразить в общем виде (Унгербоек, 1974):

$$CM(\mathbf{I}) = 2\sum_n I_n r_n - \sum_n \sum_m I_n I_m x_{n-m},$$

где $x_n = x(nT)$ – отсчёты сигнала на выходе согласованного фильтра, $\{I_n\}$ – последовательность данных, $\{r_n\}$ – принимаемая сигнальная последовательность на выходе согласованного фильтра. Определите метрики для дуобинарного сигнала.

10.7. Рассмотрим использование сигнального импульса со спектром (корень квадратный) приподнятого косинуса с коэффициентом ската, равным 1, для передачи сигнала двоичной АМ по идеальному частотно-ограниченному каналу, который передаёт импульс без искажений. Принимаемый сигнал

$$\upsilon(t) = \sum_{k=-\infty}^{\infty} I_k g_T(t - kT_b),$$

где сигнальный интервал $T_b = \frac{1}{2}T$. Таким образом, скорость передачи символов в два раза больше, чем при отсутствии МСИ.

a Определите величины МСИ на выходе согласованного фильтра демодулятора.

b Постройте решётку для последовательного детектора максимального правдоподобия и обозначьте состояния.

10.8. Двоичные противоположные сигналы передаются по неидеальному частотно-ограниченному каналу, который вызывает МСИ между двумя соседними символами. Для изолированного переданного импульса сигнала $s(t)$ выход демодулятора (при отсутствии шума) равен $\sqrt{\mathscr{E}_b}$ при $t = T$, $\sqrt{\mathscr{E}_b}/4$ при $t = 2T$ и нулю при $t = kT$, $k > 2$, где $\mathscr{E}_b$ – энергия сигнала, а T – сигнальный интервал.

a Определите среднюю вероятность ошибки, предполагая, что два сигнала имеют равные вероятности, а аддитивный шум гауссовский и белый.

b Построив кривые вероятности ошибки, полученные в (**a**) и для случая отсутствия МСИ, определите относительное различие в ОСШ при вероятности ошибки 10^{-6}.

10.9. Получите выражение (10.3.5) для коэффициентов фильтра обратной связи ЭОСР.

10.10. Двоичная АМ используется для передачи информации по невыравненному линейному фильтровому каналу. Когда передаётся снмвол $a = 1$, выход демодулятора (без шума) равен

$$x_m = \begin{cases} 0{,}3 & (m = 1), \\ 0{,}9 & (m = 0), \\ 0{,}3 & (m = -1), \\ 0 & (\text{при других } m). \end{cases}$$

a. Синтезируйте линейный ЭНВП с тремя ячейками так, чтобы выходы были равны

$$q_m = \begin{cases} 1 & (m = 0), \\ 0 & (m = \pm 1). \end{cases}$$

b. Определите q_m для $m = \pm 2, \pm 3$ путём свёртки импульсных характеристик эквалайзера и канала.

10.11. Передача сигнального импульса со спектром приподнятого косинуса через канал приводит к следующим выходным отсчётам (без учёта шума) демодулятора:

$$x_k = \begin{cases} -0{,}5 & (k = -2), \\ 0{,}1 & (k = -1), \\ 1 & (k = 0), \\ -0{,}2 & (k = 1), \\ 0{,}05 & (k = 2), \\ 0 & (\text{при других } k). \end{cases}$$

a. Определите коэффициенты ячеек трёхячеечного линейного эквалайзера, основанного на критерии обнуления.

b. Для коэффициентов, определённых в (a), найдите выход эквалайзера для случая изолированного импульса. Определите также остаточную МСИ и её протяжённость во времени.

10.12. Неидеальный с ограниченной полосой частот канал вызывает МСИ по трём соседним символам. Отсчёты отклика согласованного фильтра демодулятора в моменты стробирования kT равны

$$\int_{-\infty}^{\infty} s(t)s(t - kT)dt = \begin{cases} \varepsilon_b & (k = 0), \\ 0{,}9\varepsilon_b & (k = \pm 1), \\ 0{,}1\varepsilon_b & (k = \pm 2), \\ 0 & (\text{при других } k). \end{cases}$$

a. Определите коэффициенты ячеек трёхячеечного линейного эквалайзера, который выравнивает характеристику канала (принимаемого сигнала) до эквивалентного сигнала с парциальным откликом (дуобинарный сигнал)

$$y_k = \begin{cases} \mathscr{E}_b & (k = 0,\ 1), \\ 0 & (\text{при других } k). \end{cases}$$

b. Предположим, что за линейным эквалайзером в (**a**) следует детектор Витерби для сигнала с парциальным откликом. Дайте оценку вероятности ошибки, если аддитивный шум гауссовский и белый со спектральной плотностью мощности $\frac{1}{2}N_0$ Вт/Гц.

10.13. Определите весовые коэффициенты ячеек трёхячеечного обнуляющего эквалайзера, если МСИ простирается на 3 символа и характеризуется величинами

$$x(0)=1,\ x(-1)=0{,}3,\ x(1)=0{,}2.$$

Определите также остаточную МСИ на выходе эквалайзера для оптимальных коэффициентов ячеек.

10.14. При передаче по радиоканалу прямой видимости сигнал достигает точки приёма по двум путям распространения: прямой путь и путь с запаздыванием, обусловленный сигналом, отражённым от местных предметов. Предположим, что принимаемый сигнал имеет вид

$$r(t)=s(t)+\alpha s(t-T)+n(t),$$

где $s(t)$ – переданный сигнал, α – ослабление $(\alpha<1)$ второго луча, а $n(t)$ – АБГШ.

a. Определите выход демодулятора при $t=T$ и $t=2T$, который создаёт фильтр, согласованный с сигналом $s(t)$.

b. Определите вероятность ошибки для посимвольного детектора, если передаются двоичные противоположные сигналы, а детектор игнорирует МСИ.

c. Какова характеристика качества (вероятность ошибки) для простейшего (одноячеечного) ЭОСР, который оценивает α и устраняет МСИ? Нарисуйте структуру детектора, который применяет ЭОСР.

10.15. Повторите задачу 10.10, используя минимум СКО в качестве критерия оптимизации коэффициентов ячеек. Предположите, что спектральная плотность мощности шума равна 0,1 Вт/Гц.

10.16. В магнитном записывающем канале, где считывающий импульс, возникающий при положительном переходе записывающего тока, имеет вид

$$p(t)=\left[1+\left(\frac{2t}{T_{50}}\right)^2\right]^{-1},$$

используется линейный эквалайзер для выравнивания импульса к парциальному отклику. Параметр T_{50} определяется как ширина импульса на 50% уровне амплитуды. Битовая скорость равна $1/T_b$, а отношение $T_{50}/T_b=\Delta$ является нормированной плотностью записи.

Предположим, что импульс выравнивается к величинам парциальных откликов $x(nT)$,

$$x(nT)=\begin{cases}1 & (n=-1,\ 1),\\ 2 & (n=0),\\ 0 & (\text{при других } n),\end{cases}$$

где $x(t)$ представляет выровненную огибающую импульса.

a. Определите спектр $X(f)$ ограниченного по полосе выровненного импульса.

b. Определите возможные выходные уровни детектора, предположив, что последовательные переходы могут возникать со скоростью $1/T_b$.

c. Определите вероятность ошибки посимвольного детектора для этого сигнала, предполагая, что аддитивный шум гауссовский, с нулевым средним и дисперсией σ^2.

10.17. Нарисуйте решётку состояний для детектора Витерби выровненного сигнала из задачи 10.16 и обозначьте все состояния. Определите минимальное евклидово расстояние между сливающимися путями.

10.18. Рассмотрите задачу выравнивания в эквивалентном канале дискретного времени, показанном на рис. Р10.18. Информационная последовательность $\{I_n\}$ двоичная (± 1) и некоррелированная. Аддитивный шум $\{v_n\}$ белый и вещественный с дисперсией N_0. Принимаемая последовательность $\{y_n\}$ обрабатывается линейным трёхячеечным эквалайзером, который оптимизирован по критерию минимума СКО.

a. Определите оптимальные коэффициенты эквалайзера, как функции от N_0.

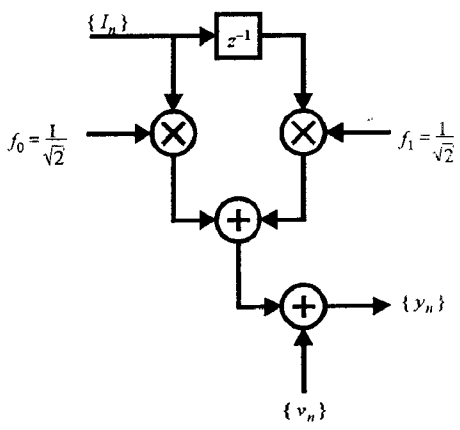


Рис. Р.10.18

b. Определите три собственных значения λ_1, λ_2 и λ_3 ковариационной матрицы **Г** и соответствующие собственные векторы $\mathbf{v}_1$, $\mathbf{v}_2$, и $\mathbf{v}_3$.**c.** Определите минимум СКО для трёхячеечного эквалайзера, как функцию от N_0.

d. Определите выходное ОСШ для трёхячеечного эквалайзера, как функцию от N_0. Что даёт сравнение с выходным ОСШ для неограниченного по длине эквалайзера? Вычислите выходные ОСШ для этих двух эквалайзеров при N_0 =0,1

10.19. Используйте принцип ортогональности для получения уравнений для коэффициентов в эквалайзере с обратной связью по решению, основанном на критерии минимума СКО и даваемом (10.3.3) и (10.3.5).

10.20. Предположим, что модель с дискретным временем для МСИ характеризуется коэффициентами ячеек $f_0, f_1, \ldots f_L$. Из уравнений для коэффициентов ячеек ЭОСР покажите, что только L ячеек необходимы в фильтре обратной связи. Это значит, что если $\{c_k\}$ являются коэффициентами фильтра обратной связи, то $c_k = 0$ для $k \geq L+1$.

10.21. Рассмотрите модель канала, показанную на рис. Р.10.21.

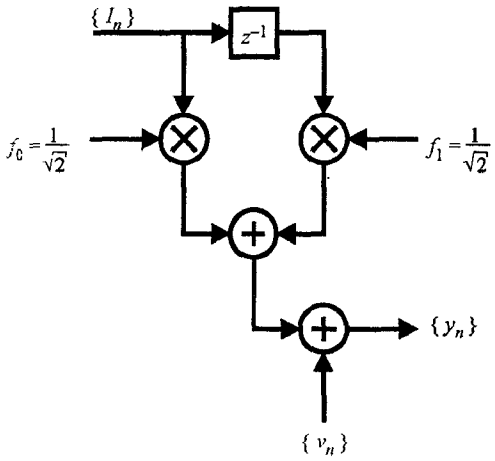


Рис. Р.10.21

$\{\nu_n\}$– белая шумовая последовательность из вещественных случайных величин с нулевым средним и дисперсией N_0. Предположим, что канал надо выровнять посредством ЭОСР, имеющего двухячеечный фильтр прямой связи (c_0, c_{-1}) и одноячеечный фильтр обратной связи (c_1). Коэффициенты $\{c_l\}$ оптимизированы на основе критерия минимума СКО.

a. Определите оптимальные коэффициенты и их приближённые значения для $N_0 << 1$.

b. Определите точное значение минимума СКО и аппроксимацию первого порядка, соответствующую случаю $N_0 << 1$.

c. Определите точное значение выходного ОСШ для трёхячеечного эквалайзера, как функцию от N_0, и аппроксимацию первого порядка, соответствующую случаю $N_0 << 1$.

d. Сравните результаты (**b**) и (**c**) с качеством ОСР неограниченной длины.

e. Рассчитайте и сравните точные значения выходного ОСШ для двухячеечного и неограниченного по длине ЭОСР для частного случая, когда $N_0 = 0{,}1$ и 0,01. Прокомментируйте, насколько трёхячеечный эквалайзер приближается к характеристикам неограниченного по длине ЭОСР.

10.22. Импульс и его спектр (типа приподнятого косинуса) показаны на рис. Р.10.22. Этот импульс используется для передачи цифровой информации по каналу с ограниченной полосой пропускания со скоростью $1/T$ символов/с.

a. Каков коэффициент ската β ?

b. Какова скорость передачи импульсов?

c. Канал искажает сигнальный импульс. Предположим, что отсчётные значения отфильтрованного принимаемого импульса $x(t)$ таковы, как на рис. Р.10.22(с). Очевидно, что имеются пять интерферирующих компонент сигнала. Укажите последовательность из символов ± 1, которая вызывает наибольшую (благоприятную и неблагоприятную) интерференцию) и соответствующие значения величины интерференции (пиковое искажение).

d. Какова вероятность возникновения наихудшей последовательности, полученной в (**c**), если предположить, что двоичные символы передаются равновероятно и независимо?

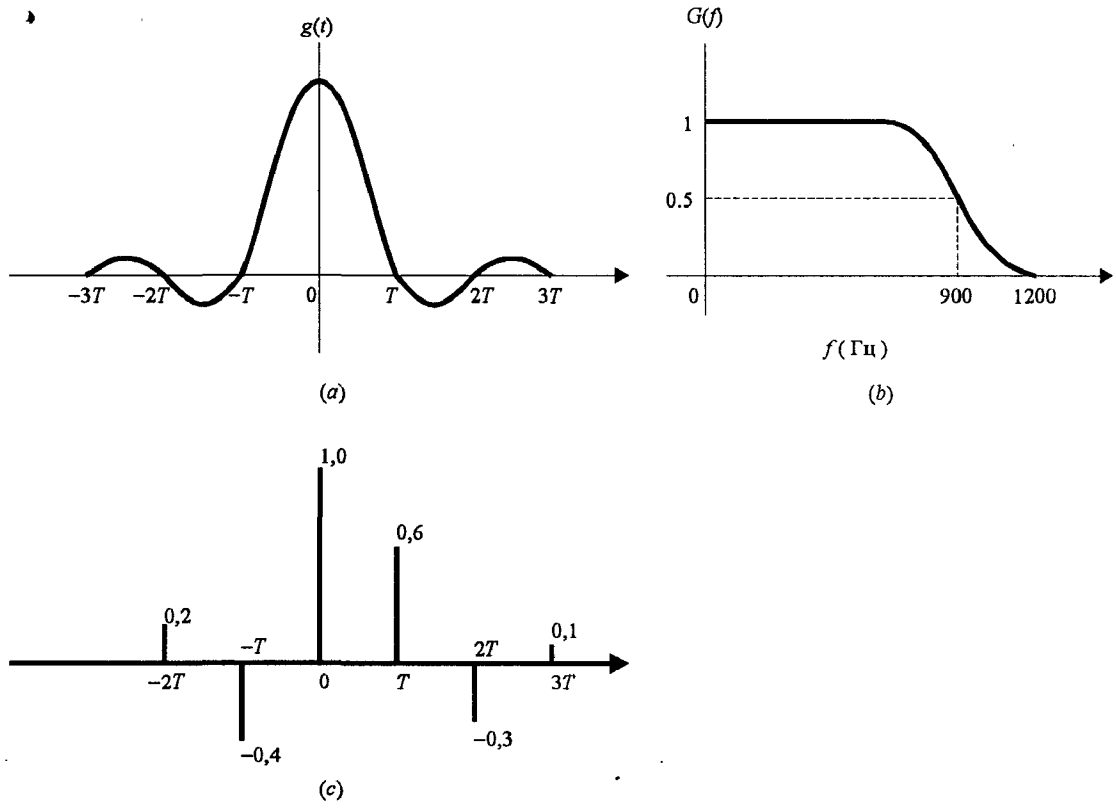


Рис. Р.10.22

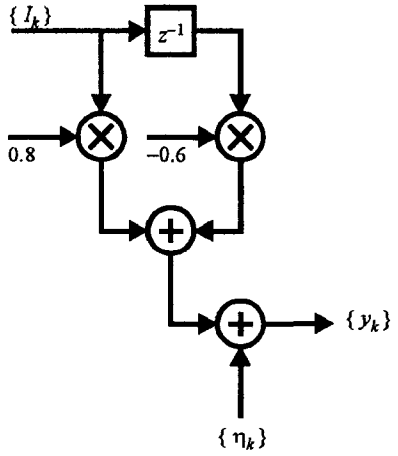


Рис. Р.10.23

10.23. Канал с рассеянием во времени с импульсным откликом $h(t)$ используется для передачи четырёхфазовой ФМ со скоростью $R = 1/T$ символов/с. Эквивалентный канал с дискретным временем показан на рис. Р.10.23. Последовательность $\{\eta_k\}$ – белый гауссовский шум с нулевым средним и дисперсией $\sigma^2 = N_0$.

a. Каковы отсчёты автокорреляционной функции $\{x_k\}$, определённые выражением

$$x_k = \int_{-\infty}^{\infty} h^*(t)h(t+kT)dt$$

для этого канала?

b. Минимум СКО линейного эквалайзера и эквалайзера с ОСР с неограниченным числом ячеек зависит от *сложенной спектральной характеристики* канала

$$\frac{1}{T}\sum_{n=-\infty}^{\infty}\left|H\left(\omega+\frac{2\pi n}{T}\right)\right|^2,$$

где $H(\omega)$ – преобразование Фурье от $h(t)$. Определите сложенную частотную характеристику канала, данного выше.

c. Используйте Ваш ответ в **(b)** для получения минимума СКО линейного эквалайзера через сложенную частотную характеристику канала (можно оставить ответ в интегральной форме).

d. Повторите **(c)** для эквалайзера с ОСР с неограниченным числом ячеек.

10.24. Рассмотрите систему четырёхуровневой АМ, которая может передавать уровни 3,1,−1,−3. Канал вносит МСИ по двум соседним символам. Модель эквивалентного канала с дискретным временем показана на рис. Р.10.24.

$\{\eta_k\}$ – это поеледовательность вещественных независимых гауссовских случайных величин с нулевым средним и с дисперсией $\sigma^2 = N_0$. Принимаемая последовательность

$$y_1 = 0{,}8I_1 + n_1,$$
$$y_2 = 0{,}8I_2 - 0{,}6I_1 + n_2,$$
$$y_3 = 0{,}8I_3 - 0{,}6I_2 + n_3,$$
$$\vdots$$
$$y_k = 0{,}8I_k - 0{,}6I_{k-1} + n_k.$$

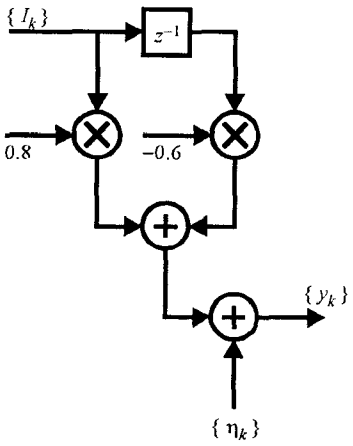


Рис. Р.10.24

a. Нарисуйте структуру дерева, показывающего возможные сигнальные последовательности для принимаемых сигналов y_1, y_2 и y_3.

b. Предположим, что используется алгоритм Витерби для детектирования информационной последовательности. Сколько вероятностей надо вычислить на каждом шаге работы алгоритма?

c. Сколько выживших последовательностей имеется в алгоритме Витерби в этом канале?

d. Предположим, что принимаемые сигналы

$$y_1 = 0{,}5;\ y_2 = 2{,}0;\ y_3 = -1{,}0.$$

Определите выжившие последовательности на шаге y_3 и соответствующие метрики.

e. Дайте плотную верхнюю границу для вероятности ошибки при передаче по каналу четырёхуровневой АМ.

10.25. Трансверсальный эквалайзер с K ячейками имеет импульсную характеристику

$$e(t) = \sum_{k=0}^{K-1} c_k \delta(t - kT),$$

где T – задержка между соседними ячейками, а передаточная функция

$$E(z) = \sum_{k=0}^{K-1} c_k z^{-k}.$$

Дискретное преобразование Фурье (ДПФ) для коэффициентов эквалайзера $\{c_k\}$ определяется выражением

$$E_n \equiv E(z)\Big|_{z=e^{j2\pi n/K}} = \sum_{k=0}^{K-1} c_k e^{-j2\pi kn/K}, \quad n = 0, 1, \ldots, K-1.$$

Обратное ДПФ определяется выражением

$$b_k = \frac{1}{K}\sum_{k=0}^{K-1} E_n e^{j2\pi nk/K}, \quad k = 0, 1, \ldots, K-1.$$

a. Покажите, что $b_k = c_k$ при подстановке E_n в предыдущее выражение.

b. Из соотношений, данных выше, определите структуру эквивалентного фильтра, имеющего z-преобразование

$$E(z) = \underbrace{\frac{1 - z^{-K}}{K}}_{E_1(z)} \sum_{n=0}^{K-1} \underbrace{\frac{E_n}{1 - e^{j2\pi n/K} z^{-1}}}_{E_2(z)}.$$

c. Если $E(z)$ представлено каскадным соединением двух отдельных фильтров $E_1(z)$ и $E_2(z)$, нарисуйте блок-схему для каждого фильтра, используя z^{-1} для обозначения единичной задержки.

d. В трансверсальном фильтре настраиваемыми параметрами являются коэффициенты эквалайзера $\{c_k\}$. Каковы настраиваемые параметры эквивалентного эквалайзера в (b) и как они связаны с $\{c_k\}$?

11

АДАПТИВНЫЕ ЭКВАЛАЙЗЕРЫ

В 10-й главе мы рассмотрели оптимальные и субоптимальные приемники, которые компенсируют МСИ при передаче цифровой информации по частотно-ограниченным, не идеальным каналам. Оптимальный приемник использует максимально-правдоподобную оценку для детектирования информационной последовательности отсчетов фильтра демодулятора. Субоптимальные приемники используют или линейный эквалайзер, или эквалайзер с обратной связью по решению. При разработке трех методов выравнивания мы безоговорочно предполагали, что характеристика канала (импульсная характеристика или частотная характеристика) известна у приёмника. Однако в большинстве систем связи, которые используют эквалайзеры, характеристики канала априори неизвестны и, во многих случаях, характеристики канала меняются во времени. В этом случае, эквалайзеры проектируются так, чтобы приспосабливаться к характеристикам канала, и в случае, если они меняются во времени, адаптироваться к этому изменению.

В этой главе мы представим алгоритм для автоматической подстройки коэффициентов эквалайзера так, чтобы оптимизировать показатель качества и осуществить адаптивную компенсацию изменений во времени характеристик канала. Мы также проанализируем характеристики качества алгоритмов обработки, включая скорость их сходимости и их вычислительной сложности.

11.1. АДАПТИВНЫЙ ЛИНЕЙНЫЙ ЭКВАЛАЙЗЕР

Напомним, что в случае линейного эквалайзера мы рассмотрели два различных критерия для определения величин коэффициентов эквалайзера $\{c_k\}$. Один критерий основан на минимизации пикового искажения на выходе эквалайзера, определяемого (10.2.4). Второй критерий основан на минимизации СКО на выходе эквалайзера, которая определяется (10.2.25). Ниже мы опишем два алгоритма для выполнения оптимизации автоматически и адаптивно.

11.1.1. Алгоритм сведения к нулю

При использовании критерия пикового искажения искажение $\mathscr{D}(\mathbf{c})$, определяемое (10.2.22), минимизируется путем выбора коэффициентов эквалайзера $\{c_k\}$. В общем, это не простой в вычислительном отношении алгоритм для осуществления оптимизации, исключая частные случаи, когда пиковое искажение на входе эквалайзера, обозначенное $\mathscr{D}_0$, в (10.2.23) меньше единицы. Если $\mathscr{D}_0 < 1$, то искажение $\mathscr{D}(\mathbf{c})$ на выходе эквалайзера минимизируется путем получения отклика эквалайзера $q_n = 0$ для $1 \le |n| \le K$ и $q_0 = 1$. В этом случае получается простой в вычислительном отношении алгоритм, названный алгоритмом сведения к нулю АСН («нуль-форсирующим» алгоритмом), который и обеспечивает указанные условия.

«Нуль-форсирующее» решение достигается путем получения взаимной корреляции между последовательностью ошибок $\varepsilon_k = I_k - \hat{I}_k$ и желательной информационной

последовательностью $\{I_k\}$, равной нулю для сдвигов в области $0 \le |n| \le K$. Доказательство того, что именно это ведет к желательному решению, совсем простое. Мы имеем

$$E\left(\varepsilon_k I^*_{k-j}\right) = E\left[\left(I_k - \hat{I}_k\right) I^*_{k-j}\right] = E\left(I_k I^*_{k-j}\right) - E\left(\hat{I}_k I^*_{k-j}\right), \qquad j = -K, \ldots, K. \tag{11.1.1}$$

Предположим, что информационные символы некоррелированы, то есть $E\left(I_k I^*_j\right) = \delta_{kj}$ и что информационная последовательность $\{I_k\}$ некоррелирована с аддитивной шумовой последовательностью $\{\eta_k\}$. Для $\hat{I}_k$, мы используем выражение даваемое (10.2.41). Тогда, после вычисления математического ожидания в (11.1.1) получим

$$E\left(\varepsilon_k I^*_{k-j}\right) = \delta_{j0} - q_j, \qquad j = -K, \ldots, K. \tag{11.1.2}$$

Следовательно, условие

$$E\left(\varepsilon_k I^*_{k-j}\right) = 0, \qquad j = -K, \ldots, K \tag{11.1.3}$$

выполняется, когда $q_0 = 1$ и $q_n = 0$, $1 \le |n| \le K$.

Если характеристика канала неизвестна, взаимная корреляция, определяемая (11.1.1), также не известна. Эта трудность может быть преодолена путем передачи известной обучающей последовательности $\{I_k\}$ до приёмника, которую можно использовать для оценки взаимных корреляций путем подстановки средних во времени вместо средних по ансамблю, определяемых (11.1.1). После начального обучения, требующего передачи тестирующей последовательности определённой длины, которая равна или превышает длину эквалайзера, можно определить коэффициенты, удовлетворяющие (11.1.3).

Простой рекуррентный алгоритм для настройки коэффициентов эквалайзера можно записать так

$$c_j^{(k+1)} = c_j^{(k)} + \Delta \varepsilon_k I^*_{k-j}, \qquad j = -K, \ldots, -1, 0, 1, \ldots, K, \tag{11.1.4}$$

где $c_j^{(k)}$-значение j-го коэффициента в момент $t=kT$, $\varepsilon_k = I_k - \hat{I}_k$ – сигнал ошибки в момент $t=kT$, а Δ – скалярный множитель, который контролирует скорость настройки, как это будет объяснено позже в этом разделе. Это *алгоритм сведения к нулю* (АСН). Слагаемое $\varepsilon_k I^*_{k-j}$ является оценкой взаимной корреляции (среднее по ансамблю) $E(\varepsilon_k I^*_{k-j})$. Операция усреднения взаимной корреляции выполняется посредством рекуррентного алгоритма разностного уравнения первого порядка (11.1.4), которое реализуется простым интегратором с дискретным временем.

Вслед за периодом обучения, после которого коэффициенты эквалайзера сошлись к своим оптимальным значениям, решения на выходе детектора обычно достаточно надёжны, так что они могут быть использованы для продолжения процесса адаптации коэффициентов. Это называется *моделью адаптации, управляемой решениями*. В этом случае взаимные корреляции в (11.1.4) включают в себя сигнал ошибки $\tilde{\varepsilon}_k = \tilde{I}_k - \hat{I}_k$ и предетектированную выходную последовательность $\tilde{I}_{k-j}$, $j = -K, \ldots, K$. Следовательно, для адаптивной модели (11.1.4) получаем

$$c_j^{(k+1)} = c_j^{(k)} + \Delta \tilde{\varepsilon}_k \tilde{I}^*_{k-j}. \tag{11.1.5}$$

Рис.11.1.1 иллюстрирует ЭНВП с моделью обучения и с адаптивной моделью функционирования. Характеристики АСН похожи на характеристики алгоритма наименьших квадратов (НК), который минимизирует СКО и который подробно описывается в следующем разделе.

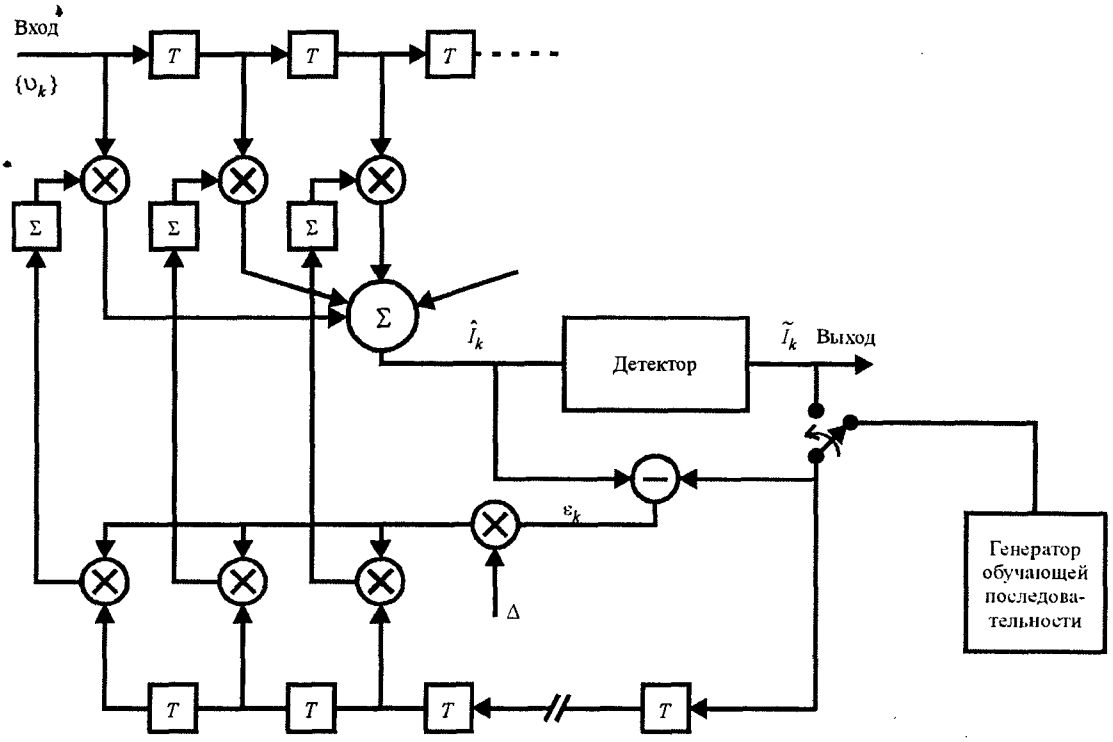


Рис. 11.1.1. Адаптивный эквалайзер с нулевыми взаимными помехами (ЭНВП)

11.1.2. Алгоритм наименьших квадратов (НК)

При минимизации СКО, обсуждённой в разделе 10.2.2, мы нашли, что оптимальные коэффициенты эквалайзера определяются из решения системы линейных уравнений, выраженной в матричной форме:

$$\mathbf{\Gamma C} = \xi, \tag{11.1.6}$$

где $\mathbf{\Gamma}$ – ковариационная матрица размером $(2K+1)\times(2K+1)$ отсчётов сигнала $\{\upsilon_k\}$, $\mathbf{C}$ – вектор-столбец из $(2K+1)$ коэффициентов эквалайзера, а ξ – вектор-столбец канальных коэффициентов фильтра размерности $(2K+1)$. Решение для вектора коэффициентов $\mathbf{C}_{\text{opt}}$ оптимального эквалайзера можно получить путём обращения ковариационной матрицы $\mathbf{\Gamma}$, что можно эффективно выполнить посредством алгоритма Левинсона–Дурбина, описанного в приложении А.

Альтернативно для вычисления $\mathbf{C}_{\text{opt}}$ можно использовать итеративную процедуру, которая избегает обращение матрицы. Вероятно простейшая итеративная процедура – это метод крутого спуска, когда можно начинать выбором произвольного начального вектора $\mathbf{C}$, скажем $\mathbf{C}_0$. Этот первоначальный выбор коэффициентов соответствует некоторой точке поверхности квадратичной функции СКО в пространстве коэффициентов размерности $(2K+1)$. Затем в этой точке на поверхности СКО вычисляется градиентный вектор $\mathbf{G}_0$, имеющий $2K+1$ градиентных компонент $\frac{1}{2}\partial J/\partial c_{0k}$, $k=-K,\ldots,-1,0,1,\ldots,K$. На каждом шаге вес меняется в направлении, противоположному соответствующей градиентной компоненте. Изменение веса на j-м шаге пропорционально объёму j-ой градиентной компоненты. Таким образом, последовательные значения коэффициентов $\mathbf{C}$ определяются согласно отношениям

$$\mathbf{C}_{k+1} = \mathbf{C}_k - \Delta\mathbf{G}_k, \qquad k = 0, 1, 2, \ldots, \tag{11.1.7}$$

а вектор градиента $\mathbf{G}_k$ равен

$$\mathbf{G}_k = \frac{1}{2}\frac{dJ}{d\mathbf{C}_k} = \mathbf{\Gamma C}_k - \xi = -E(\varepsilon_k \mathbf{V}_k^*). \qquad (11.1.8)$$

Вектор $\mathbf{C}_k$ представляет набор коэффициентов k-й итерации, $\varepsilon_k = I_k - \hat{I}_k$ является сигналом ошибки на k-й итерации. $\mathbf{V}_k$ – вектор отсчётов принимаемого сигнала, по которым делаются оценки $\hat{I}_k$, т.е. $\mathbf{V}_k = [\upsilon_{k+K} \ldots \upsilon_k \ldots \upsilon_{k-K}]^T$, а Δ – положительное число, выбираемое достаточно малым ля того, чтобы обеспечить сходимость итеративной процедуры. Если минимум СКО достигнут на некотором шаге $k = k_0$, тогда $\mathbf{G}_k = \mathbf{0}$, так что дальнейшие изменения у шаговых весов не происходят. В общем случае, при методе кратчайшего спуска точное значение $J_{\min}(K)$ нельзя получить при конечных величинах k_0. Однако, к нему можно как угодно приблизиться при некотором конечном значении k_0.

Принципиальная трудность метода кратчайшего спуска для определения оптимальных шаговых весов – это отсутствие знания вектора градиента $\mathbf{G}_k$, который зависит как от ковариационной матрицы $\mathbf{\Gamma}$, так и от вектора ξ взаимных корреляций. В свою очередь, эти величины зависят от коэффициентов $\{f_k\}$ эквивалентной модели канала с дискретным временем и от ковариаций информационной последовательности и аддитивного шума. Все эти величины могут быть, в общем, неизвестны на приёме. Чтобы преодолеть эти трудности, можно использовать оценку вектора градиентов. Это значит, что алгоритм настройки коэффициентов шаговых весов можно выразить в форме

$$\hat{\mathbf{C}}_{k+1} = \hat{\mathbf{C}}_k - \Delta\hat{\mathbf{G}}_k, \qquad (11.1.9)$$

где $\hat{\mathbf{G}}_k$ означает оценку вектора градиентов $\mathbf{G}_k$, а $\hat{\mathbf{C}}_k$ означает оценку вектора коэффициента $\mathbf{C}_k$.

Из (11.1.8) мы замечаем, что $\mathbf{G}_k$ равно обратной величине математического ожидания $\varepsilon_k \mathbf{V}_k^*$, следовательно, оценка $\mathbf{G}_k$ равна

$$\hat{\mathbf{G}}_k = -\varepsilon_k \mathbf{V}_k^*. \qquad (11.1.10)$$

Поскольку $E(\hat{\mathbf{G}}_k) = \mathbf{G}_k$, оценка $\hat{\mathbf{G}}_k$ является несмещённой оценкой правильного вектора градиента $\mathbf{G}_k$. Подстановка (11.1.10) в (11.1.9) даёт алгоритм

$$\hat{\mathbf{C}}_{k+1} = \hat{\mathbf{C}}_k + \Delta\varepsilon_k \mathbf{V}_k^*. \qquad (11.1.11)$$

Это базовый алгоритм НК (наименьших квадратов) для рекуррентной настройки коэффициентов шаговых весов эквалайзера, впервые предложенный Уидроу и Хоффом (1960). Он иллюстрируется в эквалайзере, показанном на рис.11.1.2.

Базовый алгоритм (11.1.11) и некоторые из его возможных вариантов были внедрены во многих коммерческих адаптивных эквалайзерах, которые используются в высокоскоростных модемах. Три варианта базового алгоритма были получены путём использования только информации о знаке, содержащейся в сигнале ошибки ε_k и (или) в компонентах $\mathbf{V}_k$. Таким образом, три возможных варианта алгоритма определяются так

$$c_{(k+1)j} = c_{kj} + \Delta \operatorname{csgn}(\varepsilon_k)\upsilon_{k-j}^*, \quad j = -K, \ldots, -1, 0, 1, \ldots, K \qquad (11.1.12)$$

$$c_{(k+1)j} = c_{kj} + \Delta\varepsilon_k \operatorname{csgn}(\upsilon_{k-j}^*), \quad j = -K, \ldots, -1, 0, 1, \ldots, K \qquad (11.1.13)$$

$$c_{(k+1)j} = c_{kj} + \Delta \operatorname{csgn}(\varepsilon_k)\operatorname{csgn}(\upsilon_{k-j}^*), \quad j = -K, \ldots, -1, 0, 1, \ldots, K, \qquad (11.1.14)$$

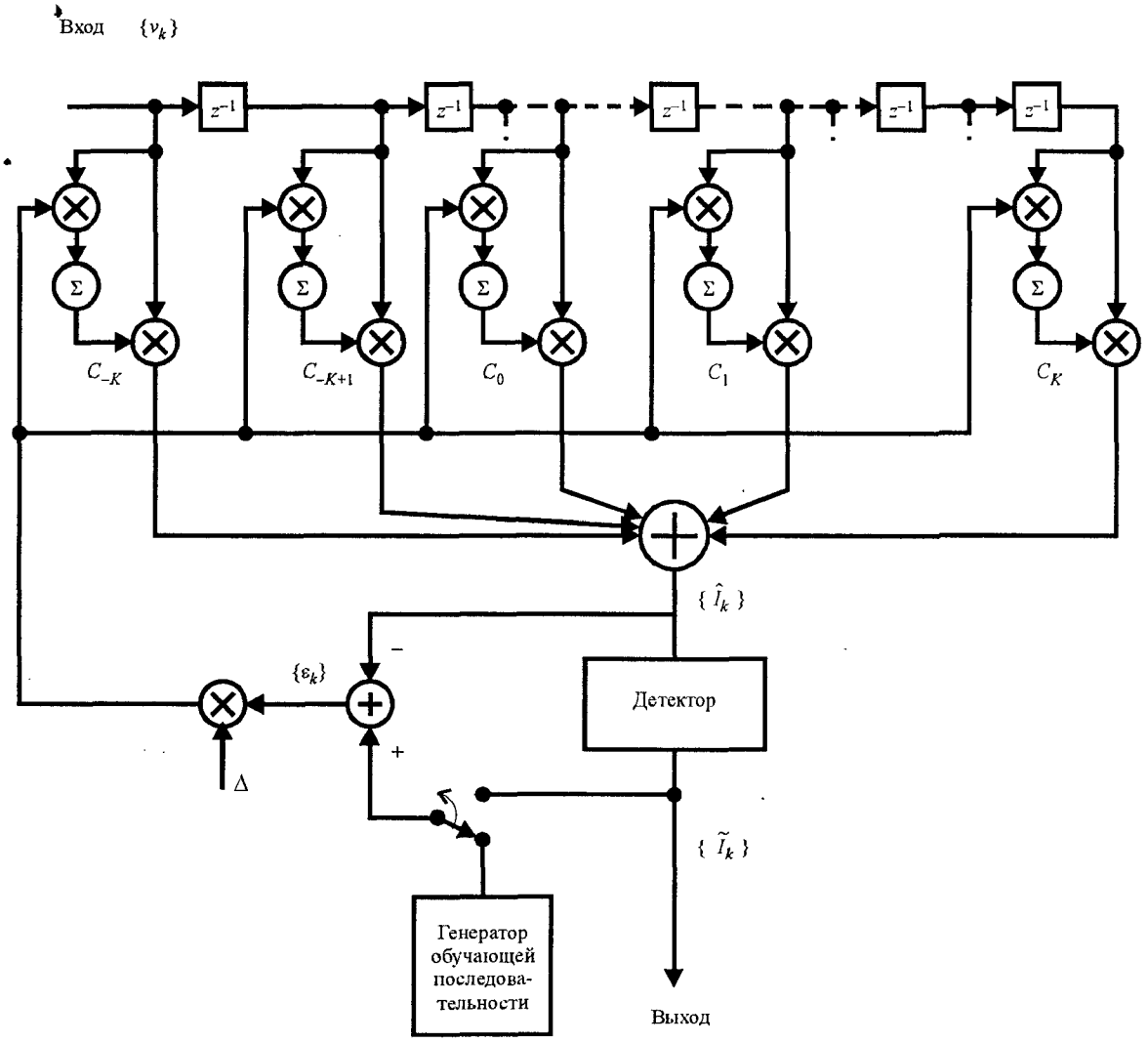


Рис. 11.1.2. Линейный адаптивный эквалайзер, основанный на критерии минимума СКО

где csgn(x) определяется так

$$\operatorname{csgn}(x)=\begin{cases}1+j & (\operatorname{Re}(x)>0,\ \operatorname{Im}(x)>0),\\ 1-j & (\operatorname{Re}(x)>0,\ \operatorname{Im}(x)<0),\\ -1+j & (\operatorname{Re}(x)<0,\ \operatorname{Im}(x)>0),\\ -1-j & (\operatorname{Re}(x)<0,\ \operatorname{Im}(x)<0),\end{cases} \qquad (11.1.15)$$

(Заметим, что в (11.1.15) $j=\sqrt{-1}$, что следует отличать от индекса j в (11.1.12)–(11.1.14)). Ясно, что алгоритм (11.1.14) реализуется существенно легче, но он дает наиболее медленную сходимость по сравнению с другими.

Несколько других вариантов алгоритма НК можно получить путем усреднения или фильтрации векторов градиентов по нескольким итерациям до выполнения настройки коэффициентов эквалайзера. Например, усреднение по N векторам градиентов дает

$$\overline{\hat{\mathbf{G}}}_{mN}=-\frac{1}{N}\sum_{n=0}^{N-1}\varepsilon_{mN+n}\mathbf{V}^{*}_{mN+n}, \qquad (11.1.16)$$

а соответствующее рекуррентное уравнение для адаптации коэффициентов эквалайзера по N итерациям имеет вид

$$\hat{\mathbf{C}}_{(k+1)N} = \hat{\mathbf{C}}_{kN} - \Delta\overline{\hat{\mathbf{G}}}_{kN}. \qquad (11.1.17)$$

Действительно, операция усреднения согласно (11.1.16) уменьшает шум в оценке вектора градиентов, как показано Гарднером (1987).

Альтернативный подход сводится к фильтрации шумовых составляющих градиента с помощью низкочастотного фильтра и использованию выхода фильтра как оценки вектора градиента. Например, простой низкочастотный фильтр для шумовых градиентов даёт выход

$$\overline{\hat{\mathbf{G}}}_k = w\overline{\hat{\mathbf{G}}}_{k-1} + (1-w)\hat{\mathbf{G}}_k, \quad \overline{\hat{\mathbf{G}}}(0) = \hat{\mathbf{G}}(0), \qquad (11.1.18)$$

где выбор $0 \le w < 1$ определяет полосу пропускания низкочастотного фильтра. Если w близко к единице, полоса фильтра мала и эффективное усреднение осуществляется по многим векторам градиентов. С другой стороны, если w мало, низкочастотный фильтр имеет большую полосу и, следовательно, он обеспечивает малое усреднение по векторам градиентов. С фильтрацией векторов градиентов по (11.1.18) вместо $\mathbf{G}_k$ мы получаем алгоритм НК с фильтрацией градиентов, определяемой так

$$\hat{\mathbf{C}}_{k+1} = \hat{\mathbf{C}}_k - \Delta\overline{\hat{\mathbf{G}}}_k. \qquad (11.1.19)$$

В вышеприведенном обсуждении было предположено, что приёмник имеет знание о переданной информационной последовательности при формировании сигналов ошибки между желательным символом и его оценкой. Такое знание можно получить в течение короткого периода обучения, когда к приемнику для первоначальной настройке шаговых весов передается известная информационная последовательность.

На практике часто в качестве обучающей последовательности выбирается периодическая псевдослучайная последовательность, такая как последовательность регистра сдвига максимальной длины с периодом N, равным длине (памяти) эквалайзера $(N = 2K+1)$. В этом случае, градиент обычно усредняется по длине последовательности как указанно в (11.1.6), а эквалайзер настраивается на каждом периоде согласно (11.1.17). Практическая схема для непрерывной настройки весов отводов может быть или основана на управлении решениями, когда решения (оценки) по информационным символам предполагаются правильными и используются вместо I_k при формировании сигнала ошибки ε_k, или такая, в которой известная псевдослучайная испытательная последовательность вставляется в информационный сигнал (путём суммирования или перемежения во времени), а веса отводов настраиваются путём сравнения принимаемых испытательных символов с известными переданными сигналами. При использовании операционной модели с управлением решениями сигнал ошибки оказывается равным $\tilde{\varepsilon}_k = \tilde{I}_k - \hat{I}_k$, где $\tilde{I}_k$ – это решение приёмника, основанное на оценке $\hat{I}_k$. До тех пор, пока приёмник работает с малыми вероятностями ошибок, редкие ошибки будут иметь несущественное влияние на сходимость алгоритма.

Если характеристики канала меняются, эти изменения отражаются в коэффициентах $\{f_k\}$ эквивалентной модели канала с дискретным временем. Они также отражаются в сигнале ошибки ε_k, поскольку он зависит от $\{f_k\}$. Таким образом, шаговые веса будут меняться согласно (11.1.11), отражая изменения в канале. Простое изменение в шаговых весах возникает, если меняется статистика шума или информационной последовательности. Таким образом, эквалайзер оказывается адаптивным.

11.1.3. Свойство сходимости алгоритма НК

Свойство сходимости алгоритма НК, определяемого (11.1.11), управляется параметром размера шага Δ. Мы теперь рассмотрим выбор параметра Δ для гарантии сходимости алгоритма кратчайшего спуска (11.1.7), который использует точные значения градиента.

Из (11.1.7) и (11.1.8) имеем

$$\mathbf{C}_{k+1} = \mathbf{C}_k - \Delta\mathbf{G}_k = (\mathbf{I} - \Delta\boldsymbol{\Gamma})\mathbf{C}_k + \Delta\boldsymbol{\xi}, \qquad (11.1.20)$$

где $\mathbf{I}$ – единичная матрица, $\boldsymbol{\Gamma}$ – автокорреляционная матрица принимаемого сигнала, $\mathbf{C}_k$ – $(2K+1)$ размерный вектор усиления ячеек эквалайзера, а $\boldsymbol{\xi}$ – вектор взаимных корреляций определяемый (10.2.45). Рекуррентное отношение (11.1.20) можно представить системой замкнутой петли управления, как показано на рис.11.1.3.

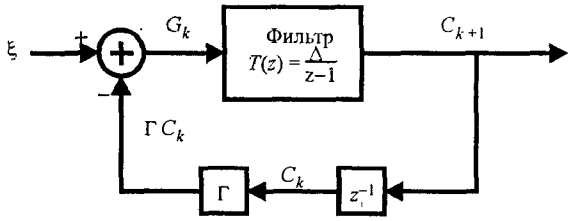


Рис. 11.1.3. Представление рекуррентного уравнения (11.1.20) системой с замкнутой петлёй управления

К сожалению, система из $2K+1$ разностных уровней первого порядка в (11.1.20) связана через матрицу автокорреляций $\boldsymbol{\Gamma}$. Чтобы решить эти уравнения и, таким образом, установить свойство сходимости рекуррентного алгоритма, математически удобно развязать уравнения путем линейного преобразования. Подходящее преобразование получается, если учесть, что матрица $\boldsymbol{\Gamma}$ является эрмитовой с комплексными элементами и, следовательно, её можно представить так:

$$\boldsymbol{\Gamma} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\mathrm{T}*}, \qquad (11.1.21)$$

где $\mathbf{U}$ – унитарная матрица для $\boldsymbol{\Gamma}$, $\boldsymbol{\Lambda}$ – диагональная матрица с элементами, равными собственным числам матрицы $\boldsymbol{\Gamma}$.

Если (11.1.21) подставить в (11.1.20) и если мы определим преобразованные (ортогонализированные) векторы $\mathbf{C}_k^0 = \mathbf{U}^{\mathrm{T}*}\mathbf{C}_k$ и $\boldsymbol{\xi}^0 = \mathbf{U}^{\mathrm{T}*}\boldsymbol{\xi}$, то получим

$$\mathbf{C}_{k+1}^0 = (\mathbf{I} - \Delta\boldsymbol{\Lambda})\mathbf{C}_k^0 + \Delta\boldsymbol{\xi}^0. \qquad (11.1.22)$$

Система дифференцированных уравнений первого порядка теперь развязана. Их сходимость определяется из однородного уравнения

$$\mathbf{C}_{k+1}^0 = (\mathbf{I} - \Delta\boldsymbol{\Lambda})\mathbf{C}_k^0. \qquad (11.1.23)$$

Мы видим, что рекуррентное отношение будет сходиться при условии, что все полюса лежат внутри единичной окружности, то есть

$$|1 - \Delta\lambda_k| < 1, \quad k = -K, \ldots, -1, 0, 1, \ldots, K, \qquad (11.1.24)$$

где $\{\lambda_k$ – набор из $2K+1$ (возможно, и не различающихся) собственных значений $\boldsymbol{\Gamma}$. Поскольку $\boldsymbol{\Gamma}$ автокорреляционная матрица она положительно определенная и, следовательно, $\lambda_k > 0$ для всех k. Следовательно, сходимость рекуррентного отношения (11.1.22) гарантируется, если Δ удовлетворяет неравенству

$$0 < \Delta < \frac{2}{\lambda_{\max}}, \qquad (11.1.25)$$

где $\lambda_{\max}$ – наибольшее собственное число $\boldsymbol{\Gamma}$.

Поскольку наибольшее собственное число положительно определенной матрицы меньше, чем сумма всех собственных чисел матрицы и, следовательно, поскольку сумма собственных чисел матрицы равна её следу, мы имеем следующую простую верхнюю границу для λ_{max}

$$\lambda_{max} < \sum_{k=-K}^{K} \lambda_k = \mathrm{tr}\mathbf{\Gamma} = (2K+1)\mathbf{\Gamma}_{kk} = (2K+1)(x_0 + N_0). \quad (11.1.26)$$

Из (11.1.23) и (11.1.24) мы видим, что быстрая сходимость возникает тогда, когда $|1-\Delta\lambda_k|$ мало, то есть когда полюсы далеки от единичной окружности. Но мы можем и не достичь это желательное условие и все же удовлетворить (11.1.25), если имеется большое различие между наибольшим и наименьшим собственными значениями матрицы $\mathbf{\Gamma}$. Другими словами, даже если мы выбираем Δ близким к верхней границе, даваемой (11.1.25), скорость сходимости рекуррентного алгоритма НК определяется наименьшим собственным числом λ_{min}. Следовательно, отношение $\lambda_{max}/\lambda_{min}$ непосредственно определяет скорость сходимости. Если $\lambda_{max}/\lambda_{min}$ мало, то Δ можно выбрать так, чтобы достичь быстрой сходимости. Однако, если отношение $\lambda_{max}/\lambda_{min}$ велико, что определяет случай, когда частотная характеристика канала имеет глубокие спектральные нули, скорость сходимости алгоритма будет медленной.

11.1.4. Излишек СКО, обусловленный зашумлёнными оценками градиентов

Рекуррентный алгоритм (11.1.11) для настройки коэффициентов в линейном эквалайзере использует несмещённые шумовые оценки вектора градиентов. Шум в этих оценках вызывает флуктуация коэффициентов около их оптимальных значений и, следовательно, ведет к увеличению СКО на выходе эквалайзера. Это означает, что финальное значение СКО равно $J_{min} + J_\Delta$, где J_Δ – дисперсия измеренного шума. Слагаемое J_Δ, обусловленное шумом оценки, было названо Уидроу (1966) *излишком среднеквадратичной ошибки*.

Суммарное СКО на выходе эквалайзера для некоторого набора коэффициентов $\mathbf{C}$ можно выразить так

$$J = J_{min} + (\mathbf{C} - \mathbf{C}_{opt})^{\mathrm{T}} * \mathbf{\Gamma}(\mathbf{C} - \mathbf{C}_{opt}), \quad (11.1.27)$$

где $\mathbf{C}_{opt}$ представляет оптимальные коэффициенты, удовлетворяющие (11.1.6). Это выражение для СКО можно упростить путем линейного ортогонального преобразования, использованного выше для установления сходимости. Результат этого преобразования к (11.1.27) даёт

$$J = J_{min} + \sum_{k=-K}^{K} \lambda_k E\,|\,c_k^0 - c^0{}_{k\,opt}\,|^2, \quad (11.1.28)$$

где $\{c_k^0\}$ – ансамбль преобразованных коэффициентов эквалайзера. Излишком СКО является величина второго слагаемого в (11.1.28), т.е.

$$J_\Delta = \sum_{k=-K}^{K} \lambda_k E\,|\,c_k^0 - c_{k\,opt}^0\,|^2 \quad (11.1.29)$$

Уидроу (1970, 1975) показал, что излишек СКО можно выразить так:

$$J_\Delta = \Delta^2 J_{min} \sum_{k=-K}^{K} \frac{\lambda_k^2}{1-(1-\Delta\lambda_k)^2}. \quad (11.1.30)$$

Выражение (11.1.30) можно упростить, если выбрать Δ так, что $\Delta\lambda_k << 1$ для всех k. Тогда

$$J_\Delta \approx \tfrac{1}{2}\Delta J_{\min}\sum_{k=-K}^{K}\lambda_k \approx \tfrac{1}{2}\Delta J_{\min}\,\mathrm{tr}\,\mathbf{\Gamma} \approx \tfrac{1}{2}\Delta(2K+1)J_{\min}(x_0+N_0). \qquad (11.1.31)$$

Заметим, что x_0+N_0 представляет принимаемый сигнал плюс мощность шума. Желательно иметь $J_\Delta < J_{\min}$. Это значит, что Δ надо выбрать так, чтобы

$$\frac{J_\Delta}{J_{\min}} \approx \tfrac{1}{2}\Delta(2K+1)(x_0+N_0) < 1,$$

или, что эквивалентно,

$$\Delta < \frac{2}{(2K+1)(x_0+N_0)}. \qquad (11.1.32)$$

Для примера, если Δ выбрать так:

$$\Delta = \frac{0{,}2}{(2K+1)(x_0+N_0)}, \qquad (11.1.33)$$

то уменьшение выходного ОСШ эквалайзера, обусловленное излишком СКО, меньше, чем 1 дБ.

Вышеприведённый анализ излишка среднеквадратичной ошибки базируется на предположении, что среднее значение коэффициентов эквалайзера сходится к оптимальной величине $\mathbf{C}_{\text{opt}}$. При этом условии размер шага Δ удовлетворяет границе (11.1.32). С другой стороны, мы определили, что сходимость вектора средних коэффициентов требует, чтобы $\Delta < 2/\lambda_{\max}$. В то время как выбор Δ вблизи верхней границы $2/\lambda_{\max}$ может вести к начальной сходимости детерминированного (известного) градиентного алгоритма крутого спуска, такая большая величина Δ обычно ведёт к нестабильности стохастического градиентного алгоритма НК.

Первоначальная сходимость или переходное поведение НК алгоритма была исследована различными исследователями. Их результаты ясно показывают, что размер шага должен быть уменьшен пропорционально длине эквалайзера, как следует из (11.1.32). Таким образом, верхняя граница, определяемая (11.1.32), также необходима, чтобы гарантировать первоначальную сходимость НК алгоритма. Статьи Гитлина и Вайнштейна (1979) и Унгербоека (1972) содержат анализ переходного поведения и свойства сходимости НК алгоритма.

Следующие примеры служат для подкрепления важных положений, высказанных выше, относительно первоначальной сходимости НК алгоритма.

Пример 11.1.1. НК алгоритм был использован для адаптивного выравнивания канала связи, для которого автокорреляционная матрица $\mathbf{\Gamma}$ имеет разброс собственных значений $\lambda_{\max}/\lambda_{\min} = 11$. Число выбранных ячеек эквалайзера $2K+1 = 11$. Входной сигнал плюс мощность шума x_0+N_0 был нормирован к единице. Как следствие, верхняя граница для Δ, определяемая (11.1.32), равна 0,18. Рис.11.1.4 иллюстрирует характеристики первоначальной сходимости НК алгоритма для Δ= 0,045, 0,09 и 0,115 при усреднении оценок СКО по 200 опытам. Мы видим, что при выборе Δ=0,09 (половина верхней границы) мы получаем относительно медленную сходимость. Если мы разделим Δ на два до Δ=0,045 скорость сходимости уменьшится, но излишек СКО также уменьшится, так что НК алгоритм работает лучше в стационарном режиме (при инвариантной во времени сигнальной среде). Наконец, мы отметим, что выбор Δ=0,115, который ещё остаётся ниже

верхней границы, вызывает большие нежелательные флуктуации выходной СКО алгоритма.

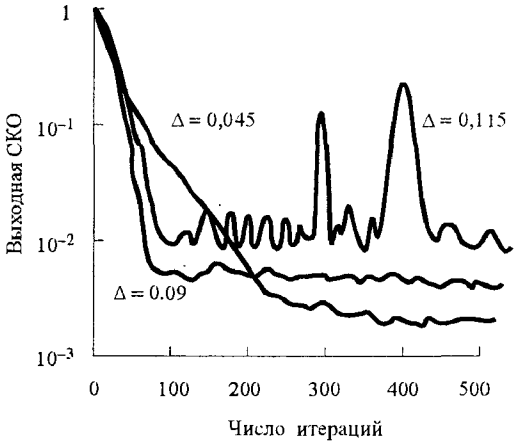


Рис. 11.1.4. Характеристика первоначальной сходимости НСК алгоритма при различных размерах шага. [Обработка цифровых сигналов, Дж.Дж.Прокис и Д.Дж.Монолакис, 1988]

При цифровой реализации НК алгоритма выбор параметра шага ячейки оказывается более критичным. В попытке уменьшить излишек СКО возможно уменьшить параметр размера шага ячейки до точки, когда суммарная СКО на самом деле увеличивается. Это условие возникает, когда оцененные градиентные компоненты вектора $\varepsilon_k \mathbf{V}_k^*$ после умножения на малый параметр шага Δ оказывается меньше, чем половина наименьших значащих бита в фиксированной точке представления коэффициентов эквалайзера. В таком случае, адаптация прекращается. Следовательно, важно, чтобы размер шага ячейки был достаточно большим для того, чтобы удержать коэффициенты эквалайзера в окрестности $\mathbf{C}_{\text{opt}}$. Если желательно существенно уменьшить размер шага ячейки, то необходимо увеличить точность коэффициентов эквалайзера. Типично, что для достаточно точного представления коэффициентов используется 16 бит, причём 10–12 наиболее значащих бита используется для арифметических операций по выравниванию сигнала. Оставшиеся наименее значащие биты требуются для обеспечения необходимой точности процесса адаптации. Таким образом, масштабированные оцененные градиентные компоненты $\Delta\varepsilon\mathbf{V}_k^*$ обычно влияют только на наименее значащие биты на любой итерации. В действительности, дополнительная точность также позволяет вести усреднение по шуму, поскольку много нарастающих изменений в наименее значимых битах требуется до того, как возникнет изменение в верхних более значащих битах, используемых в арифметических операциях для выравнивания данных. Для анализа округлённых (случайных) ошибок цифровой реализации НК алгоритма читателю рекомендуются статьи Гитлина и Вайнштейна (1979), Гитлина и др. (1982) и Карайскоса и Лайу (1984).

Наконец, необходимо указать, что НК алгоритм годится и для отслеживания медленных, инвариантных во времени, статистик сигнала. В таком случае минимум СКО и оптимальный вектор коэффициентов будут переменны во времени. Другими словами $J_{\min}(n)$ является функцией времени и $(2K+1)$-мерная поверхность ошибок передвигается с временным индексом n. НК алгоритм пытается следить за изменением минимума $J_{\min}(n)$ в $(2K+1)$-мерном пространстве, но он всегда запаздывает по отношению к значениям оцененных векторов градиента. Как следствие, НК алгоритм навлекает на себя другой вид

ошибок, называемых *ошибками запаздывания*, чьи значения средних квадратов уменьшаются с увеличением размера шага ячейки Δ. Суммарное СКО можно теперь выразить так

$$J_{\Sigma} = J_{\min}(n) + J_{\Delta} + J_{l},$$

где J_l означает средний квадрат ошибки, обусловленный запаздыванием. При заданной нестационарной адаптивной задаче выравнивания, если мы построим зависимости ошибок J_{Δ} и J_l от Δ, мы ожидаем поведение этих ошибок так, как показано на рис.11.1.5. Видим, что J_{Δ} увеличивается с ростом Δ, в то время как J_e уменьшается с ростом Δ. Суммарная ошибка имеет минимум, который определяет оптимальный выбор параметра размера шага ячейки.

Когда случайные изменения сигнала во времени происходят быстро, ошибка запаздывания будет определяющей для качества адаптивного эквалайзера. В таком случае $J_l >> J_{\min} + J_{\Delta}$ даже тогда, когда используется наибольшее возможное значение Δ. Если это условие имеет место, НК алгоритм противопоказан для применения и необходимо для получения более быстрой сходимости и отслеживания рассчитывать на более сложные рекуррентные алгоритмы наименьших квадратов, описываемые в разд. 11.4.

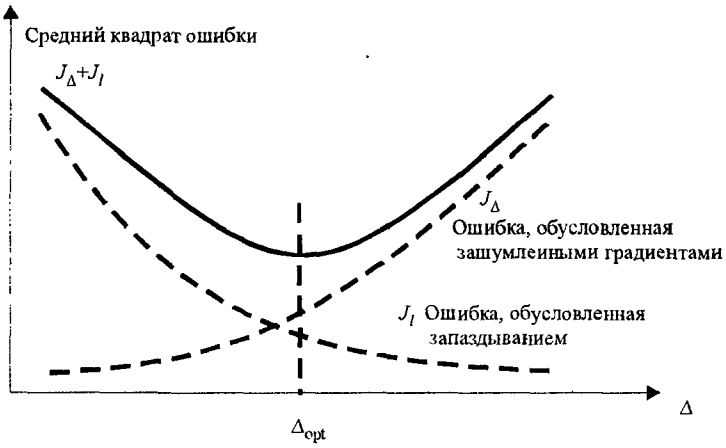


Рис. 11.1.5. Зависимость излишка среднеквадратичной ошибки и ошибки запаздывания от размера шага ячейки. [Обработка цифровых сигналов, Дж.Дж.Прокис и Д.Дж.Манолакис., 1988]

11.1.5. Линейные эквалайзеры для базовых и полосовых сигналов

Наша трактовка адаптивных линейных эквалайзеров были выполнена через эквивалентные низкочастотные сигналы. Однако в практических приложениях адаптивный эквалайзер, показанный на рис.11.1.2, можно реализовать или с базовыми или полосовыми сигналами. Например, рис.11.1.6 иллюстрирует демодуляцию сигналов КАМ (или многофазовый ФМ) при первоначальной передаче базовых сигналов и при выравнивании базовых сигналов при помощи эквалайзера с комплексными коэффициентами.

В действительности, комплексный эквалайзер с комплексными величинами (синфазных и квадратурных компонентов) по входу эквивалентен четырём параллельным эквалайзерам с вещественными коэффициентами ячеек, как показано на рис.11.1.7.

В качестве альтернативы, мы можем выравнивать сигнал как полосовой. Это выполняется, так, как показано на рис.11.1.8 для двухмерного сигнального созвездия как при КАМ и ФМ. Принимаемый сигнал фильтрируется и, параллельно, он проходит через преобразователь Гильберта, называемый *фазо-расщепляющим фильтром*.

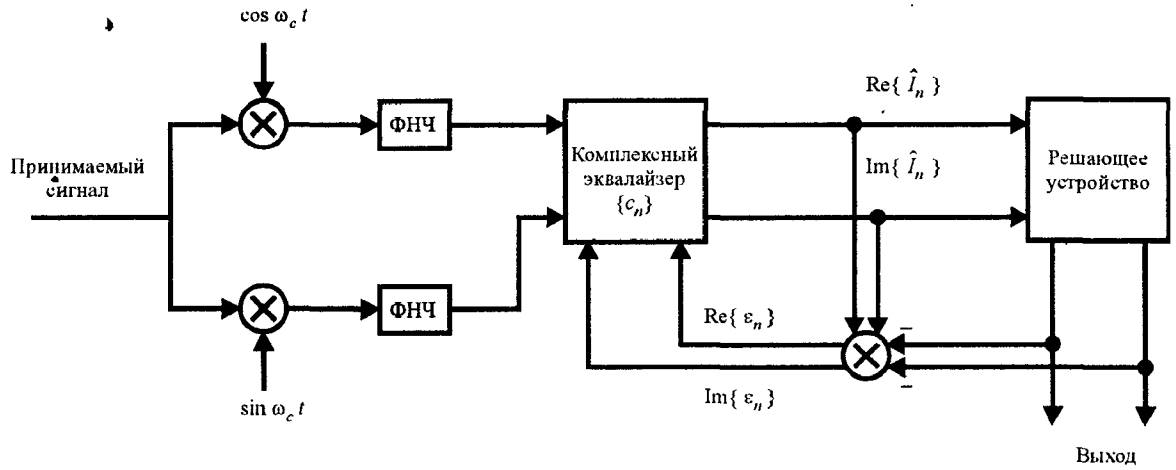


Рис. 11.1.6. Демодуляция сигналов КАМ

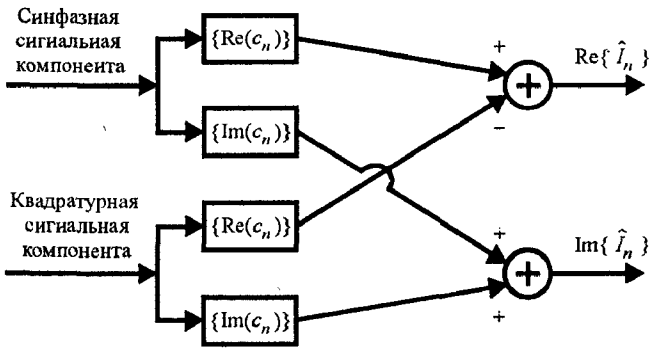


Рис.11.1.7. Комплексный эквалайзер для базового сигнала КАМ

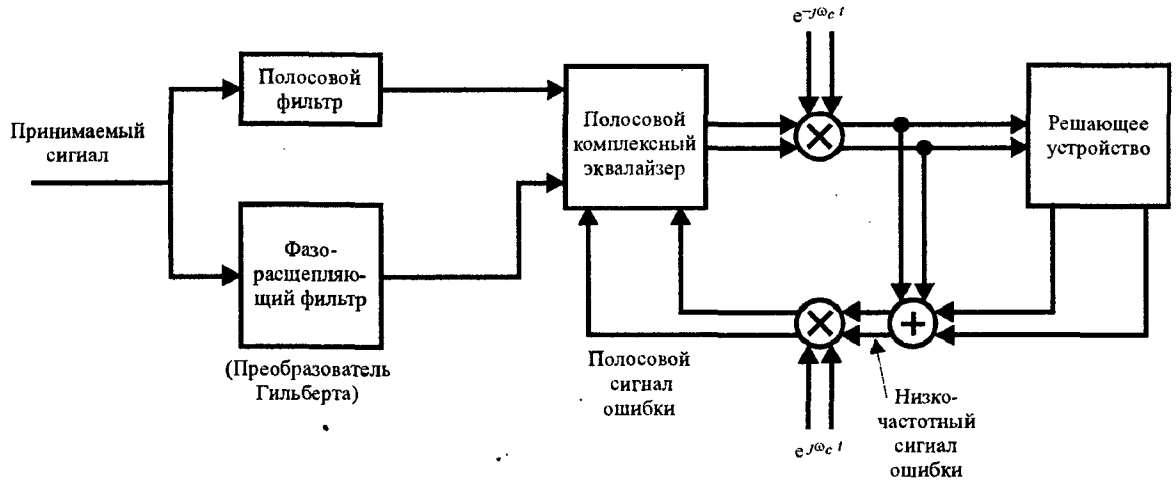


Рис.11.1.8. Выравнивание полосовых сигналов КАМ или ФМ

Таким образом, мы имеем эквивалент синфазных и квадратурных компонент в полосовом сигнале, который питает полосовой комплексный эквалайзер. После эквалайзера сигнал обратно превращается в базовый и детектируется. Сигнал ошибок, генерируемый для целей настройки коэффициентов эквалайзера, формируется, как базовый и преобразуется по частоте в полосовой, как показано на рис.11.1.8.

11.2. АДАПТИВНЫЙ ЭКВАЛАЙЗЕР С ОБРАТНОЙ СВЯЗЬЮ ПО РЕШЕНИЮ

Как в случае линейного адаптивного эквалайзера, коэффициенты фильтра прямой связи и обратной связи в эквалайзере с обратной связью по решению можно настроить рекуррентно вместо обращения матрицы, как это выполнено в (10.3.3). Основанный на минимизации СКО на выходе ЭООСР, алгоритм кратчайшего спуска принимает форму

$$\mathbf{C}_{k+1} = \mathbf{C}_k + \Delta E(\varepsilon_k \mathbf{V}_k^*), \qquad (11.2.1)$$

где $\mathbf{C}_k$ – вектор коэффициентов эквалайзера на k-ом сигнальном интервале, $E(\varepsilon_k \mathbf{V}_k^*)$ – взаимная корреляция сигнала ошибки с $\varepsilon_k = I_k - \hat{I}_k$ с $\mathbf{V}_k$, а $\mathbf{V}_k = [\upsilon_{k+K_1} \dots \upsilon_k^{\cdot} I_{k-1} \dots I_{k-K_2}]^{\mathrm{T}}$ – представляет значения сигналов в фильтрах прямой и обратной связи в момент $t = kT$. СКО минимизируется. когда вектор взаимной корреляции $E(\varepsilon_k \mathbf{V}_k^*) = 0$ при $k \to \infty$.

Поскольку точное значение вектора взаимной корреляции в любой момент времени неизвестен мы используем в качестве оценки вектор $\varepsilon_k \mathbf{V}_k^*$ и усредняем шум оценки посредством рекуррентного отношения

$$\hat{\mathbf{C}}_{k+1} = \hat{\mathbf{C}}_k + \Delta \varepsilon_k \mathbf{V}_k^*. \qquad (11.2.2)$$

Этот алгоритм НК для ЭООСР.

Как в случае линейного эквалайзера мы можем использовать обучающую последовательность для первоначальной настройки коэффициентов ЭООСР. Путём сходимости к ближайшим оптимальным коэффициентам (по минимуму СКО) мы можем переключиться на модель с управлением решениями, причём решения на выходе детектора используются для формирования сигнала ошибки ε_k, который и питает фильтр обратной связи. Это адаптивный вариант ЭООСР, который иллюстрируется на рис.11.2.1. В этом случае рекуррентное уравнение для настройки коэффициентов эквалайзера равно

$$\tilde{\mathbf{C}}_{k+1} = \tilde{\mathbf{C}}_k + \Delta \tilde{\varepsilon}_k \mathbf{V}_k^* \qquad (11.2.3)$$

$$\tilde{\varepsilon}_k = \tilde{I}_k - \hat{I}_k \quad \mathbf{V}_k = [\upsilon_{k+K_1} \dots \upsilon_k \; I_{k-1} \dots I_{k-K_2}]^{\mathrm{T}}.$$

Характеристики качества алгоритма НК для ЭООСР по существу такие же, как те, которые даны в разделах 11.1.3 и 11.1.4 для линейного адаптивного эквалайзера.

11.2.1. Адаптивное выравнивание для решётчато-кодированных сигналов

Эффективная по полосе частот решетчато-кодированная модуляция, которая была описана в разделе 8.3, часто используется в цифровой связи по телефонным каналам для уменьшения требуемого ОСШ на символ для достижения заданной вероятности ошибки. Канальные искажения решетчато-кодированных сигналов заставляет использовать адаптивное выравнивание для уменьшения межсимвольной интерференции. Выход эквалайзера затем подаётся на декодер Витерби, который выполняет декодирование мягких решений решётчато- кодированных символов.

Возникает вопрос о том, как в таком приёмнике адаптировать эквалайзер в режиме передачи данных? Одна возможность заключается в том, чтобы эквалайзер делал свои собственные решения на своем выходе исключительно для целей генерирования сигнала ошибки для настройки своих коэффициентов ячеек, как показано на блок-схеме рис.11.2.2.

Проблема, возникающая при таком подходе, заключается в том, что такие решения обычно нереализуемы, поскольку ОСШ при предварительном декодировании кодового символа относительно низкое, значение вероятности ошибки вызывает существенные нарушения в работе эквалайзера, которые в конечном счёте воздействуют на надёжность решений на выходе декодера.

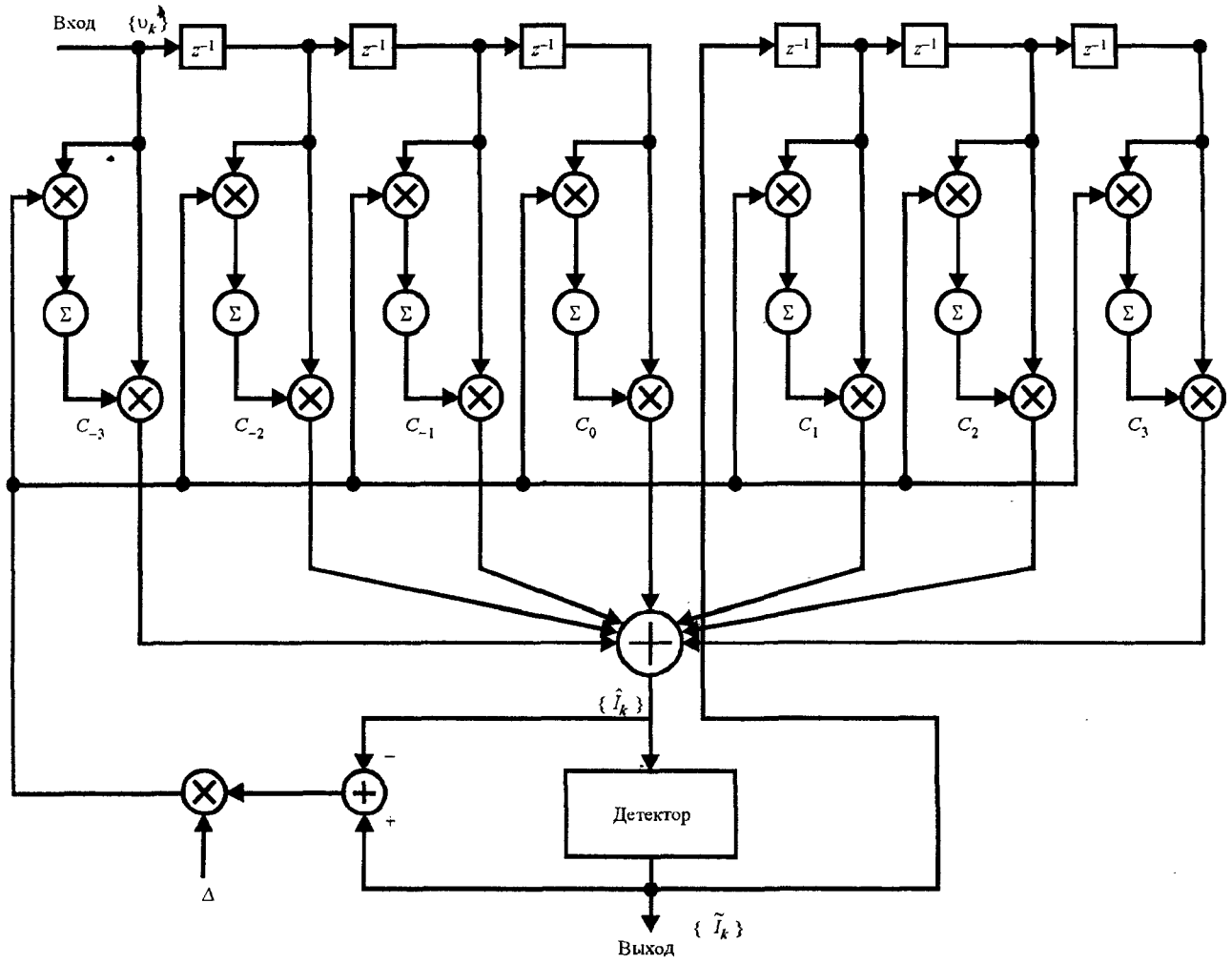


Рис.11.2.1. Эквалайзер с обратной связью по решению

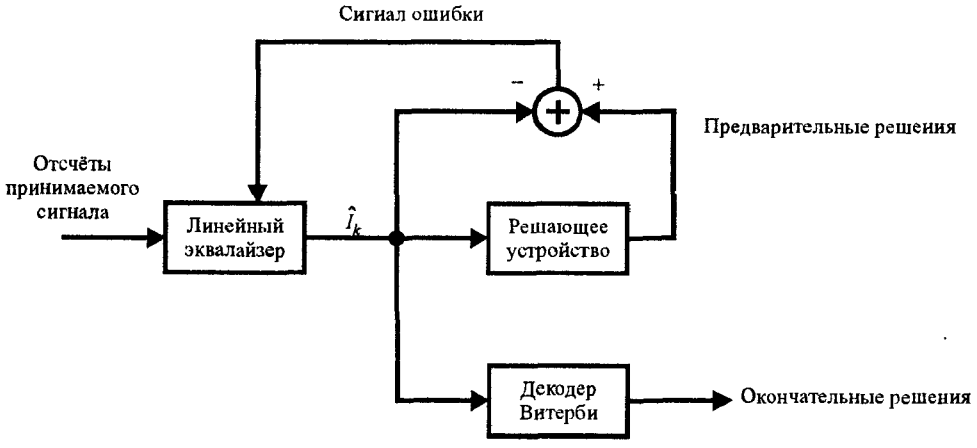


Рис.11.2.2. Настройка эквалайзера, основанная на предварительных решениях

Более желательная альтернатива сводится к использованию для непрерывной адаптации эквалайзера решений декодера Витерби, которые намного более надёжны. Такой подход, конечно, более предпочтителен и жизнеспособен, когда линейный эквалайзер используется до декодера Витерби. Задержку при декодировании, присущую

декодеру Витерби, можно преодолеть путём введения идентичной задержки в настраиваемых весах ячеек для коэффициентов эквалайзера, как показано на рис.11.2.3. Основную цену, которую нужно платить за дополнительную задержку сводится к уменьшению параметра размера шага в алгоритме НК, для достижения стабильности алгоритма, как описано Лонгом и др.(1987, 1989).

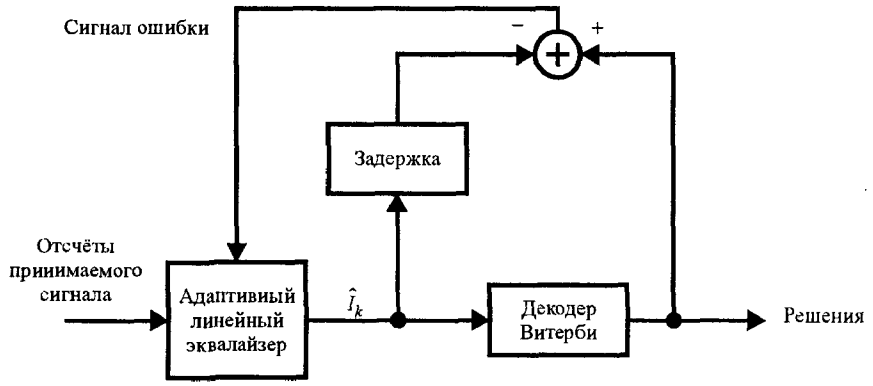


Рис.11.2.3. Настройка эквалайзера, основанного на решениях декодера

В каналах с одним или большим числом спектральных нулей по полосе частот, когда линейный эквалайзер не годится для компенсации канальной МСИ, мы используем ЭОСР. Но ЭОСР требует надёжные решения в своём фильтре обратной связи для того, чтобы погасить МСИ от предварительно продетектированных символов. Предварительные решения до декодирования в большей степени ненадёжны и, следовательно, не подходящие. К сожалению, общепринятый ЭОСР не может быть каскадно соединён с алгоритмом Витерби, чтобы решение после декодера вернуть по цепи обратной связи в ЭОСР.

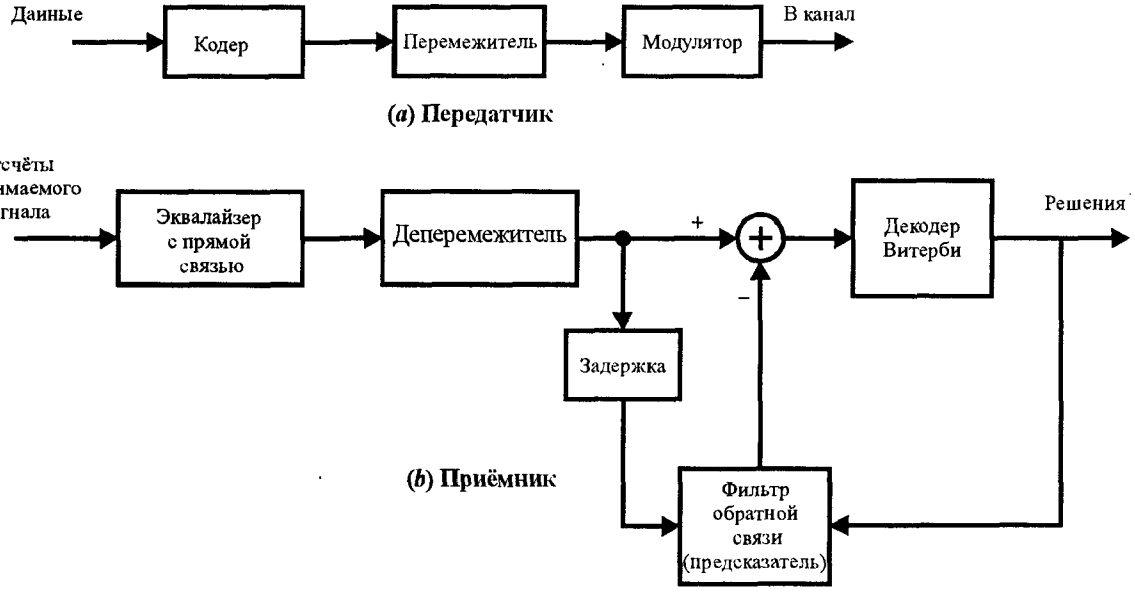


Рис.11.2.4. Использование ЭОСР с предсказанием с перемежением и решёточно-кодовой модуляцией

Одна альтернатива сводится к использованию ЭОСР с предсказанием, описанного в разд.10.3.3. Для того, чтобы приспособиться к задержке при декодировании, поскольку она влияет на линейный предсказатель, мы вводим в систему периодическую пару перемежитель-деперемежитель, которые имеют ту же задержку, что и декодер Витерби, и

таким образом делают возможным генерировать соответствующий сигнал ошибки для предсказателя, как показано на блок-схеме рис.11.2.4.

Новый путь комбинирования ЭОСР с предсказанием с декодером Витерби для выравнивания решётчато-кодированых сигналов описан и проанализирован Эйбоглу (1988). Та же идея использована для выравнивания многопутевых каналов с замираниями Зоу и др. (1988, 1990), но структура ЭОСР была модифицирована для использования рекурсивных минимально-квадратичных лестничных фильтров, которые обеспечивают более быструю адаптацию к изменениям во времени, возникающим в канале.

11.3. АДАПТИВНЫЙ ОЦЕНИВАТЕЛЬ КАНАЛА ДЛЯ МАКСИМАЛЬНО ПРАВДОПОДОБНОГО ПОСЛЕДОВАТЕЛЬНОГО ДЕТЕКТИРОВАНИЯ

Правило МП последовательного детектирования, реализованное алгоритмом Витерби, воплощенное в расчёте метрик (10.1.23), и вероятностный алгоритм посимвольного детектирования, описанный в разделе 5.1.5, требуют знания коэффициентов $\{f_k\}$ эквивалентного канала с дискретным временем. Чтобы приспособить их к неизвестному каналу или каналу с медленными изменениями во времени, можно включить оцениватель канала, работающий параллельно с алгоритмом детектирования, как показано на рис.11.3.2.

Оцениватель канала, показанный на рис.11.3.2 идентичен по структуре линейному трансверсальному эквалайзеру, обсужденному ранее в разделе 11.1.

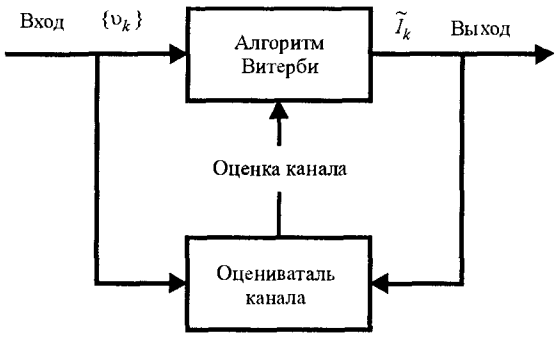


Рис.11.3.1. Блок- схема метода оценивания характеристик канала для алгоритма Витерби

Действительно, такой оцениватель канала является воплощением (replica?) эквивалентного канала с дискретным временем, который моделирует МСИ. Оцененные коэффициенты ячеек, обозначаемые $\{\hat{f}_k\}$, подстраиваются рекуррентно для минимизации СКО между действительно принимаемой последовательностью и выходом оценивателя. Для примера, алгоритм кратчайшего спуска в варианте выполнения операций с управлением решениями можно записать так:

$$\hat{\mathbf{f}}_{k+1} = \hat{\mathbf{f}}_k + \Delta\varepsilon_k \tilde{\mathbf{I}}_k^*, \qquad (11.3.1)$$

где $\hat{\mathbf{f}}_k$ – вектор коэффициентов усиления ячеек на k-й итерации, Δ – размер шага ячейки, $\varepsilon_k = \upsilon_k - \hat{\upsilon}_k$ – сигнал ошибки, а $\tilde{\mathbf{I}}_k$ означает вектор детектированных информационных символов в оценивателе канала на к-й итерации.

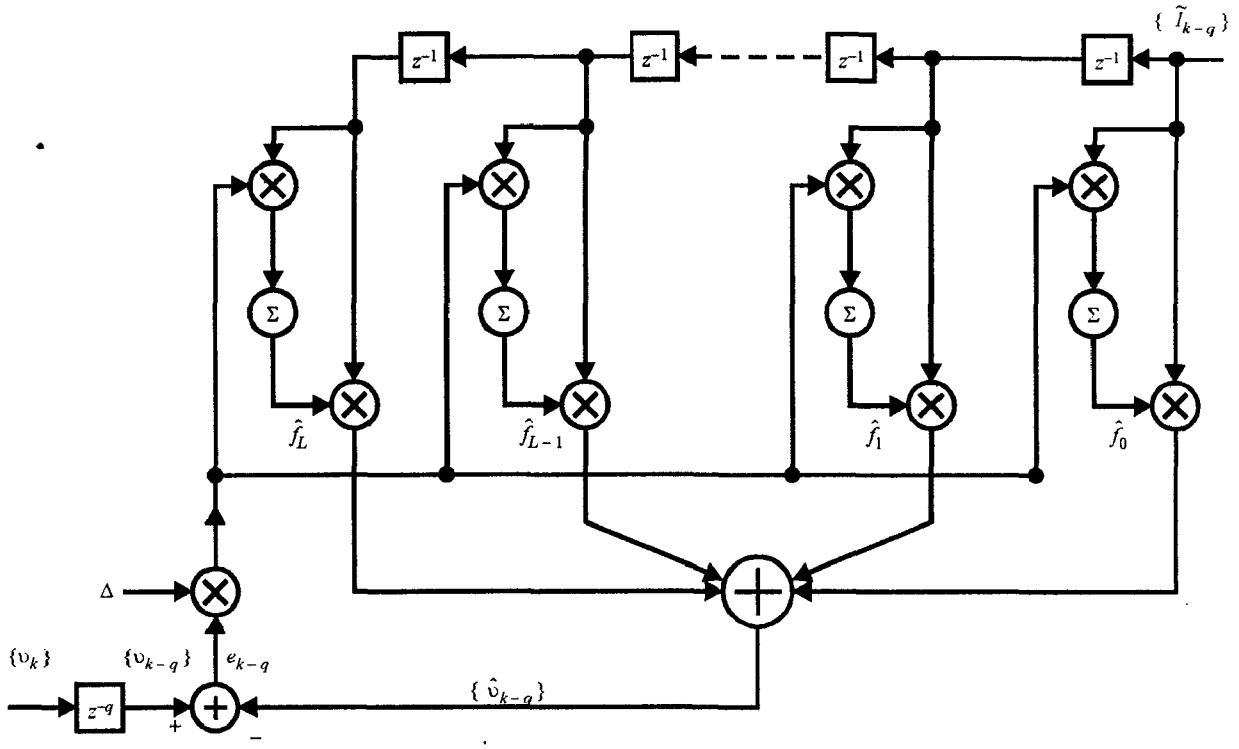


Рис.11.3.2. Адаптивный трансверсальный фильтр для оценивания канального рассеяния

Теперь покажем. что когда СКО между υ_k и $\hat{\upsilon}_k$ минимизируется, результирующие значения коэффициентов усиления ячеек оценивателя канала являются величинами модели канала с дискретным временем. Для математической трактуемости предположим, что продетектированная информационная последовательность $\{\tilde{I}_k\}$ правильная, то есть $\{\tilde{I}_k\}$ идентична переданной последовательности $\{I_k\}$. Это разумное предположение, когда система работает при малой вероятности ошибок. Таким образом, СКО между принимаемым сигналом υ_k и оценкой $\hat{\upsilon}_k$ равна

$$J(\hat{\mathbf{f}}) = E\left(\left|\upsilon_k - \sum_{j=0}^{N-1} \hat{f}_j I_{k-j}\right|^2\right). \qquad (11.3.2)$$

Коэффициенты ячеек $\{\hat{f}_k\}$, которые минимизируют $J(\hat{\mathbf{f}})$ в (11.3.2), удовлетворяют системе из N линейных уравнений

$$\sum_{j=0}^{N-1} \hat{f}_j \phi_{kj} = d_k, \qquad k = 0, 1, \ldots, N-1, \qquad (11.3.3)$$

где

$$\phi_{kj} = E(I_k I_j^*), \quad d_k = \sum_{j=0}^{N-1} f_j \phi_j . \qquad (11.3.4)$$

Из (11.3.3) и (11.3.4) следует, что пока информационная последовательность некоррелирована, оптимальные коэффициенты точно равны соответствующим величинам эквивалентного канала дискретного времени. Также очевидно. что когда число ячеек оценивателя канала N больше или равно $L+1$, оптимальные коэффициенты усиления ячеек

$\{\hat{f}_k\}$ равны соответствующим значением $\{f_k\}$, даже если информационная последовательность коррелированна. С учетом указанных выше условий, минимум СКО просто равно дисперсии шума N_0.

В вышеприведенном обсуждении, оцененная информационная последовательность на выходе алгоритма Витерби или вероятностного посимвольного алгоритма была использована для выполнения настройки оценивателя канала. Для начала операции можно послать короткую обучающую последовательность для формирования начальной настройки коэффициентов ячеек, как это обычно делается в случае линейного трансверсального эквалайзера. При адаптивном варианте обработки сигналов приемник просто использует свои собственные решения для формирования сигнала ошибки.

11.4. РЕКУРРЕНТНЫЕ АЛГОРИТМЫ МИНИМАЛЬНЫХ КВАДРАТОВ ДЛЯ АДАПТИВНОГО ВЫРАВНИВАНИЯ

НК алгоритм, который был описан в разделе 11.1 и 11.2 для адаптивной настройки коэффициентов ячеек линейного эквалайзера или ЭОСР является по существу (стохастическим) алгоритмом кратчайшего спуска, в котором правильный вектор градиентов аппроксимируется оценкой, полученной непосредственно по данным.

Важнейшее преимущество алгоритма кратчайшего спуска определяется его вычислительной простотой. Однако плата за простоту – медленная сходимость особенно, если характеристики канала отражены в матрице автокорреляций $\mathbf{\Gamma}$, чьи собственные значения имеют большой разброс, т.е. $\lambda_{max}/\lambda_{min} >> 1$. Если посмотреть с другой точки зрения то градиентный алгоритм имеет единственный настраиваемый параметр для управления скорости сходимости, а именно параметр Δ. Следовательно, медленная сходимость обусловлена фундаментальным ограничением.

Для получения быстрой сходимости необходимо разработать более сложные алгоритмы, включающие дополнительные параметры. В частности, если матрица $\mathbf{\Gamma}$ размером $N \times N$ имеет собственные значения $\lambda_1, \lambda_2, ..., \lambda_N$, мы можем использовать алгоритм, который содержит N параметров – один для каждого собственного значения, Оптимальный выбор этих параметров для достижения быстрой сходимости является темой этого раздела.

При разработке алгоритмов быстрой сходимости мы будем использовать подход минимальных квадратов. Таким образом, мы будем работать непосредственно с принимаемыми данными для минимизации квадратичного показателя качества, в то время как раньше мы минимизировали ожидаемую величину среднеквадратичной ошибки. Проще говоря, это значит, что показатель качества выражается через временное среднее вместо статистического среднего.

Удобно выразить рекуррентные алгоритмы наименьших квадратов в матричном виде. С этой целью определим некоторое число векторов и матриц, которые необходимы в этом исследовании. Так поступая, мы не значительно изменим привычные обозначения. Конкретнее, оценку информационных символов в момент t, где t целое, в линейном эквалайзере теперь выразим так:

$$\tilde{I}(t) = \sum_{j=-K}^{K} c_j(t-1)\upsilon_{t-j} .$$

Изменяя индекс j для $c_j(t-1)$ от $j=0$ до $j=N-1$ и определив на время

$$y(t) = \upsilon_{t+K},$$

оценку $\tilde{I}(t)$ можно выразить так

$$\tilde{I}(t)=\sum_{j=0}^{N-1}c_j(t-1)y(t-j)=\mathbf{C}_N^{\mathrm{T}}(t-1)\mathbf{Y}_N(t), \qquad (11.4.1)$$

где $\mathbf{G}_N(t-1)$ и $\mathbf{Y}_N(t)$ – соответственно векторы-столбцы коэффициентов эквалайзера $c_j(t-1)$, $j=0,1,\ldots,N-1$ и входных сигналов $y(t-j)$, $j=0,1,2,\ldots,N-1$.

Аналогично, в эквалайзере с обратной связью по решению мы имеем коэффициенты ячеек $c_j(t)$, $j=0,1,\ldots,N-1$, где первые K_1+1 является коэффициентами фильтра с прямой связью, а оставшиеся $K_2=N-K_1-1$ являются коэффициентами фильтра обратной связи. Данные в оценке $\hat{I}(t)$ являются $\upsilon_{t+K_1}\ldots\upsilon_{t+1}$, $\tilde{I}_{t-1}\ldots\tilde{I}_{t-K_2}$, где $\tilde{I}_{t-j}$, $1\le j\le K_2$ означают оценки по ранее продетектированным символам. В этом исследовании мы пренебрегаем влиянием ошибочных решений в алгоритме. Итак, мы считаем, что $\tilde{I}_{t-j}=I_{t-j}$, $1\le j\le K_2$.

Для удобства обозначений мы также определим

$$y(t-j)=\begin{cases}\upsilon_{t+K_1-j} & (0\le j\le K_1)\\ I_{t+K_1-j} & (K_1<j\le N-1)\end{cases}. \qquad (11.4.2)$$

Следовательно,

$$\mathbf{Y}_N(t)=\begin{bmatrix}y(t) & y(t-1)\ldots & y(t-N+1)\end{bmatrix}^{\mathrm{T}}=\begin{bmatrix}\upsilon_{t+K_1}\ldots\upsilon_{t+1} & \upsilon_t & I_{t-1}\ldots I_{t-K_2}\end{bmatrix}^{\mathrm{T}}. \qquad (11.4.3)$$

11.4.1. Рекуррентный алгоритм (Калмана) наименьших квадратов

Рекуррентное по наименьшим квадратам (РНК) оценивание $\hat{I}(t)$ можно сформулировать следующим образом. Допустим, мы наблюдаем векторы $\mathbf{Y}_N(n)$, $n=0,1,\ldots,t$ и желаем определить вектор коэффициентов $\mathbf{C}_N(t)$ эквалайзера (линейного или с обратной связью по решению), который минимизирует усреднённые во времени взвешенные квадраты ошибок

$$\mathscr{E}_N^{\mathrm{НК}}=\sum_{n=0}^{t}w^{t-n}\,|\,e_N(n,t)\,|^2, \qquad (11.4.4)$$

где ошибки определяются так

$$e_N(n,t)=I(n)-\mathbf{C}_N^{\mathrm{T}}(t)\mathbf{Y}_N(n), \qquad (11.4.5)$$

а w представляет взвешивающий множитель $0<w<1$. Таким образом, мы ввели показательное взвешивание по последним данным, которое приемлемо, когда характеристики канала меняются во времени. Минимизация $\mathscr{E}_N^{MK}$ по вектору коэффициентов $\mathbf{C}_N(t)$ ведет к системе линейных уравнений

$$\mathbf{R}_N(t)\mathbf{C}_N(t)=\mathbf{D}_N(t), \qquad (11.4.6)$$

где $\mathbf{R}_N(t)$ – матрица корреляций сигнала, определенная так

$$\mathbf{R}_N(t)=\sum_{n=0}^{t}w^{t-n}\mathbf{Y}_N^{*}(n)\mathbf{Y}_N^{\mathrm{T}}(n), \qquad (11.4.7)$$

а $\mathbf{D}_N(t)$ – вектор взаимных корреляций

$$\mathbf{D}_N(t)=\sum_{n=0}^{t}w^{t-n}I(n)\mathbf{Y}_N^{*}(n). \qquad (11.4.8)$$

Решение (11.4.6) равно

$$\mathbf{C}_N(t) = \mathbf{R}_N^{-1}(t)\mathbf{D}_N(t). \qquad (11.4.9)$$

Матрица $\mathbf{R}_N(t)$ родственна матрице статистических автокорреляций $\mathbf{\Gamma}_N$, в то время как вектор $\mathbf{D}_N(t)$ родственен вектору ξ_N взаимных корреляций, определенных раньше. Подчеркнем, однако, что $\mathbf{R}_N(t)$ не является тёплицевой матрицей. Мы также хотим напомнить, что для малых значений t, $\mathbf{R}_N(t)$ может быть плохо обусловленной; следовательно, это обычно приводит к первоначальной добавке матрицы $\delta\mathbf{I}_N$ к $\mathbf{R}_N(t)$, где δ – малая положительная константа, а $\mathbf{I}_N$ – единичная матрица. При показательном взвешивании по последним данным влияние прибавления $\delta\mathbf{I}_N$ ослабевает со временем.

Теперь предположим, что мы имеем решение (11.4.9) для момента $t-1$, то есть $\mathbf{C}_N(t-1)$, и мы желаем вычислить $\mathbf{C}_N(t)$. Неэффективно и, следовательно, непрактично решать систему N линейных уравнений для каждой новой сигнальной компоненты, которая принимается. Чтобы преодолеть эту трудность, мы поступим так. Сначала $\mathbf{R}_N(t)$ можно вычислить рекуррентно следующим образом

$$\mathbf{R}_N(t) = w\mathbf{R}_N(t-1) + \mathbf{Y}_N^*(t)\mathbf{Y}_N^{\mathrm{T}}(t). \qquad (11.4.10)$$

Мы назовем формулу (11.4.10) обновляющее уравнение для $\mathbf{R}_N(t)$.

Поскольку в (11.4.9) требуется обращение $\mathbf{R}_N(t)$, используем соотношение для обращённых матриц

$$\mathbf{R}_N^{-1}(t) = \frac{1}{w}\left[\mathbf{R}_N^{-1}(t-1) - \frac{\mathbf{R}_N^{-1}(t-1)\mathbf{Y}_N^*(t)\mathbf{Y}_N^{\mathrm{T}}(t)\mathbf{R}_N^{-1}(t-1)}{w + \mathbf{Y}_N^{\mathrm{T}}(t)\mathbf{R}_N^{-1}(t-1)\mathbf{Y}_N^*(t)}\right]. \qquad (11.4.11)$$

Таким образом, $\mathbf{R}_N^{-1}(t)$ можно вычислить рекуррентно согласно (11.4.11).

Для удобства определим $\mathbf{P}_N(t) = \mathbf{R}_N^{-1}(t)$. Также удобно определить N-мерный вектор, называемый *вектор калмановского усиления,* так

$$\mathbf{K}_N(t) = \frac{1}{w + \mu_N(t)}\mathbf{P}_N(t-1)\mathbf{Y}_N^*(t), \qquad (11.4.12)$$

где $\mu_N(t)$ является скаляром, определяемым так

$$\mu_N(t) = \mathbf{Y}_N^T(t)\mathbf{P}_N(t-1)\mathbf{Y}_N^*(t). \qquad (11.4.13)$$

С этими определениями (11.4.11) приводится к виду

$$\mathbf{P}_N(t) = \frac{1}{w}\left[\mathbf{P}_N(t-1) - \mathbf{K}_N(t)\mathbf{Y}_N^t(t)\mathbf{P}_N(t-1)\right]. \qquad (11.4.14)$$

Предположим, что мы умножаем обе части (11.4.14) на $Y_N^*(t)$. Тогда

$$\mathbf{P}_N(t)\mathbf{Y}_N^*(t) = \frac{1}{w}\left[\mathbf{P}_N(t-1)\mathbf{Y}_N^*(t) - \mathbf{K}_N(t)\mathbf{Y}_N^{\mathrm{T}}(t)\mathbf{P}_N(t-1)\mathbf{Y}_N^*(t)\right] =$$

$$= \frac{1}{w}\left\{[w + \mu_N(t)]\mathbf{K}_N(t) - \mathbf{K}_N(t)\mu_N(t)\right\} = \mathbf{K}_N(t). \qquad (11.4.15)$$

Следовательно, вектор калмановского усиления можно также определить как $\mathbf{P}_N(t)\mathbf{Y}_N(t)$.

Теперь используем соотношение для обратных матриц для получения уравнения, определяющего $\mathbf{C}_N(t)$ по $\mathbf{C}_N(t-1)$. Поскольку

$$\mathbf{C}_N(t) = \mathbf{P}_N(t)\mathbf{D}_N(t)$$

и

$$\mathbf{D}_N(t) = w\mathbf{D}_N(t-1) + I(t)\mathbf{Y}_N^*(t), \qquad (11.4.16)$$

имеем

$$\mathbf{C}_N(t)=\frac{1}{w}\left[\mathbf{P}_N(t-1)-\mathbf{K}_N(t)\mathbf{Y}_N^{\mathrm{T}}(t)\mathbf{P}_N(t-1)\right]\left[w\mathbf{D}_N(t-1)+I(t)\mathbf{Y}_N^{*}(t)\right]$$

$$=\mathbf{P}_N(t-1)\mathbf{D}_N(t-1)+\frac{1}{w}I(t)\mathbf{P}_N(t-1)\mathbf{Y}_N^{*}(t)$$

$$-\mathbf{K}_N(t)\mathbf{Y}_N^{\mathrm{T}}(t)\mathbf{P}_N(t-1)\mathbf{D}_N(t-1)$$

$$-\frac{1}{w}I(t)\mathbf{K}_N(t)\mathbf{Y}_N^{\mathrm{T}}(t)\mathbf{P}_N(t-1)\mathbf{Y}_N^{*}(t)$$

$$=\mathbf{C}_N(t-1)+\mathbf{K}_N(t)[I(t)-\mathbf{Y}_N^{\mathrm{T}}(t)\mathbf{C}_N(t-1)] \qquad (11.4.17)$$

Заметим, что $\mathbf{Y}_N^{t}(t)\mathbf{C}_N(t-1)$ это выход эквалайзера в момент t, то есть

$$\hat{I}(t)=\mathbf{Y}_N^{\mathrm{T}}(t)\mathbf{C}_N(t-1) \qquad (11.4.18)$$

и

$$e_N(t,t-1)=I(t)-\hat{I}(t)\equiv e_N(t) \qquad (11.4.19)$$

– это ошибка между желательным символом и его оценкой. Таким образом, $\mathbf{C}_N(t)$ определяется рекуррентно согласно отношению

$$\mathbf{C}_N(t)=\mathbf{C}_N(t-1)+\mathbf{K}_N(t)e_N(t). \qquad (11.4.20)$$

Остаточный СКО, получаемый при такой оптимизации, равен

$$\mathscr{E}_{N\min}^{MK}=\sum_{n=0}^{t}w^{t-n}\,|I(n)|^2-\mathbf{C}_N^{\mathrm{T}}(t)\mathbf{D}_N^{*}(t). \qquad (11.4.21)$$

Для суммирования предположим, что мы имеем $\mathbf{C}_N(t-1)$ и $\mathbf{P}_N(t-1)$. Когда принимается новая сигнальная компонента имеем $\mathbf{Y}_N(t)$. Затем рекуррентные вычисления для обновления $\mathbf{C}_N(t)$ и $\mathbf{P}_N(t)$ продолжаются так:

– вычисляется выход:

$$\hat{I}(t)=\mathbf{Y}_N^{\mathrm{T}}(t)\mathbf{C}_N(t-1),$$

– вычисляется ошибка:

$$e_N(t)=I(t)-\hat{I}(t),$$

– вычисляется вектор калмановского усиления:

$$\mathbf{K}_N(t)=\frac{\mathbf{P}_N(t-1)\mathbf{Y}_N^{\mathrm{T}}(t)}{w+\mathbf{Y}_N^{\mathrm{T}}(t)\mathbf{P}_N(t-1)\mathbf{Y}_N^{*}(t)},$$

– вычисляется соотношение обратных матриц

$$\mathbf{P}_N(t)=\tfrac{1}{w}[\mathbf{P}_N(t-1)-\mathbf{K}_N(t)\mathbf{Y}_N^{\mathrm{T}}(t)\mathbf{P}_N(t-1)],$$

– определяются коэффициенты

$$\mathbf{C}_N(t)=\mathbf{C}_N(t-1)+\mathbf{K}_N(t)e_N(t)=\mathbf{C}_N(t-1)+\mathbf{P}_N(t)\mathbf{Y}_N^{*}(t)e_N(t). \qquad (11.4.22)$$

Алгоритм, определяемый (11.4.22), назван прямой формой РНК или алгоритмом Калмана. Он подходит, когда эквалайзер имеет трансверсальную (прямая форма) структуру. Заметим, что коэффициенты эквалайзера меняются со временем на величину, равной ошибки $e_N(t)$, уменьшенной на вектор усиления Калмана $\mathbf{K}_N(t)$. Поскольку $\mathbf{K}_N(t)$ является N-мерным каждый коэффициент ячейки в действительности управляется одним из элементов $\mathbf{K}_N(t)$. Как следствие получается быстрая сходимость. В противоположность этому, алгоритм кратчайшего спуска, выраженный в новых обозначениях, определяется так

$$\mathbf{C}_N(t) = \mathbf{C}_N(t-1) + \Delta \mathbf{Y}_N^*(t) e_N(t), \qquad (11.4.23)$$

и единственным переменным параметром является размер шага ячейки Δ.

Рис.11.4.1 иллюстрирует начальную скорость сходимости этих двух алгоритмов для канала с фиксируемыми параметрами $f_0 = 0{,}26$, $f_1 = 0{,}93$, $f_2 = 0{,}26$ и линейного эквалайзера с 11 ячейками. Отношение собственных значений для этого канала $\lambda_{max} / \lambda_{min} = 11$. Все коэффициенты эквалайзера были первоначально обнулены. Алгоритм кратчайшего спуска был реализован с $\Delta = 0{,}020$. Превосходство алгоритма Калмана очевидно. Это особенно важно при отслеживании меняющегося во времени канала. Для примера, изменения во времени высокочастотного (ВЧ или КВ) ионосферного радиоканала слишком быстрые, чтобы их выравнивать градиентным алгоритмом, но алгоритм Кальмана адаптируется достаточно быстро для отслеживания таких изменений.

Несмотря на прекрасные качества отслеживания алгоритм Калмана, описанный выше, имеет два недостатка. Один – его сложность, второй – его чувствительность к случайному шуму, который накапливается при рекуррентных операциях. Последний может вызвать нестабильность алгоритма.



Рис. 11.4.1. Сравнение скорости сходимости алгоритма Калмана и градиентного алгоритма.

Число вычислений или операций (умножений, делений и вычитаний) при расчете величин (11.4.22), пропорционально N^2. Большинство из этих операций используется при определении $\mathbf{P}_N(t)$. Эта часть вычислений также чувствительна к случайному шуму. Чтобы решить эту проблему были разработаны алгоритмы, которые избегают вычисления $\mathbf{P}_N(t)$ согласно (11.4.14). Основа этих алгоритмов сводится к декомпозиции $\mathbf{P}_N(t)$ в виде

$$\mathbf{P}_N(t) = \mathbf{S}_N(t) \mathbf{\Lambda}_N(t) \mathbf{S}_N^{\mathrm{T}}(t), \qquad (11.4.24)$$

где $\mathbf{S}_N(t)$ – нижняя треугольная матрица, чьи диагональные элементы – единицы, а $\mathbf{\Lambda}_N(t)$ – диагональная матрица. Такая декомпозиция называется *факторизацией* (см. Бирман, 1977). Эта факторизация описывается в приложении D. В алгоритме факторизации $\mathbf{P}_N(t)$ не обновляется, как в (11.4.14), а вычисляется. Вместо этого при помощи рекуррентного обновления формируются $\mathbf{S}_N(t)$ и $\mathbf{\Lambda}_N(t)$.

Алгоритмы НК часто используются в системах управления, которые включают в себя калмановскую фильтрацию. В цифровой связи, алгоритм НК Калмана используется в ФМ модеме с выравниванием на основе обратной связи по решению, спроектированном для передачи с высокой скоростью по ВЧ каналу с номинальной полосой частот 3 кГц. Этот алгоритм описан в статье Хшу (1982). Он имеет вычислительную сложность порядка $1{,}5N^2 + 6{,}5N$ (умножения комплексных величин и делений на выходные символы). Для

подробного ознакомления с алгоритмами НК в последовательном оценивании читателю рекомендуется книга Бирмана (1977).

Возможно также разработать РНК алгоритмы с вычислительной сложностью, которая возрастает линейно с числом коэффициентов эквалайзера N. Такие алгоритмы обычно называются *быстрыми РНК алгоритмами*, и они были описаны в статьях Караяниса и др. (1983), Чиффи и Кайлата (1989) и Слока и Кайлата (1988).

11.4.2. Линейное предсказание и лестничные фильтры

В главе 3 мы рассмотрели линейное предсказание сигнала в плане кодирования речи. В этом разделе мы хотим установить связь межу линейным предсказанием и лестничным фильтром.

Проблему линейного предсказания можно сформулировать так: по значениям набора данных $y(t-1), y(t-2), \ldots, y(t-p)$ надо предсказать значение данных в последующей точке $y(t)$. Предсказатель порядка p определяется так

$$\hat{y}\ (t) = \sum_{k=1}^{p} a_{pk} y(t-k). \tag{11.4.25}$$

Минимизация СКО

$$\mathscr{E}_p = E\left[y(t) - \hat{y}\ (t)\right]^2 = E\left[\ y(t) - \sum_{k=1}^{p} a_{pk} y(t-k)\right]^2 \tag{11.4.26}$$

по коэффициентам предсказания $\{a_{pk}\}$ ведет к системе линейных уравнений

$$\sum_{k=1}^{p} a_{pk}\phi(k-l) = \phi(l), \quad l = 1, 2, \ldots, p, \tag{11.4.27}$$

где

$$\phi(l) = E[y(t)y(t+l)].$$

Их называют *нормальными уравнениями Юли-Волкера.*

Матрица коэффициентов $\mathbf{\Phi}$ с элементами $\phi(k-l)$ является теплицевой матрицей. Следовательно, алгоритм Левинсона-Дурбина, описанный в приложения А, дает эффективный способ для рекуррентного решения линейных уравнений, начиная с предсказателя первого порядка и продолжая рекуррентно для нахождения коэффициентов предсказателя порядка p. Рекуррентные соотношения для алгоритма Левинсона-Дурбина таковы

$$\begin{aligned} a_{11} &= \frac{\phi(1)}{\phi(0)}, \qquad \mathscr{E}_0 = \phi(0) \\ a_{mm} &= \frac{\phi(m) - \mathbf{A}_m^{\mathrm{T}} \boldsymbol{\varphi}_{m-1}^{r}}{\mathscr{E}_{m-1}} \\ a_{mk} &= a_{m-1\,k} - a_{mm} a_{m-1\,m-k} \\ \mathscr{E}_m &= \mathscr{E}_{m-1}(1 - a_{mm}^2) \end{aligned} \tag{11.4.28}$$

для $m = 1, 2, \ldots, p$, причём векторы $\mathbf{A}_{m-1}$ $\boldsymbol{\varphi}_{m-1}^{r}$ определены так:

$$\mathbf{A}_{m-1} = [a_{m-11} \quad a_{m-12} \quad \ldots \quad a_{m-1\,m-1}]^{\mathrm{T}},$$

$$\boldsymbol{\varphi}_{m-1}^{r} = [\phi(m-1) \quad \phi(m-2) \quad \ldots \quad \phi(1)]^{\mathrm{T}}.$$

Линейный предсказывающий фильтр порядка m можно реализовать как трансверсальный фильтр с передаточной функцией

$$A_m(z) = 1 - \sum_{k=1}^{m} a_m z^{-k} \tag{11.4.29}$$

Его входом являются данные $\{y(t)\}$, а его выходом – ошибка $e(t) = y(t) - \hat{y}(t)$ Предсказывающий фильтр можно также реализовать в лестничном виде, как мы теперь покажем.

Начнём с использования алгоритма Левинсона-Дурбина для коэффициентов предсказателя a_{mk} в (11.4.29). Подстановка даёт

$$\begin{aligned} A_m(z) &= 1 - \sum_{k=1}^{m-1} (a_{m-1\,k} - a_{mm} a_{m-1\,m-k}) z^{-k} - a_{mm} z^{-m} \\ &= 1 - \sum_{k=1}^{m-1} (a_{m-1\,k} z^{-k} - a_{mm} z^{-m} \left(1 - \sum_{k=1}^{m-1} a_{m-1\,k} z^{k} \right) \\ &= A_{m-1}(z) - a_{mm} z^{-m} A_{m-1}(z^{-1}) . \end{aligned} \tag{11.4.30}$$

Таким образом, мы получили передаточную функцию предсказателя m-го порядка через передаточную функцию $(m-1)$-го порядка. Теперь предположим, что мы определяем фильтр с передаточной функцией $G_m(z)$, равной

$$G_m(z) = z^{-m} A_m(z^{-1}) . \tag{11.4.31}$$

Тогда (11.4.30) можно выразить так

$$A_m(z) = A_{m-1}(z) - a_{mm} z^{-1} G_{m-1}(z) . \tag{11.4.32}$$

Заметим, что $G_{m-1}(z)$ представляет трансверсальный фильтр с коэффициентами в отводах $(-a_{m-1\,m-1}, -a_{m-1\,m-2}, \ldots, -a_{m-1\,1}, 1)$, в то время как коэффициенты для $A_{m-1}(z)$ такие же за исключением того, что они даны в обратном порядке.

Больше понимания соотношения между $A_m(z)$ и $G_m(z)$ можно получить путем вычисления выхода этих двух фильтров при подачи ко входу последовательности $y(t)$. Используя z-преобразование, имеем

$$A_m(z)Y(z) = A_{m-1}(z)Y(z) - a_{mm} z^{-1} G_{m-1}(z)Y(z) . \tag{11.4.33}$$

Выходы фильтров мы определяем так

$$\begin{aligned} F_m(z) &= A_m(z)Y(z) \\ B_m(z) &= G_m(z)Y(z). \end{aligned} \tag{11.4.34}$$

Тогда (11.4.33) можно записать

$$F_m(z) = F_{m-1}(z) - a_{mm} z^{-1} B_{m-1}(z) . \tag{11.4.35}$$

Во временной области соотношение (11.4.35) можно выразить так

$$f_m(t) = f_{m-1}(t) - a_{mm} b_{m-1}(t-1), \qquad m \geq 1, \tag{11.4.36}$$

где

$$f_m(t) = y(t) - \sum_{k=1}^{m-1} a_{mk} y(t-k) , \tag{11.4.37}$$

$$b_m(t) = y(t-m) - \sum_{k=1}^{m-1} a_{mk} y(t-m+k) . \tag{11.4.38}$$

Для детальной разработки отметим, что $f_m(t)$ в (11.4.37) представляет ошибку предсказания m-го порядка по более ранним отсчётам (ошибка вперёд), в то время как $b_m(t)$ представляет ошибку предсказания m-го порядка по более поздним отсчётам (ошибка назад).

Соотношение (11.4.36) – одно из двух, определяющих лестничный фильтр. Второе соотношение получается из $G_m(z)$ следующим образом:

$$G_m(z) = z^{-m} A_m(z^{-1}) = z^{-m}[A_{m-1}(z^{-1}) - a_{mm} z^m A_{m-1}(z)] = z^{-1} G_{m-1}(z) - a_{mm} A_{m-1}(z). \quad (11.4.39)$$

Теперь мы умножим левую и правые части (11.4.39) на $Y(z)$ и выразим результаты через $F_m(z)$ и $B_m(z)$, используя определение (11.4.34), мы получим

$$B_m(z) = z^{-1} B_{m-1}(z) - a_{mm} F_{m-1}(z). \quad (11.4.40)$$

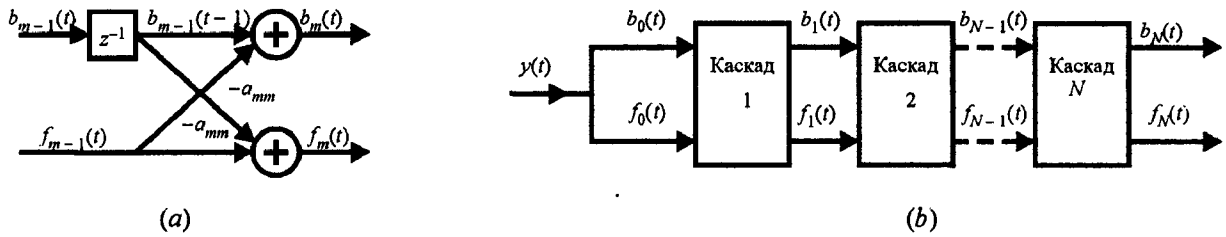
Путем преобразования (11.4.40) во временную область мы получим второе отношение, которое соответствует лестничному фильтру, а именно

$$b_m(t) = b_{m-1}(t-1) - a_{mm} f_{m-1}(t), \qquad m \geq 1. \quad (11.4.41)$$

Начальные условия

$$f_0(t) = b_0(t) = y(t). \quad (11.4.42)$$

Лестничные фильтры, описанные рекуррентными отношениями (11.4.36) и (11.4.40), иллюстрируются на рис.11.4.2.



(*a*) (*b*)

Рис. 11.4.2. Лестничный фильтр.

Каждый каскад характеризуется собственным коэффициентом умножения $\{a_{ii}\}$, $i = 1, 2, \ldots, N$, который определяются алгоритмом Левинсона-Дурбина. Ошибки вперёд и назад $f_m(t)$ и $b_m(t)$ обычно называют *остатками*. Средний квадрат этих остатков равен

$$\mathscr{E}_m = E\left[f_m^2(t)\right] = E\left[b_m^2(t)\right], \quad (11.4.43)$$

а $\mathscr{E}_m$ определяется рекуррентно согласно алгоритму Левинсона-Дурбина:

$$\mathscr{E}_m = \mathscr{E}_{m-1}(1 - a_{mm}^2) = \mathscr{E}_0 \prod_{i=1}^{m} (1 - a_{ii}^2), \quad (11.4.44)$$

где

$$\mathscr{E}_0 = \phi(0).$$

Остатки $\{f_m(t)\}$ и $\{b_m(t)\}$ удовлетворяют ряду интересных свойств, как описано Макхоулом (1978). Наиболее важные из них – свойства ортогональности

$$\begin{aligned} E[b_m(t) b_n(t)] &= \mathscr{E}_m \delta_{mn}, \\ E[f_m(t+m) f_n(t+n)] &= \mathscr{E}_m \delta_{mn}. \end{aligned} \quad (11.4.45)$$

Далее, взаимные корреляции между $f_m(t)$ и $b_n(t)$ определяются так

$$E[f_m(t) b_n(t)] = \begin{cases} a_{nn} \mathscr{E}_m & (m \geq n) \\ 0 & (m < n) \end{cases} \qquad m, n \geq 0. \quad (11.4.46)$$

Вследствие свойств ортогональности остатков, различные секции лестницы проявляют форму независимости, которая позволяет нам прибавить или удалить одну или больше последних каскадов без влияния на параметры оставшихся каскадов. Поскольку остаточный средний квадрат ошибки $\mathscr{E}_m$ уменьшается монотонно с числом секций, $\mathscr{E}_m$

можно использовать как показатель качества того, каким числом ячеек можно ограничиться.

Из вышеприведенного обсуждения мы видим, что линейный предсказывающий фильтр можно реализовать или как линейный трансверсальный фильтр или как лестничный фильтр. Лестничный фильтр рекуррентен по порядку и, как следствие, число его секций (каскадов) можно легко увеличить или уменьшить без влияния на параметры оставшихся секций. В противоположность этому коэффициенты трансверсального фильтра, полученные на основе РНК, взаимозависимы. Это значит, что увеличение или уменьшение размера фильтра приведет к изменению всех коэффициентов. Следовательно, алгоритм Калмана, описанный в разделе 11.4.1, рекуррентен во времени, но не по порядку реализующих его звеньев.

Основываясь на оптимизации по критерию наименьших квадратов, лестничные алгоритмы РНК были разработаны так, что их вычислительная сложность растет линейно с ростом числа коэффициентов фильтра N (числа каскадов). Таким образом, структура лестничного эквалайзера в вычислительном отношении конкурентоспособна с алгоритмом быстрого РНК эквалайзера прямой формы. Лестничные алгоритмы РНК описаны в статьях Морфа и др. (1973), Саториуса и Александера (1979). Саториуса и Пака (1981), Линга и Прокиса (1984) и Линга и др. (1986).

Лестничные алгоритмы РНК имеют несомненное будущее, поскольку они численно нечувствительны по отношению к случайным ошибкам, свойственным цифровой реализации алгоритма. Трактовку их количественных характеристик можно найти в статьях Линга и других (1984, 1986).

11.5. СЛЕПОЕ ВЫРАВНИВАНИЕ

В общепринятых эквалайзерах, работающих по алгоритму сведения к нулю или минимума СКО, мы предположили, что к приёмнику передаётся известная обучающая последовательность для целей начальной настройки коэффициентов эквалайзера. Однако имеется ряд приложений, таких как многопользовательские сети связи, когда желательно для приёмника синхронизироваться от принимаемого сигнала и настроить эквалайзер без использования обучающей последовательности.

Техника выравнивания, основанная на первоначальной настройке коэффициентов без использования обучающей последовательности, названа *самонастраивающейся* или *слепой*.

Начиная со статьи Сато (1975), за последние два десятилетия были разработаны три различных класса адаптивных алгоритмов слепого выравнивания. Один класс алгоритмов основан на методе кратчайшего спуска для адаптации эквалайзера. Второй класс алгоритмов основан на использовании статистики второго и более высокого порядка (обычно четвёртого порядка) принимаемого сигнала для оценки начальных характеристик и синтеза эквалайзера. Совсем недавно был изобретён третий класс алгоритмов слепого выравнивания, основанный на правиле максимального правдоподобия. В этом разделе мы вкратце опишем эти подходы и дадим несколько относящихся к теме ссылок на литературу.

11.5.1. Слепое выравнивание, основанное на критерии максимального правдоподобия

Удобно использовать эквивалентную модель канала с дискретным временем, описанную в разделе 10.1.2. Напомним, что выход этой модели канала с МСИ равен

$$\upsilon_n = \sum_{k=0}^{L} f_k I_{n-k} + \eta_n, \qquad (11.5.1)$$

где $\{f_k\}$– коэффициенты эквивалентного канала с дискретным временем, $\{I_n\}$ представляет информационную последовательность, а $\{\eta_n\}$ последовательность отсчетов белого гауссовского шума. Для блока из N принимаемых сигнальных точек совместная ФПВ (вектора $\mathbf{v} = [\upsilon_1\ \upsilon_2 \ldots \upsilon_n]^{\mathrm{T}}$) при условии известного вектора импульсной характеристики канала $\mathbf{f} = [f_0\ f_1 \ldots f_L]^{\mathrm{T}}$ и известного вектора данных $\mathbf{I} = [I_1\ I_2 \ldots I_N]^{\mathrm{T}}$, равна

$$p(\mathbf{v}\,|\,\mathbf{f},\mathbf{I}) = \frac{1}{(2\pi\sigma^2)^N} \exp\left(-\frac{1}{2\sigma^2} \sum_{n=1}^{N} \left| \upsilon_n - \sum_{k=0}^{L} f_k I_{n-k} \right|^2 \right). \qquad (11.5.2)$$

Совместные максимально правдоподобные оценки $\mathbf{f}$ и $\mathbf{I}$ – это такие значения этих векторов, которые максимизирует совместную ФПВ $p(\mathbf{v}\,|\,\mathbf{f},\mathbf{I})$ или, что эквивалентно, это величины $\mathbf{f}$ и $\mathbf{I}$, которые минимизируют показатель экспоненты. Таким образом, максимально правдоподобное решение определяется минимумом по $\mathbf{f}$ и $\mathbf{I}$ метрики

$$DM(\mathbf{I},\mathbf{f}) = \sum_{n=1}^{N} \left| \upsilon_n - \sum_{k=0}^{L} f_k I_{n-k} \right|^2 = \|\mathbf{v} - \mathbf{A}\mathbf{f}\|^2, \qquad (11.5.3)$$

где матрица $\mathbf{A}$ называется матрицей данных и определяется так

$$\mathbf{A} = \begin{bmatrix} I_1 & 0 & 0 & \ldots & 0 \\ I_2 & I_1 & 0 & \ldots & 0 \\ I_3 & I_2 & I_1 & \ldots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ I_N & I_{N-1} & I_{N-2} & \ldots & I_{N-L} \end{bmatrix}. \qquad (11.5.4)$$

Мы сделаем несколько наблюдений. Прежде всего заметим, что, когда вектор данных $\mathbf{I}$ (или матрица данных $\mathbf{A}$) известен, как в случае, когда на приёме используется обучающая последовательность, максимально правдоподобная оценка импульсной характеристики канала, полученная минимизацией (11.5.3) по $\mathbf{f}$, равна

$$\mathbf{f}_{\text{МП}}(\mathbf{I}) = (\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{v}. \qquad (11.5.5)$$

С другой стороны, когда импульсная характеристика канала известна, оптимальной МП детектор для последовательности данных $\mathbf{I}$ осуществляет поиск по решётке (или поиск по дереву), используя алгоритм Витерби для канала с МСИ.

Если не известны как $\mathbf{I}$ так и $\mathbf{f}$ минимизацию показателя качества $DM(\mathbf{I},\mathbf{f})$ можно выполнить совместно по $\mathbf{I}$ и $\mathbf{f}$. Альтернативно $\mathbf{f}$ можно оценить по ФПВ $p(\mathbf{v}\,|\,\mathbf{f})$, которую можно получить усреднением $p(\mathbf{v},\mathbf{f}\,|\,\mathbf{I})$ по всем возможным последовательностям данных

$$p(\mathbf{v}\,|\,\mathbf{f}) = \sum_{m} p(\mathbf{v},\mathbf{I}^{(m)}\,|\,\mathbf{f}) = \sum_{m} p(\mathbf{v}\,|\,\mathbf{I}^{(m)},\mathbf{f}) P(\mathbf{I}^{(m)}), \qquad (11.5.6)$$

где $P(\mathbf{I}^{(m)})$ – вероятность последовательности $\mathbf{I} = \mathbf{I}^{(m)}$, $m = 1, 2, \ldots, M^N$, а M – размер символьного созвездия.

Оценка канала, основанная на усреднении последовательностей данных. Как указанно в приведенном выше обсуждении, когда и $\mathbf{I}$ и $\mathbf{f}$ не известны один из подходов сводится к оценке импульсной характеристики $\mathbf{f}$ после усреднения ФПВ $p(\mathbf{v},\mathbf{I}\,|\,\mathbf{f})$ по всем последовательностям данных. Таким образом, имеем

$$p(\mathbf{v}\,|\,\mathbf{f}) = \sum_m p(\mathbf{v}, \mathbf{I}^{(m)}, \mathbf{f}) P(\mathbf{I}^{(m)}) = \sum_m \left[\frac{1}{(2\pi\sigma^2)^N} \exp\left(-\frac{\left\| \mathbf{v} - \mathbf{A}^{(m)}\mathbf{f} \right\|^2}{2\sigma^2} \right) \right] P(\mathbf{I}^{(m)}). \qquad (11.5.7)$$

Затем, оценка $\mathbf{f}$, которая максимизирует $p(\mathbf{v}\,|\,\mathbf{f})$ определяется решением уравнения

$$\frac{\partial p(\mathbf{v}\,|\,\mathbf{f})}{\partial \mathbf{f}} = \sum_m P(\mathbf{I}^{(m)}) (\mathbf{A}^{(m)\mathrm{T}} \mathbf{A}^{(m)} \mathbf{f} - \mathbf{A}^{(m)\mathrm{T}} \mathbf{v}) \exp\left(-\frac{\left\| \mathbf{v} - \mathbf{A}^{(m)}\mathbf{f} \right\|^2}{2\sigma^2} \right) = 0. \qquad (11.5.8)$$

Следовательно, оценку $\mathbf{f}$ можно выразить так:

$$\mathbf{f} = \left[\sum_m P(\mathbf{I}^{(m)}) \mathbf{A}^{(m)\mathrm{T}} \mathbf{A}^{(m)} g(\mathbf{v}, \mathbf{A}^{(m)}, \mathbf{f}) \right]^{-1} \cdot \sum_m P(\mathbf{I}^{(m)}) g(\mathbf{v}, \mathbf{A}^{(m)}, \mathbf{f}) \mathbf{A}^{(m)\mathrm{T}} \mathbf{v}, \qquad (11.5.9)$$

где функция $g(\mathbf{v}, \mathbf{A}^{(m)}, \mathbf{f})$ определяется так

$$g(\mathbf{v}, \mathbf{A}^{(m)}, \mathbf{f}) = \exp\left(-\frac{\left\| \mathbf{v} - \mathbf{A}^{(m)}\mathbf{f} \right\|^2}{2\sigma^2} \right). \qquad (11.5.10)$$

Результирующее решение для оптимального $\mathbf{f}$ обозначим $\mathbf{f}_{\text{МП}}$.

Уравнение (11.5.9) является нелинейным уравнением для оценки импульсной характеристики канала при заданном векторе принимаемого сигнал $\mathbf{v}$. В общем, трудно получить оптимальное решение непосредственным решением (11.5.9). С другой стороны относительно легко разработать численный метод для рекуррентного решения $\mathbf{f}_{\text{МП}}$. В частности, можем написать

$$\mathbf{f}^{(k+1)} = \left[\sum_m P(\mathbf{I}^{(m)}) \mathbf{A}^{(m)\mathrm{T}} \mathbf{A}^{(m)} g(\mathbf{v}, \mathbf{A}^{(m)}, \mathbf{f}^{(k)}) \right]^{-1} \cdot \sum_m P(\mathbf{I}^{(m)}) g(\mathbf{v}, \mathbf{A}^{(m)}, \mathbf{f}^{(k)}) \mathbf{A}^{(m)\mathrm{T}} \mathbf{v}. \qquad (11.5.11)$$

Когда $\mathbf{f}_{\text{МП}}$ получено из решения (11.5.9) или (11.5.11), мы можем просто использовать эту оценку при минимизации метрики $DM(\mathbf{I}, \mathbf{f}_{\text{МП}})$, определенной (11.5.3), по всем возможным последователям данных. Поскольку $\mathbf{I}_{\text{МП}}$ – это последовательность, которая минимизирует $DM(\mathbf{I}, \mathbf{f}_{\text{МП}})$, она определяется из уравнения

$$\min_{\mathbf{I}} DM(\mathbf{I}, \mathbf{f}_{\text{МП}}) = \min_{\mathbf{I}} \left\| \mathbf{v} - \mathbf{A}\mathbf{f}_{\text{МП}} \right\|^2. \qquad (11.5.12)$$

Мы знаем, что алгоритм Витерби является эффективным алгоритмом для выполнения минимизации $DM(\mathbf{I}, \mathbf{f}_{\text{МП}})$ по $\mathbf{I}$.

Обсуждаемый алгоритм имеет два главных недостатка. Первый – это то, что рекуррентная обработка (11.5.11) для нахождения $\mathbf{f}_{\text{МП}}$ в вычислительном отношении сложна. Второй и, вероятно, более важный, – оценка $\mathbf{f}_{\text{МП}}$ не так хороша по сравнению с максимально правдоподобной оценкой $\mathbf{f}_{\text{МП}}(\mathbf{I})$, которая получается, когда последовательность $\mathbf{I}$ известна. Следовательно, качество слепого эквалайзера (с алгоритмами Витерби), основанного на оценке $\mathbf{f}_{\text{МП}}$ хуже, чем того, который основан на $\mathbf{f}_{\text{МП}}(\mathbf{I})$. Ниже мы рассмотрим совместные оцениватели канала и данных.

Совместная оценка канала и данных. Здесь мы рассмотрим совместную оптимизацию показателя качества $DM(\mathbf{I}, \mathbf{f})$, определяемого (11.5.3). Поскольку элементы вектора импульсной характеристики канала $\mathbf{f}$ непрерывные, а элементы вектора данных $\mathbf{I}$ дискретные, возможный подход сводится к определению максимально правдоподобной оценки $\mathbf{f}$ для каждой возможной последовательности и затем к выбору последовательности данных, которая минимизирует $DM(\mathbf{I}, \mathbf{f})$ для каждой

соответствующей оценки канала. Итак оценка канала, соответствующая m-ой последовательности данных $\mathbf{I}^{(m)}$, равна

$$\mathbf{f}_{\text{МП}}(\mathbf{I}^{(m)}) = \left(\mathbf{A}^{(m)\text{T}}\mathbf{A}^{(m)}\right)^{-1}\mathbf{A}^{(m)\text{T}}\mathbf{v}. \qquad (11.5.13)$$

Для m-й последовательности данных метрика $DM(\mathbf{I},\mathbf{f})$ равна

$$DM(\mathbf{I}^{(m)},\mathbf{f}_{\text{МП}}(\mathbf{I}^{(m)})) = \left\|\mathbf{v} - \mathbf{A}^{(m)}\mathbf{f}_{\text{МП}}(\mathbf{I}^{(m)})\right\|^2. \qquad (11.5.14)$$

Затем из ансамбля M^N возможных последовательностей мы выберем последовательность данных, которая минимизирует функцию цены в (11.5.14), то есть мы определяем

$$\min_{\mathbf{I}^{(m)}} DM(\mathbf{I}^{(m)},\mathbf{f}_{\text{МП}}(\mathbf{I}^{(m)})). \qquad (11.5.15)$$

Подход, описанный выше, является исчерпывающим исследовательским вычислительным методом с вычислительной сложностью, которая растет экспоненциально с длиной блока данных. Мы можем выбрать $N = L$ и тогда мы будем иметь одну оценку канала для каждой из M^L выживших последовательностей. С этого момента можно продолжить поиск, сохраняя отдельную оценку канала для каждого выжившего пути при осуществлении поиска по алгоритму Витерби по решетке.

Подобный подход был предложен Сешадри (1991). В сущности, алгоритм Сешадри – это разновидность обобщённого алгоритма Витерби (ОАВ), который сохраняет $K \geq 1$ наилучших оценок переданной последовательности в каждом состоянии решётки и соответствующие оценки канала. В ОАВ Сешадри поиск идентичен обычному АВ, начиная с L-го шага по решётке, т.е. начиная с точки, когда обработана принятая последовательность $(\upsilon_1, \upsilon_2, ..., \upsilon_L)$. Так начиная с L-го шага формируется исчерпывающий поиск. С каждой последовательностью данных $\mathbf{I}^{(m)}$ связана соответствующая оценка канала $\mathbf{f}_{\text{МП}}(\mathbf{I}^{(m)})$. Начиная с этого шага, поиск модифицируется с тем, чтобы сохранить $K \geq 1$ выживших последовательностей и соответствующих оценок канала на состояние вместо только одной последовательности на состояние. Таким образом, ОАВ используется для обработки принимаемой сигнальной последовательности $\{\upsilon_n, n \geq L+1\}$. Оценки канала улучшаются рекуррентно на каждом шаге, используя алгоритм минимума СКО для дополнительного сокращения вычислительной сложности. Результаты моделирования, данные в статье Сешарди (1991), указывают на то, что этот ОАВ для реализации слепого выравнивания работает хорошо при умеренном отношении сигнал/шум с $K = 4$. Затем имеется умеренный рост вычислительной сложности ОАВ по сравнению с обычным АВ. Однако здесь имеется дополнительные вычисления, связанные с оцениванием и обновлением начальных оценок канала $\mathbf{f}(\mathbf{I}^{(m)})$, связанных с каждой из выживших оценок данных.

Альтернативный алгоритм совместного оценивания, который избегает вычисления наименьших квадратов при оценивании канала, был предложен Зервасом и др. (1991). В этом алгоритме порядок формирования совместной минимизации показателя качества $DM(\mathbf{I},\mathbf{f})$ обратный. Это значит, сначала выбирается импульсная характеристика канала, скажем $\mathbf{f} = \mathbf{f}^{(1)}$, а затем используется обычный АВ для нахождения оптимальной последовательности данных для этой импульсной характеристики канала. Затем мы можем модифицировать $\mathbf{f}^{(1)}$ до $\mathbf{f}^{(2)} = \mathbf{f}^{(1)} + \Delta\mathbf{f}^{(1)}$ и повторить оптимизацию по последовательностям данных $\{\mathbf{I}^{(m)}\}$.

Основываясь на этом общем подходе Зервас разработал новый МП алгоритм слепого выравнивания, который назван *алгоритмом с квантованием канала*. Алгоритм работает по

решетке пространства канала, причем он становится лучше и лучше при использовании МП правила для сохранения оцененного канала в окрестности действительного неизвестного канала. Этот алгоритм приводит к эффективной параллельной реализации, а его требования к памяти такие же, как в АВ.

11.5.2. Стохастический градиентный алгоритм

Другим классом алгоритмов слепого выравнивания являются схемы стохастически-градиентного итеративного выравнивания, которые содержат на выходе линейного КИХ-выравнивающего фильтра безынерционную нелинейность, для того чтобы генерировать «желательную характеристику» на каждой итерации.

Начнем с первоначального предположения, что коэффициенты оптимального эквалайзера равны $\{c_n\}$. Затем свёртку импульсной характеристики канала и импульсных откликов эквалайзера можно выразить так

$$\{c_n\}\otimes\{f_n\}=\{\delta_n\}+\{e_n\}, \tag{11.5.16}$$

где $\{\delta_n\}$ – единичная отсчётная последовательность, а $\{e_n\}$ означает последовательность ошибок, возникающая из нашего первоначального предположения коэффициентов эквалайзера.

Если мы возьмем свертку импульсного отклика эквалайзера и принимаемой последовательности $\{\upsilon_n\}$ мы получим

$$\begin{aligned}\{\hat{I}_N\}&=\{\upsilon_n\}\otimes\{c_n\}=\{I_n\}\otimes\{f_n\}\otimes\{c_n\}+\{\eta_n\}\otimes\{c_n\}=\{I_n\}\otimes(\{\delta_n\}+\{e_n\})+\{\eta_n\}\otimes\{c_n\}\\&=\{I_n\}+\{I_n\}\otimes\{e_n\}+\{\eta_n\}\otimes\{c_n\}.\end{aligned} \tag{11.5.17}$$

Слагаемое $\{I_n\}$ в (11.5.17) представляет желательную последовательность данных, слагаемое $\{I_n\}\otimes\{e_n\}$ представляет остаточную МСИ, а слагаемое $\{\eta_n\}\otimes\{c_n\}$ представляет аддитивный шум. Наша задача сводится к использованию «развернутой» последовательности $\{\hat{I}_n\}$, чтобы найти наилучшую оценку «желательного» отклика, которую обозначим в общем $\{d_n\}$. В случае адаптивного выравнивания, использующего обучающую последовательность $\{d_n\}=\{I_n\}$. При варианте слепого выравнивании мы хотим генерировать «желательный» отклик из $\{\hat{I}_n\}$.

Для определения наилучшей оценки $\{I_n\}$ по наблюдаемой на выходе эквалайзера последовательности $\{\hat{I}_n\}$ можно использовать критерий минимума среднего квадрата ошибки (СКО). Поскольку передаваемые последовательности $\{I_n\}$ имеет негауссовскую ФПВ, оценка по минимуму СКО определяется нелинейным преобразованием $\{\hat{I}_n\}$. В общем «наилучшая» оценка $\{d_n\}$ определяется так:

$$\begin{aligned}d_n&=g(\hat{I}_n) \qquad\qquad (\text{без памяти})\\ d_n&=g(\hat{I}_n,\hat{I}_{n-1},\ldots,\hat{I}_{n-m}) \quad (\text{память } m-\text{го порядка}),\end{aligned} \tag{11.5.18}$$

где $g(\cdot)$ – нелинейная функция. Последовательность $\{d_n\}$ затем используется для генерирования сигнала ошибки, который подается обратно на фильтр адаптивного выравнивания, как показано на рис.11.5.1.

Как хорошо известно, классическая задача оценивания формулируется так. Если выход эквалайзера $\hat{I}_n$ выразить так

$$\hat{I}_n=I_n+\tilde{\eta}_n, \tag{11.5.19}$$

где $\tilde{\eta}_n$ предполагается гауссовским с нулевым средним (здесь использована центральная предельная теорема теории вероятностей для остаточной МСИ и аддитивного шума), $\{I_n\}$ и $\{\tilde{\eta}_n\}$ статистически независимы, а $\{I_n\}$– статистически независимые и одинаково распределенные случайные величины, тогда оценка по минимуму СКО равна

$$d_n = E(I_n | \tilde{I}_n), \qquad (11.5.20)$$

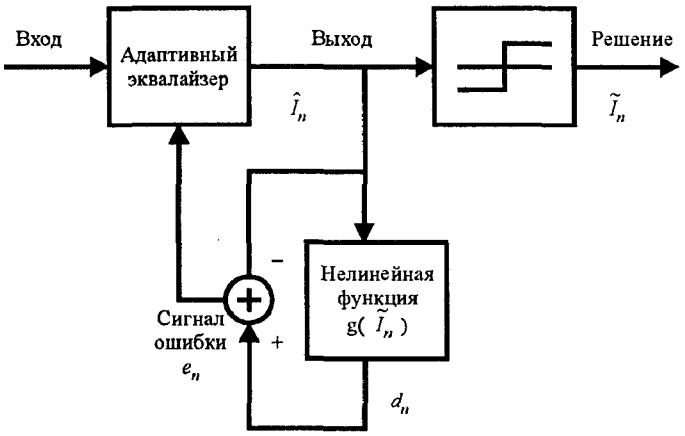


Рис.11.5.1. Адаптивное слепое выравнивание со стохастическим градиентным алгоритмом

которая является нелинейной функцией выхода эквалайзера, если $\{I_n\}$ негауссовские случайные величины.

Таблица 11.5.1. Стохастические градиентные алгоритмы для слепого выравнивания

Коэффициенты ячеек эквалайзера	$\{c_n, 0 \le n \le N-1\}$
Последовательность принятого сигнала	$\{\upsilon_n\}$
Последовательность на выходе эквалайзера	$\{\hat{I}_n\} = \{\upsilon_n\} \otimes \{c_n\}$
Последовательность ошибок эквалайзера	$\{e_n\} = g(\hat{I}_n) - \hat{I}_n$
Уравнение обновления оценок	$\mathbf{c}_{n+1} = \mathbf{c}_n + \Delta \mathbf{v}_n^* e_n$

Алгоритм	**Нелинейность:** $g(\tilde{I}_n)$
Годарда	$\frac{\hat{I}_n}{\lvert\tilde{I}_n\rvert}\left(\lvert\tilde{I}_n\rvert + R_2\lvert\hat{I}_n\rvert - \lvert\hat{I}_n\rvert^3\right), R_2 = \frac{E\{\lvert I_n\rvert^4\}}{E\{\lvert I_n\rvert^2\}}$
Сато	$\zeta\,\mathrm{csgn}(\hat{I}_n), \zeta = \frac{E\{[\mathrm{Re}(I_n)]^2\}}{E\{\lvert\mathrm{Re}(I_n)\rvert\}}$
Бенвенисто–Горсата	$\hat{I}_n + k_1(\hat{I}_n - I_n) + k_2\lvert\hat{I}_n - \tilde{I}_n\rvert[\zeta\,\mathrm{csgn}(\hat{I}_n) - \tilde{I}_n], k_1$ и k_2
Старт–стоп	$\hat{I}_n + \frac{1}{2}A(\hat{I}_n - \tilde{I}_n) + \frac{1}{2}B(\hat{I}_n - \tilde{I}_n)^*$ $(A,B) = (2,0)$, $(1,1)$, $(1,-1)$ или $(0,0)$, в зависимости от знака ошибки $\hat{I}_n - \tilde{I}_n$ и ошибки $\zeta\,\mathrm{csgn}(\hat{I}_n) - \hat{I}_n$

Табл.11.5.1 иллюстрирует общую форму существующих алгоритмов слепого выравнивания, базирующиеся на НК адаптации. Мы видим, что базовое отличие этих алгоритмов заключается в выборе инерционной нелинейности. Наиболее широко используемым на практике алгоритмом является *алгоритм Годарда*, иногда называемый *алгоритмом с постоянным модулем* (АПМ).

Из табл.11.5.1 очевидно, что выходная последовательность $\{d_n\}$, получаемая при использовании нелинейной функции от выхода эквалайзера, играет роль желательного отклика или обучающей последовательности. Также очевидно, что рассматриваемые алгоритмы просты для реализации, поскольку они являются базовыми алгоритмами типа НК. Раз так, мы ожидаем, что характеристики сходимости этих алгоритмов будут зависеть от матрицы автокорреляции принимаемых данных $\{\upsilon_n\}$.

С учетом сходимости адаптивные алгоритмы вида НК сходятся в среднем, когда

$$E[\upsilon_n g^*(\hat{I}_n)] = E[\upsilon_n \hat{I}_n^*], \tag{11.5.21}$$

и в средне квадратичном смысле, когда (верхний индекс H означает сопряжённое транспонирование)

$$\begin{aligned} E[\mathbf{c}_n^H \upsilon_n g^*(\hat{I}_n)] &= E[c_n^H \upsilon_n \hat{I}_n^*], \\ E[\hat{I}_n g^*(\hat{I}_n)] &= E[|\hat{I}_n|^2] \end{aligned} \tag{11.5.22}$$

Следовательно, требуется, чтобы выход эквалайзера $\{\hat{I}_n\}$ удовлетворял условию (11.5.22). Заметим, что (11.5.22) устанавливает, что автокорреляция $\{\hat{I}_n\}$ (правая часть) равна взаимной корреляции между $\hat{I}_n$ и нелинейного преобразования $\hat{I}_n$ (левая часть). Процесс, удовлетворяющий этому свойству, назван Беллини (1986) процессом Базганга (1952). В целом алгоритмы, данные в табл.11.5.1, сходятся, когда выходная последовательность эквалайзера $\hat{I}_n$ удовлетворяет свойству Базганга.

Базовое ограничение стохастических градиентных алгоритмов относительно медленная сходимость. Некоторые улучшения в скоростях сходимости можно достичь модификацией адаптивных алгоритмов типа НК в рекуррентный тип наименьших квадратов (РНК).

Алгоритм Годарда. Как указанно выше, алгоритм слепого выравнивания Годарда является алгоритмом скорейшего спуска, который широко используется на практике, когда обучающая последовательность нежелательна. Опишем этот алгоритм более подробно.

Годард рассматривает задачу об объединении выравнивания и восстановления фазы несущей и её отслеживания. Отслеживание фазы несущей выполняется на базовом сигнале после эквалайзера, как показано на рис.11.5.2.

Основываясь на этой структуре, мы можем выразить выход эквалайзера так

$$\hat{I}_k = \sum_{n=-K}^{K} c_n \upsilon_{k-n}, \tag{11.5.23}$$

а вход на устройство решения так $\hat{I}_k \exp(-j\hat{\phi}_k)$, где $\hat{\phi}_k$ – оценка фазы несущей на k-м символьном интервале.

Если желательный символ известен мы можем формировать сигнал ошибки

$$\varepsilon_k = I_k - \hat{I}_k e^{-j\hat{\phi}_k} \tag{11.5.24}$$

и минимизировать СКО по $\hat{\phi}_k$ и $\{c_n\}$

$$\min_{\phi_k, \mathbf{c}} E\left(|I_k - \hat{I}_k e^{-j\hat{\phi}_k}|^2\right). \tag{11.5.25}$$

Этот критерий ведет нас к использованию алгоритма НК для рекуррентного оценивания $\mathbf{C}$ и ϕ_k. Алгоритм НК, базирующийся на знании переданной последовательности можно записать

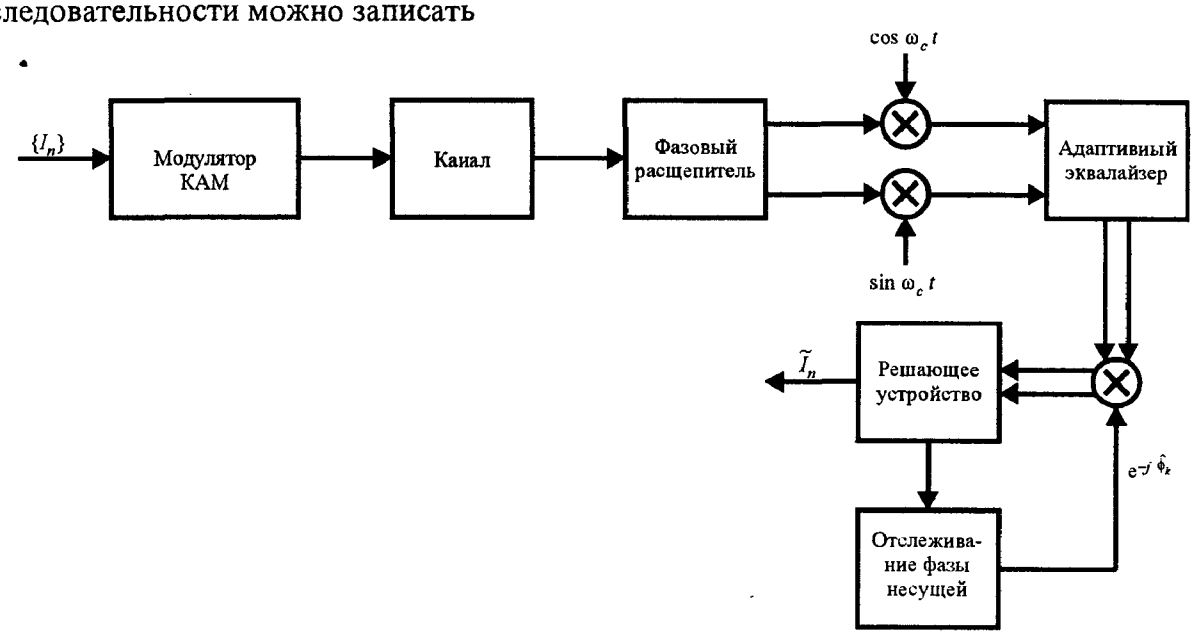


Рис. 11.5.2. Схема Годарда для объединения адаптивного (слепого) выравнивания и отслеживания фазы несущей

$$\hat{\mathbf{C}}_{k+1} = \hat{\mathbf{C}}_k + \Delta_c (I_k - \hat{I}_k e^{-j\hat{\phi}_k}) \mathbf{V}_k^* e^{j\hat{\phi}_k}, \tag{11.5.26}$$

$$\hat{\phi}_{k+1} = \hat{\phi}_k + \Delta_\phi \operatorname{Im}(I_k \hat{I}_k^* e^{j\hat{\phi}_k}), \tag{11.5.27}$$

где Δ_c и Δ_ϕ – параметры размера шага для двух рекуррентных уравнений. Заметим, что эти рекуррентные уравнения объединены вместе. К сожалению, эти уравнения, в общем не сходятся, когда желательная последовательность символов $\{I_k\}$ неизвестна.

Подход, предложенный Годардом, сводится к использованию критерия, который зависит от уровня МСИ на выходе эквалайзера, но который независим от сигнального созвездия КАМ и фазы несущей. Для примера, функция стоимости, которая не зависит от фазы несущей и имеет свойство, что ее минимум ведет к малой величине СКО, равна

$$G^{(p)} = E(|\hat{I}_k|^p - |I_k|^p)^2, \tag{11.5.28}$$

где p – положительное, вещественное и целое число. Минимизация $G^{(p)}$ относительно коэффициентов эквалайзера ведет к выравниванию только сигнальных амплитуд. Основываясь на этих наблюдениях, Годард выбрал более общую функцию сходимости, названной дисперсией порядка p, и определяемой так:

$$D^{(p)} = E(|\hat{I}_k|^p - R_p)^2, \tag{11.5.29}$$

где R_p – положительная вещественная константа. Как и в случае $G^{(p)}$, мы видим, что $D^{(p)}$ не зависит от фазы несущей. Минимизация $D^{(p)}$ по коэффициентам эквалайзера можно выполнить рекуррентно согласно алгоритму скорейшего спуска

$$C_{k+1} = C_k - \Delta_p \frac{dD^{(p)}}{dC_k}, \tag{11.5.30}$$

где Δ_p – параметр размера шага. Дифференцируя $D^{(p)}$ и применяя операцию усреднения, мы получаем следующий алгоритм типа НК для настройки коэффициентов эквалайзера:

$$\hat{\mathbf{C}}_{k+1} = \hat{\mathbf{C}}_k + \Delta_p \mathbf{V}_k^* \hat{I}_k \, |\hat{I}_k|^{p-2} (R_p - |\hat{I}_k|^p), \qquad (11.5.31)$$

а оптимальный выбор R_p дает

$$R_p = \frac{E(|I_k|^{2p})}{E(|I_k|^p)}. \qquad (11.5.32)$$

Как ожидалось, рекуррентное соотношение (11.5.31) для $\hat{\mathbf{C}}_k$ не требует знания фазы несущей. Отслеживание фазы несущей можно выполнить по варианту управления решениями согласно (11.5.27).

Особенно важен случай, когда p=2, который ведет к относительно простому алгоритму:

$$\begin{aligned} \hat{\mathbf{C}}_{k+1} &= \hat{\mathbf{C}}_k + \Delta_p \mathbf{V}_k^* \hat{I}_k (R_p - |\hat{I}_k|^2) \\ \hat{\phi}_{k+1} &= \hat{\phi}_k + \Delta_\phi \operatorname{Im}(\tilde{I}_k \hat{I}_k^* e^{j\hat{\phi}_k}), \end{aligned} \qquad (11.5.33)$$

где $\tilde{I}_k$ – выход решающего устройства, использующего $\hat{I}_k$, и

$$R_2 = \frac{E(|I_k|^4)}{E(|I_k|^2)}. \qquad (11.5.34)$$

Сходимость алгоритма (11.5.33) была показана в статье Годарда (1980). Первоначально коэффициенты эквалайзера выбираются нулевыми, исключая коэффициент центральной (опорной) ячейки, который выбирается согласно условию

$$|c_0|^2 > \frac{E|I_k|^4}{2|x_0|^2 [E(|I_k|^2)]^2}, \qquad (11.5.35)$$

которое является достаточным, но не необходимым для сходимости алгоритма. Результаты моделирования, выполненные Годардом на моделях телефонных каналов с типичными частотными характеристиками при скоростях передачи 7200...12000 бит/с, указывает на то, что алгоритм (11.5.33) хорошо реализуется и ведёт к сходимости на интервале 5000...20 000 итераций, в зависимости от сигнального созвездия. Первоначально, до выравнивания, глазковая диаграмма была закрыта. Число итераций, требуемых для сходимости, примерно на порядок величины больше, чем число, требуемое для выравнивания канала с известной обучающей последовательностью. Никаких существенных трудностей не возникает при использовании алгоритма с управлением решениями для оценивания фазы в (11.5.33) для начала процесса настройки эквалайзера.

11.5.3. Алгоритмы слепого выравнивания, основанные на статистике сигнала второго и более высокого порядка

Хорошо известно, что статистики второго порядка (автокорреляция) принимаемой сигнальной последовательности дает информацию об амплитудно-частотной характеристике, но не о фазочастотной характеристике. Однако это утверждение неправильно, если автокорреляционная функция принимаемого сигнала периодическая, что является случаем цифрового модулированного сигнала. В таком случае возможно получить измерение АЧХ и ФЧХ канала по принимаемому сигналу. Это свойство циклостационарности принимаемого сигнала образует базу для алгоритмов оценки канала, предложенных Тонгом и др. (1993).

Также возможно оценить характеристики канала по принимаемому сигналу, используя методы статистики более высокого порядка. В частности, импульсную характеристику линейной не меняющейся во времени системы с дискретным временем можно получить полностью по кумулянтам принимаемого сигнала, при условии, что вход канала не

гауссовский. Мы опишем простой метод оценки импульсной характеристики канала (отклик канала импульс?) при помощи кумулянтов четвёртого порядка принимаемой сигнальной последовательности. Кумулянт четвёртого порядка определяется так:

$$c(\upsilon_k, \upsilon_{k+m}, \upsilon_{k+n}, \upsilon_{k+l}) \equiv c_\upsilon(m,n,l) = E(\upsilon_k \upsilon_{k+m} \upsilon_{k+n} \upsilon_{k+l}) -$$
$$- E(\upsilon_k \upsilon_{k+m}) E(\upsilon_{k+n} \upsilon_{k+l}) - E(\upsilon_k \upsilon_{k+n}) E(\upsilon_{k+m} \upsilon_{k+l}) - E(\upsilon_k \upsilon_{k+l}) E(\upsilon_{k+m} \upsilon_{k+n}) \qquad (11.5.36)$$

(кумулянт четвертого порядка для гауссовского сигнала равен нулю). Как следствие, получаем

$$c_r(m,n,l) = c(I_k, I_{k+m}, I_{k+n}, I_{k+l}) \sum_{k=0}^{\infty} f_k f_{k+m} f_{k+n} f_{k+l} . \qquad (11.5.37)$$

Для статистически независимой и одинаково распределенной входной последовательности $\{I_k\}$ канала $c(I_k, I_{k+m}, I_{k+n}, I_{k+l}) = k$ – константа, называемая *куртосисом*. Затем, если память канала равна $L+1$, мы можем положить $m = n = l = -L$, так что

$$c_r(-L,-L,-L) = k f_L f_0^3 . \qquad (11.5.38)$$

Аналогично, если положим $m = 0$, $n = L$, и $l = p$, получим

$$c_r(0, L, p) = k f_l f_0^2 f_p . \qquad (11.5.39)$$

Если скомбинируем (11.5.38) и (11.5.39), мы получим отсчёты импульсной характеристики канала с точностью до скалярного множителя

$$f_p = f_0 \frac{c_r(0, L, p)}{c_r(-L,-L,-L)}, \qquad p = 1, 2, \ldots, L . \qquad (11.5.40)$$

Кумулянты $c_r(m,n,l)$ оцениваются усреднением по отсчетам принимаемой сигнальной последовательности $\{\upsilon_n\}$.

Другой подход, основанный на статистике высших порядков, использовали Хатзинакос и Никиас (1991). Они первыми ввели полиспектральный метод адаптивного слепого выравнивания, называемый трикепстр-алгоритмом выравнивания (ТАВ).

Этот метод оценивает характеристики канального отклика путем использования сложного кепстрального представления кумулянта четвертого порядка (трикепстр) принимаемой сигнальной последовательности $\{\upsilon_n\}$. ТАВ зависит только от кумулянтов четвёртого порядка от $\{\upsilon_n\}$ и способен отдельно реконструировать минимально-фазовые и максимально-фазовые характеристики канала. Затем вычисляются коэффициенты канального эквалайзера по измеренным канальным характеристикам. Базовый подход, использующий ТАВ, сводится к вычислению трикепстра принимаемый сигнальной последовательности $\{\upsilon_n\}$, который является обратным (трёхмерным) преобразованием Фурье логарифма трикепстра от $\{\upsilon_n\}$. (*Трикепстр* является трёхмерным дискретным преобразованием Фурье кумулянтной последовательности четвёртого порядка $c_r(m,n,l)$). Коэффициенты эквалайзера вычисляются по кепстр-коэффициентам.

Путём отделения оценивания канала от канального выравнивания, возможно использовать любой тип эквалайзера для МСИ, то есть или линейной, или с обратной связью по решению, или максимально правдоподобное последовательное детектирование. Главный недостаток этого класса алгоритмов – большое количество данных и свойственная им вычислительная сложность, включая оценивание моментов высоких порядков (кумулянтов) принимаемого сигнала.

В заключении мы дали обзор трёх классов алгоритмов слепого выравнивания, которые нашли применение в цифровой связи. Из трёх семейств описанных алгоритмов те, которые

базируются на правиле максимального правдоподобия для совместного оценивания импульсной характеристики канала и последовательности данных, являются оптимальными и требуют относительно немного принимаемых отсчётов сигнала для выполнения оценивания канала. Однако вычислительная сложность этих алгоритмов велика, когда МСИ простирается на много символов. В некоторых каналах, таких как каналы мобильной радиосвязи, когда протяжённость МСИ относительно короткая, эти алгоритмы легко выполнять, однако в телефонных каналах, где МСИ простирается на многие символы, но она не является очень тяжёлой, обычно используются алгоритмы типа НК (стохастического градиента).

11.6. БИБЛИОГРАФИЧЕСКИЕ ЗАМЕЧАНИЯ И ССЫЛКИ

Адаптивное выравнивание для цифровой связи было разработано Лакки (1965, 1966). Его алгоритм, базирующийся на критерии минимума пикового искажения, вёл к АСН. Работа Лакки была главным прорывом, который привёл к быстрому развитию высокоскоростных модемов в пределах пяти лет после публикации его работ. Конкурентный алгоритм НК был разработан Уидроу (1966) и его использование для адаптивного выравнивания комплексных сигналов (синфазных и квадратурных компонент) было описано и проанализировано в учебных статьях Прокиса и Миллера (1969).

Учебная трактовка алгоритмов адаптивного выравнивания, разработанных за период 1965–1975 дана Прокисом (1975). Более новая учебная трактовка адаптивного выравнивания дана в статье Куреши (1985) Важный прорыв в технике адаптивного выравнивания, начатый работой Лакки в 1965, связан с разработкой решетчато-кодовой модуляции, которая была предложена Унгербоеком и Чайка (1976), что привело к разработке коммерчески приемлемых высокоскоростных модемов со скоростями передачи 9600–28800 бит/с по телефонным каналам.

Использование алгоритмов более быстрой сходимости для адаптивного выравнивания было предложены Годардом (1974). Наше изложение НК (Калмана) алгоритма, описанного в разделе 11.4.1, следует подходу Пикинбоно (1978). РНК-лестничные алгоритмы для общего оценивания сигналов были разработаны Морфом и др. (1977, 1979). Приложения этих алгоритмов были переработаны несколькими исследователями, включая Макхоула (1978), Саториуса и Пака (1981), Саториуса и Александера (1979) и Лингом и Прокисом (1982, 1984 a-e, 1985). Быстрый НК алгоритм Калмана для адаптивного выравнивания был впервые описан Фальконером и Льюнгом (1978). Выше сделанные ссылки являются как раз теми немногими важными статьями, опубликованными по НК алгоритмам для адаптивного выравнивания и других применений.

Оригинальная работа Сато (1975) по слепому выравниванию была сконцентрированы на АМ (одномерных) сигнальных созвездий. Впоследствии она была обобщенна на двух мерные и многомерные сигнальные созвездия в алгоритмах, открытых (?) Годардом (1980), Бенвинистом и Гоурсатом (1984), Сато, (1986) Фошини (1985), Пичи и Прати (1987) и Шалви и Вайнштейном (1990). Методы слепого выравнивания, основанные на использовании моментов второго и более высоких порядков принимаемого сигнала были предложены Хатзинакосом и Никиасом (1991) и Тонгом и другими (1994). Использование правила максимального правдоподобия для совместного оценивания канала и детектирование данных были исследованы и обсуждены в статьях Сешадри (1991), Гоша и Вебера (1991), Зерваса и других (1991) и Рахели и других (1995). Наконец, характеристики сходимости алгоритмов стохастического градиентного слепого выравнивания были исследованы Дингом (1990), Дингом и другими (1989) и Джонсоном (1991).

ЗАДАЧИ

11.1. Эквивалентный канал с дискретным временем с белым гаусовским шумом показан на рис. Р11.1.

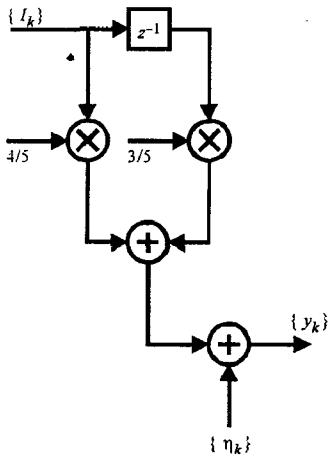


Рис. Р.11.1

a) Предположим, что мы используем линейный эквалайзер для выравнивания канала. Определите коэффициенты ячеек c_{-1}, c_0, c_1 для трехячеечного эквалайзера. Для упрощения расчетов считается, что АБГШ имеет нулевое среднее.

b) Коэффициенты ячеек линейного эквалайзера в (a) определяются рекуррентно по алгоритму

$$\mathbf{C}_{k+1} = \mathbf{C}_k + \Delta g_k\,,\ \mathbf{C}_k = [c_{-1k}\ c_{0k}\ c_{1k}]^{\mathrm{T}},$$

где $\mathbf{g}_k = \mathbf{\Gamma C}_k - \mathbf{b}$ – вектор градиента, а Δ – размер шага ячейки. Определить диапазон возможных величин Δ, чтобы гарантировать сходимость рекуррентного алгоритма. Для упрощения вычислений считайте, что АБТШ имеет нулевое среднее.

c) Определите веса ячеек ЭОСР с двумя ячейками в цепи прямой связи и одной в цепи обратной связи. Для упрощения вычислений считайте, что АБТШ имеет нулевое среднее.

11.2. Ссылаясь на задачу 10.18, ответьте на следующие вопросы.

a) Определите максимальное значение Δ, которую можно использовать, чтобы обеспечить сходимость коэффициентов эквалайзера в процессе работы в адаптивном режиме.

b) Какова зависимость от Δ дисперсии собственного шума, создаваемого трех ячеечным эквалайзером, когда он работает в адаптивном режиме. Предположим, что желательно ограничить дисперсию собственного шума до 10 % от минимального СКО в трех ячеечном эквалайзере, когда $N_0 = 0{,}1$. Какую величину Δ вы выберете?

c) Если оптимальные коэффициенты эквалайзера вычисляются рекуррентно методом крутого спуска, рекуррентное уравнение можно записать в виде

$$\mathbf{C}_{(n+1)} = (\mathbf{I} - \Delta\mathbf{\Gamma})\mathbf{C}_{(n)} + \Delta\xi,$$

где $\mathbf{I}$ – единичная матрица. Уравнение определяет систему из трёх связанных разностных уравнений первого порядка. Они могут быть развязаны линейным преобразованием, которое диагонализирует матрицу $\mathbf{\Gamma}$. Это значит $\mathbf{\Gamma} = \mathbf{U}\Lambda\mathbf{U}^{\mathrm{T}}$, где Λ – диагональная матрица, имеющая собственные значения матрицы $\mathbf{\Gamma}$, как свои диагональные элементы, а $\mathbf{U}$ является унитарной матрицей, которую можно получить из вашего ответа на задачу 10.18(b). Пусть $\mathbf{C}^{\mathrm{T}} = \mathbf{U}^{\mathrm{T}}\mathbf{C}$ определяет устойчивое решение для $\mathbf{C}^{\mathrm{T}}$. Через нее оцените $(?)\,\mathbf{C} = (\mathbf{U}^{\mathrm{T}})^{-1}\mathbf{C}^{\mathrm{T}} = \mathbf{U}\mathbf{C}^{\mathrm{T}}$ и, таким образом покажите, что ваш ответ соответствует результату, полученному 10.18(a).

11.3. Если используется периодическая псевдослучайная последовательность длины N для настройки коэффициентов N ячеечного линейного эквалайзера, вычисления можно выполнять эффективно в частотной области путем использования дискретного преобразования Фурье (ДПФ). Предположим, что $\{y_n\}$ это последовательность из N принятых отсчётов (взятых со скоростью передачи символов) на входе эквалайзера. Тогда вычисление коэффициентов эквалайзера выполните так

a) Вычислите ДПФ для одного периода входной последовательности $\{y_n\}$ эквалайзера

$$Y_k = \sum_{n=0}^{N-1} y_n e^{-j2\pi nk/N}$$

b) Вычислите желательный спектр эквалайзера

$$C_k = \frac{X_k Y_k^*}{|Y_k|^2}, \qquad k = 0, 1, \ldots, N-1,$$

где $\{X_k\}$ – предварительно вычисленное ДПФ обучающейся последовательности.

c) Вычислите обратные ДПФ от $\{C_k\}$ для получения коэффициентов эквалайзера $\{c_n\}$. Покажите, что эта процедура в отсутствие шума дает эквалайзер, чья частотная характеристика в отсчётных точках $f_k = k/NT$, $k = 0, 1, \ldots, N-1$ обратна отсчётам сложенной частотной характеристики канала

11.4. Покажите, что вектор градиентов при минимизации СКО можно выразить так

$$\mathbf{G}_k = -E(\varepsilon_k \mathbf{V}_k^*),$$

где ошибка $\varepsilon_k = I_k - \hat{I}_k$, а оценка $\mathbf{G}_k$, то есть $\hat{\mathbf{G}}_k = -\varepsilon_k \mathbf{V}_k^*$ удовлетворяет условию

$$E(\hat{\mathbf{G}}_k) = \mathbf{G}_k$$

11.5. Алгоритм НК с *пошаговой утечкой*, предложенный в статье Гитлина и другнх (1982), можно выразить так

$$\mathbf{C}_N(n+1) = w\mathbf{C}_N(n) + \Delta\varepsilon(n)\mathbf{V}_n^*(n),$$

где 0 <w<1, Δ– размер шага, а $\mathbf{V}_N(n)$ – вектор данных в момент n. Определите условие для сходимости среднего значения $C_N(n)$

11.6. Рассмотрите случайный процесс

$$x(n) = g\upsilon(n) + w(n), \qquad n = 0, 1, \ldots, M-1,$$

где v_n– известная последовательность, g– случайная величина с $E(g)$=0 и $E(g^2) = G$. Процесс w(n) – последовательность белого шума с функцией корреляции

$$\gamma_{ww}(m) = \sigma_w^2 \delta_m.$$

Определите коэффициенты линейного оценивателя для g:

$$\hat{g} = \sum_{n=0}^{M-1} h(n)x(n),$$

которые минимизируют СКО

11.7. Цифровой трансверсальный фильтр можно реализовать через частотные отсчёты посредством системной функции (см. задачу 10.25)

$$H(z) = \frac{1-z^{-M}}{M} \sum_{k=0}^{M-1} \frac{H_k}{1-e^{j2\pi k+M} z^{-1}} = H_1(z)H_2(z),$$

где $H_1(z)$ – гребенчатый фильтр, $H_2(z)$ – параллельный блок резонаторов, а $\{H_k\}$ – значение дискретного преобразования Фурье (ДПФ).

a) Предположим, что эта структура реализована как адаптивный фильтр использующий алгоритм НК для настройки параметров фильтров (ДПФ) $\{H_k\}$. Дайте обновляющее уравнение для этих параметров. Нарисуйте структуру адаптивного фильтра.

b) Предположим, что эта структура использована как адаптивный канальный эквалайзер, в котором желательный сигнал

$$d(n) = \sum_{k=0}^{M-1} A_k \cos \omega_k n, \qquad \omega_k = \frac{2\pi k}{M}.$$

При такой форме желательного сигнала, какое преимущество имеет адаптивный алгоритм НК для коэффициентов ДПФ $\{H_k\}$ относительно прямой структурной формы с коэффициентами $\{h(n)\}$ (см. Прокис, 1970).

11.8. Рассмотрите показатель качества $J = h^2 + 40h + 28$.

Предположим, что мы ищем минимум J используя алгоритм крутого спуска

$$h(n+1) = h(n) - \tfrac{1}{2}\Delta g(n),$$

где $g(n)$ – градиент.

a) Определите диапазон значений Δ, которые обеспечивают сходимость системы в процессе настройки.

b) Нарисуйте J, как функцию n для значений Δ в этой области.

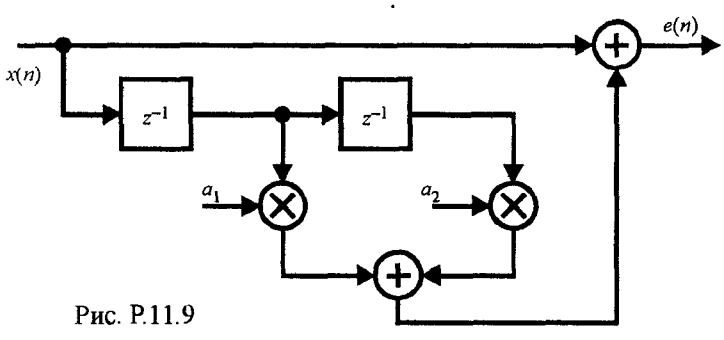


Рис. P.11.9

11.9. Определите коэффициенты a_1 и a_2 для линейного предсказателя, показанного на рис. P11.9, при заданной автокорреляционной функции $\gamma_{xx}(m)$ входного сигнала

$$\gamma_{xx}(m) = b^m, \qquad 0 < b < 1$$

11.10. Определите лестничный фильтр и его оптимальные коэффициенты, соответствующие линейному предсказателю задачи 11.9.

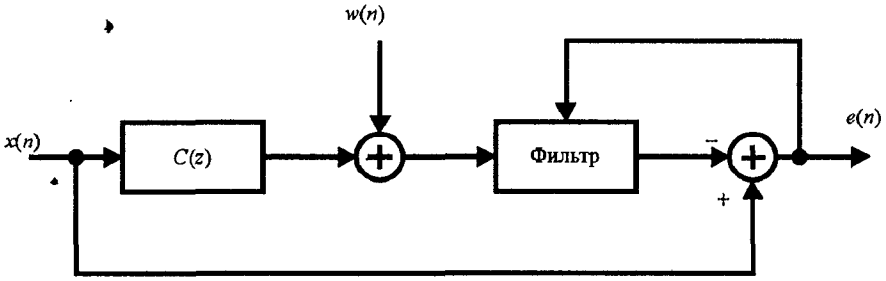


Рис.Р.11.11

11.11. Рассмотрите адаптивный КИХ фильтр, показанный на рис. Р11.11.

Система характеризуется системной функцией

$$C(z)=\frac{1}{1-0{,}9z^{-1}}$$

Определите оптимальные коэффициенты адаптивного трансверсального (КИХ) фильтра $B(z)=b_0+b_1z^{-1}$, которые минимизируют СКО. Аддитивный шум белый с дисперсией $\sigma_w^2=0.1$.

11.12. Матрица корреляций $\mathbf{\Gamma}$ размерности $N\times N$ имеет собственные числа $\lambda_1>\lambda_2>\ldots>\lambda_N>0$ и соответствующие собственные векторы $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N$. Такую матрицу можно представить так

$$\mathbf{\Gamma}=\sum_{t=1}^{N}\lambda_t\mathbf{v}_t\mathbf{v}_t^{*\mathrm{T}}.$$

a) Если $\mathbf{\Gamma}=\mathbf{\Gamma}^{1/2}\mathbf{\Gamma}^{1/2}$, где $\mathbf{\Gamma}^{1/2}$ – квадратный корень из $\mathbf{\Gamma}$, покажите, что $\mathbf{\Gamma}^{1/2}$ можно представить так

$$\mathbf{\Gamma}^{1/2}=\sum_{t=1}^{N}\lambda_t^{1/2}\mathbf{v}_t\mathbf{v}_t^{*\mathrm{T}}.$$

b) Используя это представление, определите процедуру для вычисления $\mathbf{\Gamma}^{1/2}$.

12

МНОГОКАНАЛЬНЫЕ СИСТЕМЫ И СИСТЕМЫ С МНОГИМИ НЕСУЩИМИ

В некоторых применениях желательно передать сигнал с одной и той же информацией по нескольким каналам. Этот вариант передачи используется прежде всего в ситуациях, когда имеется большая вероятность того, что один или более каналов будут ненадежными от времени ко времени. Для примера, радиоканалы, такие как с ионосферным и тропосферным рассеиванием, обусловленным многопутёвостью, вызывают замирания сигналов, что делает каналы ненадёжными на некоторых временных интервалах. Второй пример, когда многоканальная передача используется в военной связи как средство преодоления радиопротиводействия. Передавая одну и ту же информацию по многим каналам, мы осуществляем разнесение сигнала, которое приёмник может использовать для восстановления информации.

Другой вид многоканальной передачи – это передача на многих несущих, когда полоса частот канала разделяется на определенное число подканалов, и по каждому из подканалов передаётся различная информация. Целесообразный способ разделения полосы частот канала на определенное число узкополосных подканалов даётся ниже.

В этой главе мы рассмотрим как многоканальную передачу, так и передачу на многих несущих. Мы начинаем с трактовки многоканальной передачи.

12.1. МНОГОКАНАЛЬНАЯ ЦИФРОВАЯ СВЯЗЬ В КАНАЛАХ С АБГШ

В этом разделе мы ограничим наше внимание многоканальной передачей по фиксированным каналам, которые отличаются только ослаблениями (затуханиями) и фазовыми сдвигами. Специфическую модель для систем многоканальной передачи можно описать так. Сигнал в общем случае выражается так:

$$s_m^{(n)}(t) = \mathrm{Re}\left[s_{lm}^{(n)}(t)e^{j2\pi f_0 t}\right], \qquad 0 \le t \le T, \quad n = 1, 2, \ldots, L, \quad m = 1, 2, \ldots, M, \tag{12.1.1}$$

где L – число каналов, а M – число сигналов. Считается что сигналы имеют равные энергии и равные априорные вероятности. Сигналы $\left\{s_m^{(n)}(t)\right\}$ переданные по L каналам, получают множители ослабления (затухания) $\{\alpha_n\}$, фазовые сдвиги $\{\phi_n\}$ и искажаются аддитивным шумом. Эквивалентные низкочастотные сигналы, принимаемые по L каналам, можно выразить так

$$r_l^{(n)}(t) = \alpha_n e^{-j\phi_n} s_{lm}^{(n)}(t) + z_n(t), \quad 0 \le t \le T, \quad n = 1, 2, \ldots, L, \quad m = 1, 2, \ldots, M, \tag{12.1.2}$$

где $\left\{s_{lm}^{(n)}(t)\right\}$ – эквивалентные низкочастотные переданные сигналы, а $\{z_n(t)\}$ представляют аддитивные шумовые процессы по L каналам. Мы предполагаем, что $\{z_n(t)\}$ при различных n – взаимно статистически независимые и одинаково распределённые гауссовские шумовые случайные процессы.

Мы рассмотрим два типа обработки сигналов на приёме, именно, когерентное и некогерентное детектирование. Приёмник при когерентном детектировании оценивает канальные параметры $\{\alpha_n\}$ и $\{\phi_n\}$ и использует оценки при вычислении величин для принятия решения. Допустим, мы определим $g_n = \alpha_n e^{-j\phi_n}$ и пусть $\hat{g}_n$ – оценки g_n. Приёмник в многоканальной системе коррелирует каждый из L принимаемых сигналов с образцами соответствующих переданных сигналов, умножает выход каждого коррелятора на соответствующую оценку $\{\hat{g}_n^*\}$ и суммирует результирующие сигналы. Таким образом, величины для решения при когерентном детектировании являются корреляционные метрики

$$CM_m = \sum_{n=1}^{L} \mathrm{Re}\left[\hat{g}_n^* \int_0^T r_l^{(n)}(t)\, s_{lm}^{(n)*}(t)dt\right], \quad m = 1, 2, \ldots, M. \qquad (12.1.3)$$

При некогерентном детектировании не делается усилий по оцениванию параметров канала. Демодулятор основывает своё решение или по сумме огибающих (детектирование огибающей) или по сумме квадратов огибающих (квадратичное детектирование) выходов согласованных фильтров (корреляторов). В общем, качество, полученное при детектировании огибающей, отличается немного от качества, полученного при квадратичном детектировании в канале АБГШ. Однако квадратичное детектирование при многоканальной передаче в каналах с АБГШ значительно проще для анализа, чем детектирование огибающей. Поэтому мы ограничим наше внимание квадратичным детектированием принимаемых сигналов по L каналам, что приводит к величинам для решения

$$CM_m = \sum_{n=1}^{L} \left| \int_0^T r_l^{(n)}(t)\, s_{lm}^{(n)*}(t)dt \right|^2, \quad m = 1, 2, \ldots, M. \qquad (12.1.4)$$

Сначала рассмотрим двоичные сигналы и положим, что $s_{l1}^{(n)}, n = 1, 2, \ldots, L$ являются L переданными сигналами. Тогда ошибка совершается, если $CM_2 > CM_1$ или, что эквивалентно, если разность $D = CM_1 - CM_2 < 0$.

Для некогерентного декодирования эту разность можно выразить так

$$D = \sum_{n=1}^{L} \left(|X_n|^2 - |Y_n|^2 \right), \qquad (12.1.5)$$

где величины $\{X_n\}$ и $\{Y_n\}$ определяются так:

$$\begin{aligned} X_n &= \int_0^T r_l^{(n)}(t)\, s_{l1}^{(n)*}(t)dt, \quad n = 1, 2, \ldots, L \\ Y_n &= \int_0^T r_l^{(n)}(t)\, s_{l2}^{(n)*}(t)dt, \quad n = 1, 2, \ldots, L. \end{aligned} \qquad (12.1.6)$$

Величины $\{X_n\}$ – взаимно независимые и одинаково распределенные гауссовские случайные величины. Такие же свойства и у величин $\{Y_n\}$. Однако для определённого n, X_n и Y_n могут быть коррелированы. При когерентном детектировании разность $D = CM_1 - CM_2$ можно выразить так:

$$D = \tfrac{1}{2} \sum_{n=1}^{L} \left(X_n Y_n^* + X_n^* Y_n \right), \qquad (12.1.7)$$

где, по определению,

$$Y_n = \hat{g}_n, \qquad n = 1, 2, \ldots, L$$
$$X_n = \int_0^T r_1^{(n)}(t)\left[s_{11}^{(n)*}(t) - s_{12}^{(n)*}(t)\right]dt. \tag{12.1.8}$$

Если оценки $\{\hat{g}_n\}$ получены при наблюдении принимаемого сигнала по одному или многим сигнальным интервалам, как описано в приложении С, их статистические характеристики описываются гауссовским распределением. Тогда $\{Y_n\}$ характеризуются как взаимно независимые и одинаково распределенные гауссовские случайные величины. Такие же свойства у величин $\{X_n\}$. Как при некогерентном детектировании, мы допускаем корреляцию между X_n и Y_n, но не между X_m и Y_n для $m \neq n$.

12.1.1. Двоичные сигналы

В приложении В мы получили вероятность того, что общая квадратичная форма

$$D = \sum_{n=1}^{L}(A|X_n|^2 + B|Y_n|^2 + CX_nY_n^* + C^*X_nY_n) \tag{12.1.9}$$

комплексных гауссовских случайных величин меньше нуля.

Эта вероятность, которая дана формулой (В.21) приложения В – это вероятность ошибки двоичной многоканальной передачи в каналах с АБГШ. Определённое число частных случаев имеет важное значение.

Если двоичные сигналы противоположны, а оценки $\{g_n\}$ точные, как при когерентной ФМ вероятность ошибки определяется простейшей формулой

$$P_b = Q\left(\sqrt{2\gamma_b}\right), \tag{12.1.10}$$

где

$$\gamma_b = \frac{\mathscr{E}}{N_0}\sum_{n=1}^{L}|g_n|^2 = \frac{\mathscr{E}}{N_0}\sum_{n=1}^{L}\alpha_n^2 \tag{12.1.11}$$

– ОСШ на бит. Если все каналы идентичны $\alpha_n = \alpha$ для всех n, и, следовательно,

$$\gamma_b = \frac{L\mathscr{E}}{N_0}\alpha^2. \tag{12.1.12}$$

Видим, что $L\mathscr{E}$ – это общая энергия L переданных сигналов. Интерпретация этих результатов сводится к тому, что приёмник складывает энергии L каналов оптимальным образом. Это значит, что нет потери качества при делении суммарной энергии переданных сигналов по L каналам. Такое же качество будет в случае, когда единственный сигнал с энергией $L\mathscr{E}$ передается по одному каналу. Такое поведение имеет место только, если оценка точная $\hat{g}_n = g_n$ для всех n. Если оценки неправильные, то возникает потеря в качестве, величина которой зависит от качества оценки, что описывается в приложении С.

Точные оценки $\{g_n\}$ соответствуют экстремальному случаю. В другом экстремальном случае мы имеем двоичную ДФМ. В ДФМ оценки $\{\hat{g}_n\}$ являются просто нормированными отсчётами сигнала в смеси с шумом на выходе согласованных фильтров на предыдущем сигнальном интервале. Это наихудшая оценка, которую можно рассмотреть, используя оценки $\{g_n\}$. Для двоичной ДФМ вероятность ошибки, следующая из (В.21), равна

$$P_b = \frac{1}{2^{2L-1}}e^{-\gamma_b}\sum_{n=1}^{L-1}c_n\gamma_b^n, \tag{12.1.13}$$

где, по определению,

$$c_n = \frac{1}{n!}\sum_{k=0}^{L-1-n}\binom{2L-1}{k} \qquad (12.1.14)$$

и γ_b – ОСШ на бит, определенное (12-1-11), а для идентичных каналов в (12.1.12). Этот результат можно сравнить с вероятностью ошибки при использовании одного канала $(L=1)$. Для упрощения сравнения мы предположим, что L каналов имеют одинаковые множители ослабления. При том же значении γ_b качество многоканальной системы хуже одноканальной. Это значит что распределение общей переданной энергии по L каналам ведет к потере качества, величина которой зависит от L.

Потеря в качестве также возникает при квадратичном детектировании ортогональных сигналов, переданных по L каналам. Для двоичных ортогональных сигналов выражение для вероятности ошибки идентично по форме той, что для двоичной ДФМ, данное в (12-1-13), исключая того, что γ_b заменяется на $\frac{1}{2}\gamma_b$. Это значит, что двоичные ортогональные сигналы при некогерентном детектировании на 3 Дц хуже, чем двоичная ДФМ. Однако потеря в качестве, обусловленная некогерентным сложением сигналов, принимаемых по L каналам идентична той, что для двоичной ДФМ.

Рис.12.1.1 иллюстрирует потери, обусловленные некогерентным (квадратичным) сложением L сигналов, как функция от L.

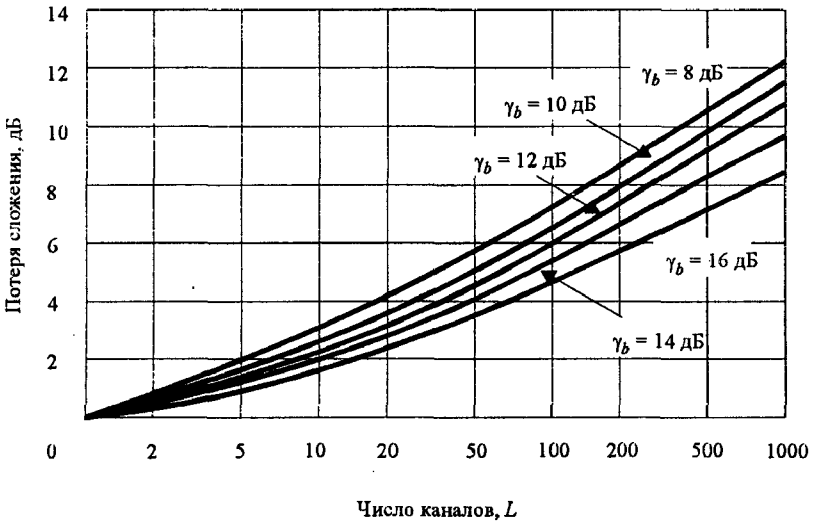


Рис. 12.1.1. Потери сложения при некогерентном детектировании и сложении двоичных многоканальных сигналов

Вероятность ошибки на рисунке не показана, но её легко определить по формуле

$$P_b = \tfrac{1}{2}e^{-\gamma_b}, \qquad (12.1.15)$$

которая определяет вероятность ошибки двоичной ДФМ, показанная на рис.5.2.12. Затем определяются изменения требуемого ОСШ на бит, γ_b, при учёте потерь некогерентного сложения, соответствующие данной величине L.

12.1.2. *M*-ичные ортогональные сигналы

Теперь рассмотрим *M*-ичные ортогональные сигналы с квадратичным детектированием и сложением сигналов L каналов. Величины для решения определяются (12-1-4). Предположим, что сигналы $s_{l1}^{(n)}(t)$, $n=1,2,\ldots,L$, переданы по L каналам с АБГШ. Тогда величины для решения выражаются так:

$$U_1 = \sum_{n=1}^{L}\left|2\mathscr{E}\alpha_n + N_{n1}\right|^2$$
$$U_m = \sum_{n=1}^{L}\left|N_{nm}\right|^2, \quad m=2,3,\ldots,M, \qquad (12.1.16)$$

где $\{N_{nm}\}$ – комплексные гауссовские случайные величины с нулевым средним и

дисперсией $\sigma^2 = \frac{1}{2}E\left(|N_{nm}|^2\right) = 2\mathscr{E}N_0$. Поэтому U_1 описывается статистически нецентральным хи-квадрат распределением с $2L$ степенями свободы и параметром нецентральности

$$s^2 = \sum_{n=1}^{L}\left(2\mathscr{E}\alpha_n\right)^2 = 4\mathscr{E}^2\sum_{n=1}^{L}\alpha_n^2. \qquad (12.1.17)$$

Используя (2.1.17), получим для ФПВ U_1

$$p(u_1) = \frac{1}{4\mathscr{E}N_0}\left(\frac{u_1}{s^2}\right)^{(L-1)/2}\exp\left(-\frac{s^2+u_1}{4\mathscr{E}N_0}\right)I_{L-1}\left(\frac{s\sqrt{u_1}}{2\mathscr{E}N_0}\right), \quad u_1 \geq 0. \qquad (12.1.18)$$

С другой стороны, распределением $\{U_m\}$, $m = 2, 3, \ldots, M$ статистически независимые и одинаково распределённые случайные величины с центральным хи-квадрат распределением с $2L$ степенями свободы для каждой. Используя (2.1.10), мы получаем ФПВ для U_m

$$p(u_m) = \frac{1}{\left(4\mathscr{E}N_0\right)^L(L-1)!}u_m^{L-1}\mathrm{e}^{-u_m/4\mathscr{E}N_0}, \quad u_m \geq 0, \qquad m = 2, 3, \ldots, M. \qquad (12.1.19)$$

Вероятность ошибочного приёма

$$\begin{aligned} P_M &= 1 - P_c = 1 - P(U_2 < U_1, U_3 < U_1, \ldots\ldots, U_M < U_1) = \\ &= 1 - \int_0^\infty [P(U_2 < u_1 | U_1 = u_1)]^{M-1} p(u_1)\, du_1. \end{aligned} \qquad (12.1.20)$$

Но

$$P(U_2 < u_1 | U_1 = u_1) = 1 - \exp\left(-\frac{u_1}{4\mathscr{E}N_0}\right)\sum_{k=0}^{L-1}\frac{1}{k!}\left(\frac{u_1}{4\mathscr{E}N_0}\right)^k. \qquad (12.1.21)$$

Таким образом

$$\begin{aligned} P_M &= 1 - \int_0^x\left[1 - \mathrm{e}^{-u_1/4\mathscr{E}N_0}\sum_{k=0}^{L-1}\frac{1}{k!}\left(\frac{u_1}{4\mathscr{E}N_0}\right)^k\right]^{M-1} p(u_1)\, du_1 \\ &= 1 - \int_0^x\left(1 - \mathrm{e}^{-\upsilon}\sum_{k=0}^{L-1}\frac{\upsilon^k}{k!}\right)^{M-1}\left(\frac{\upsilon}{\gamma}\right)^{(L-1)/2}\mathrm{e}^{-(\gamma+\upsilon)}I_{L-1}\left(2\sqrt{\gamma\upsilon}\right)d\upsilon, \end{aligned} \qquad (12.1.22)$$

где

$$\gamma = \mathscr{E}\sum_{n=1}^{L}\alpha_n^2/N_0.$$

Интеграл в (12.1.22) можно вычислить численно. Возможно также разложить $(1-x)^{M-1}$ в (12.1.22) в ряд и выполнить интегрирование почленно. Такой подход дает выражение для P_M через ограниченные суммы.

Альтернативный подход сводится к использованию объединённой границы

$$P_M < (M-1)P_2(L), \qquad (12.1.23)$$

где $P_2(L)$ – вероятность ошибки при выборе вместо U_1 одну из $M-1$ величин для решения $\{U_m\}$, $m = 2, 3, \ldots, M$. С учётом нашего предыдущего обсуждения качества двоичной ортогональной системы сигналов, имеем

$$P_2(L) = \frac{1}{2^{2L-1}}\mathrm{e}^{-k\gamma_b/2}\sum_{n=0}^{L-1}c_n\left(\tfrac{1}{2}k\gamma_b\right)^n, \qquad (12.1.24)$$

где c_n определяется (12.1.14). При относительно малых значениях M, объединённая граница в (12-1-23) достаточно плотная для большинства практических приложений.

12.2. СВЯЗЬ СО МНОГИМИ НЕСУЩИМИ

Из нашей трактовки неидеальных линейных фильтровых каналов в главах 10 и 11 мы видели, что такие каналы приводят к МСИ, которая ухудшает качество по сравнению с идеальным каналом. Степень ухудшения качества зависит от частотной характеристики канала. Далее, сложность приёмника увеличивается по мере увеличения протяжённости МСИ.

При заданной частотной характеристике канала разработчик системы связи должен решать, как эффективно использовать имеющуюся полосу частот канала для надёжной передачи информации в пределах ограничений на мощность передатчика и сложности приемника. Для неидеального линейного фильтрового канала один из выборов сводится к использованию системы с одной несущей, в которой информационная последовательность передаётся последовательно с некоторой определённой скоростью R символов/бит. В таком канале время рассеивания обычно намного больше, чем символьный интервал и, следовательно возникает МСИ из-за неидеальной частотной характеристики канала. Как мы уже отметили ранее, в этом случае необходим эквалайзер для компенсации искажений в канале.

Альтернативный подход к синтезу частотно-эффективной системы связи при наличии искажений в канале сводится к разделению имеющейся в распоряжении полосы частот на определенное число подканалов так, что каждый подканал почти идеален. Для детальной проработки предположим, что $C(f)$ – это частотная характеристика неидеального частотно-ограниченного канала с полосой W и что спектральная плотность мощности аддитивного гауссовского шума $\Phi_{nn}(f)$. Затем мы делим полосу W на $N = W/\Delta f$ подканалов с шириной Δf, где Δf выбирается достаточно малой, так что $|C(f)|^2/\Phi_{nn}(f)$ примерно постоянно в пределах каждого подканала. Далее мы хотим выбрать распределение мощности переданного сигнала по частоте $P(f)$ так, чтобы удовлетворять ограничению

$$\int_W P(f)\,df \le P_{\text{ср}}, \tag{12.2.1}$$

где $P_{\text{ср}}$ – имеющаяся в распоряжении средняя мощность передатчика. Определим пропускную способность неидеального канала с аддитивным гауссовским шумом.

12.2.1. Пропускная способность неидеального линейного фиксированного канала

Напомним, что пропускная способность идеального, частотно-ограниченного, канала с АБГШ равна

$$C' = W \log_2\left(1 + \frac{P_{\text{ср}}}{WN_0}\right), \tag{12.2.2}$$

где C' – пропускная способность в бит/с, W – полоса канала, а $P_{\text{ср}}$ – средняя мощность переданного сигнала. В системе с многими несущими с достаточно малой величиной Δf пропускная способность подканала

$$C'_i = \Delta f \log_2\left[1 + \frac{\Delta f\, P(f_i)|C(f)|^2}{\Delta f\, \Phi_{nn}(f_i)}\right]. \tag{12.2.3}$$

Тогда суммарная пропускная способность канала

$$C' = \sum_{i=1}^{N} C'_i = \Delta f \sum_{i=1}^{N} \log_2\left[1 + \frac{P(f_i)|C(f_i)|^2}{\Phi_{nn}(f_i)}\right]. \tag{12.2.4}$$

В пределе, когда $\Delta f \to 0$, мы получим пропускную способность в бит/с:

$$C' = \int_W \log_2\left[1 + \frac{P(f)|C(f)|^2}{\Phi_{nn}(f)}\right] df. \qquad (12.2.5)$$

С учётом ограничений на $P(f)$, определяемые (12.2.1) выбор $P(f)$, который максимизирует C', можно сделать путем максимизации интеграла

$$\int_W \left\{\log_2\left[1 + \frac{P(f)|C(f)|^2}{\Phi_{nn}(f)}\right] + \lambda P(f)\right\} df, \qquad (12.2.6)$$

где λ – множитель Лагранжа, который выбирается так, чтобы удовлетворить заданному ограничению (12.2.1). Используя метод вариации для обеспечения максимизации, мы находим, что оптимальное распределение мощности переданного сигнала определяется решением уравнения

$$\frac{1}{|C(f)|^2 P(f) + \Phi_{nn}(f)} + \lambda = 0. \qquad (12.2.7)$$

Следовательно, $P(f) + \Phi_{nn}(f)/|C(f)|^2$ должна быть константой (K), чья величина подстраивается с тем, чтобы удовлетворить ограничению на среднюю мощность (12.2.1). Это значит, что

$$P(f) = \begin{cases} K - \Phi_{nn}(f)/|C(f)|^2 & (f \in W), \\ 0 & (f \notin W). \end{cases} \qquad (12.2.8)$$

Выражение для пропускной способности неидеального линейного фильтрового канала с АБГШ принадлежит Шеннону (1949). Фундаментальная интерпретация этого результата сводится к тому, что мощность сигнала должна быть велика, когда канальное ОСШ $|C(f)|^2/\Phi_{nn}(f)$ велико, она должны быть мала, когда ОСШ мало.

Этот результат о распределении переданной мощности иллюстрируется на рис.12.2.1.

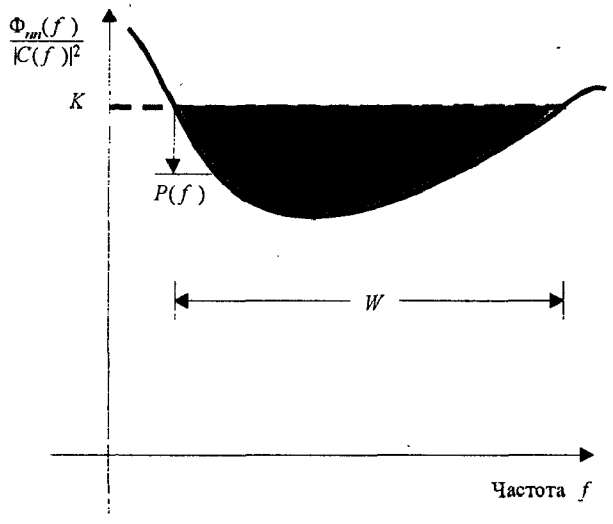


Рис. 12.2.1. Оптимальный спектр: иллюстрация решения с помощью заполнения чаши водой

Видим, что если $\Phi_{nn}(f)/|C(f)|^2$ интерпретировать как дно чаши единичной глубины и мы налили определенное количество воды равное $P_{ср}$ в чашу, вода сама распределится по чаше так чтобы достичь пропускную способность. Это называется *водо-наполняемая интерпретация* оптимального распределения мощности, как функции частоты.

Интересно отметить, что пропускная способность наименьшая, когда канальное ОСШ $|C(f)|^2/\Phi_{nn}(f)$ является константой для всех $f \in W$.

В этом случае $P(f)$ – константа для всех $f \in W$. Эквивалентно, если частотная характеристика канала идеальная, т.е. $C(f) = 1$ для $f \in W$, тогда наихудшее распределение мощности гауссовского шума, с точки зрения максимизации пропускной способности, соответствует белому гауссовскому шуму.

Вышеизложенное наводит на мысль, что передача со многими несущими, при которой доступная полоса канала делится на подполоски с относительно узкой полосой $\Delta f = W/N$, даёт решение, которое может обеспечить скорость передачи, близкую к пропускной способности канала. Сигнал в каждой подполоске можно независимо кодировать и модулировать с синхронной скоростью передачи символов $1/\Delta f$ и с оптимальным распределением мощности $P(f)$. Если Δf очень мало, тогда $C(f)$ по существу константа во всей подполоске, так что выравнивание не требуется, поскольку МСИ пренебрежимо мала.

Модуляция со многими несущими используется в модемах как радио-, так и телефонных каналах. Модуляция со многими несущими также предполагается для применения в будущем цифровом звуковом вещании.

Особо подходящее применение для систем со многими несущими – это цифровая передача по медным проводным соединительным линиям.

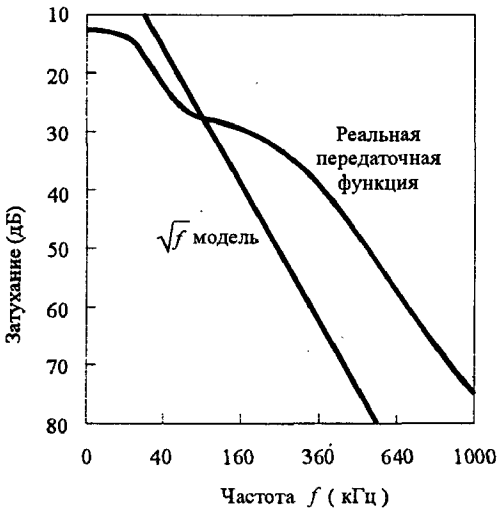


Рис.12.2.2. Коэффициент передачи 24 gauge 12 ktf PIC loop [*Werner* (1991) ©*IEEE*]

Типичная характеристика канального затухания для такой соединительной линии иллюстрируется на рис.12.2.2.

Мы видим, что коэффициент передачи быстро падает с ростом частоты. Такая характеристика сильно затрудняет достичь высокую скорость передачи при использовании одной несущей и эквалайзера на приеме[1]. Ухудшение качества из-за МСИ очень велико. С другой стороны, модуляция со многими несущими с оптимальным распределением мощности обеспечивает возможность для высокой скорости передачи.

Доминирующий шум при передаче по соединительным линиям это взаимная интерференция от сигналов других телефонных линий, расположенных в том же кабеле. Распределение мощности этого типа шума также зависит от частоты, что можно принять во внимание при распределении возможной передаваемой мощности.

Процедура синтеза системы КАМ со многими несущими для неидеального линейного фильтрованного канала была дана Калетом (1989). В этой процедуре суммарная битовая скорость максимизируется посредством расчета оптимального деления мощности по поднесущим и оптимального выбора числа бит на символ (размера сигнального созвездия КАМ) для каждой поднесущей при учете ограничений на среднюю мощность передатчика и при условии, что вероятности ошибочного приёма символов для всех поднесущих равны.

Ниже мы представим разработку модулятора КАМ со многими несущими и демодулятора, которая базируется на дискретном преобразовании Фурье (ДПФ)для генерации многих несущих.

---

[1] Разработки последних лет показывают, что системы с одной несущей вполне конкурентоспособны по сравнению с системами на многих несущих при цифровой передаче по медным проводным соединительным линиям (VDSL) с учётом реальных свойств канала (прп).

12.2.2. Система со многими поднесущими, основанная на быстром преобразовании Фурье

В этом разделе мы опишем систему связи со многими поднесущими которая использует алгоритм быстрых преобразований Фурье (БПФ) для синтеза сигнала на передаче и для демодуляции принимаемых сигналов на приеме. БПФ просто эффективный вычислительный инструмент для разработки дискретного преобразования Фурье (ДПФ).

Рис.12.2.3 иллюстрирует блок-схему системы связи со многими несущими.

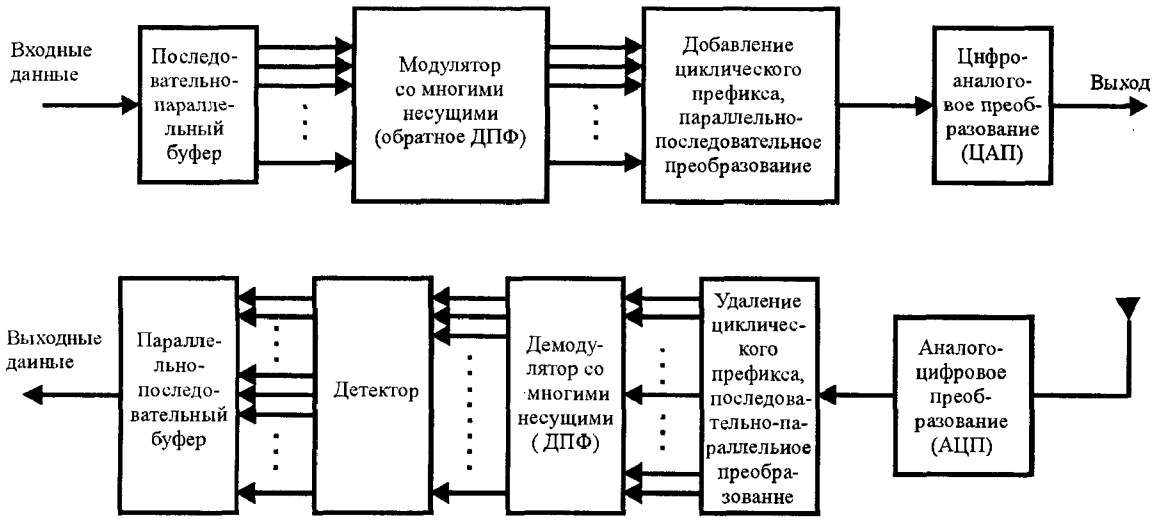


Рис.12.2.3. Система связи со многими несущими

Буферное устройство превращает информационную последовательность в параллельно передаваемые блоки по N_f бит. Каждый фрагмент из N_f бит делятся на $\tilde{N}$ групп, причём i-й группе назначается $\tilde{n}_i$ бит и

$$\sum_{i=1}^{\tilde{N}} \tilde{n}_i = N_f . \qquad (12.2.9)$$

Каждая группа может кодироваться отдельно, так что число выходных символов кодера для i-й группы $n_i \geq \tilde{n}_i$.

Удобно рассматривать модуляцию со многими несущими как модуляцию, составленную из $\tilde{N}$ независимых КАМ каналов, работающих с одинаковой скоростью $1/T$, но имеющих независимые КАМ созвездия, т.е. каждый i -й канал использует $M_i = 2^{n_i}$ сигнальных точек. Обозначим комплексные точки сигнала, соответствующие информационным символам в подканалах X_k, $k = 0, 1, \ldots, \tilde{N}-1$. Для получения многочастотного сигнала с $\tilde{N}$ поднесущими по информационным символам $\{X_k\}$ воспользуемся обратным ДПФ (ОДПФ).

Однако, если мы вычислим $\tilde{N}$-точечное ОДПФ по $\{\hat{X}_k\}$, то получим комплексную последовательность чисел, которая не эквивалентна $\tilde{N}$ КАМ-модулированным поднесущим. Вместо этого введём $N = 2\tilde{N}$ информационных символов, определённых следующим образом:

$$X_{N-k} = X_k^*, \quad k = 1, \ldots, \tilde{N}-1 \qquad (12.2.10)$$

и $X_0' = \mathrm{Re}(X_0)$, $X_N = \mathrm{Im}(X_0)$. Таким образом, символ X_0 распадается на две части,

причём обе части – вещественные. Теперь N-точечное ОДПФ порождает вещественную последовательность

$$x_n = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X_k e^{j2\pi nk/N}, \quad n = 0, 1, \ldots, N-1, \qquad (12.2.11)$$

где $1/\sqrt{N}$ – масштабный множитель.

Последовательность $\{x_n, 0 \le n \le N-1\}$ соответствует отсчётам суммы $x(t)$ $\tilde{N}$ сигналов по всем поднесущим:

$$x(t) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X_k e^{j2\pi kt/T}, \quad 0 \le t \le T, \qquad (12.2.12)$$

где T – символьный интервал. Видно, что частоты поднесущих равны $f_k = k/T$, $k = 0, 1, \ldots, \tilde{N}$. Далее дискретная во времени последовательность $\{x_n\}$ по (12.2.11) представляет собой отсчёты $x(t)$, взятые в моменты времени $t = nT/N$, где $n = 0, 1, \ldots, N-1$.

Вычисление ОДПФ по $\{X_k\}$, как видно из (12.2.11), можно рассматривать как умножение каждой точки данных X_k на соответствующий вектор

$$\mathbf{v}_k = [\upsilon_{k0}, \upsilon_{k1} \cdots \upsilon_{k(N-1)}], \qquad (12.2.13)$$

где

$$\upsilon_{kn} = \frac{1}{\sqrt{N}} e^{j(2\pi/N)kn}, \qquad (12.2.14)$$

как показано на рис.12.2.4. Во многих случаях вычисление ДПФ может быть эффективно выполнено с использованием алгоритма быстрого преобразования Фурье (БПФ).

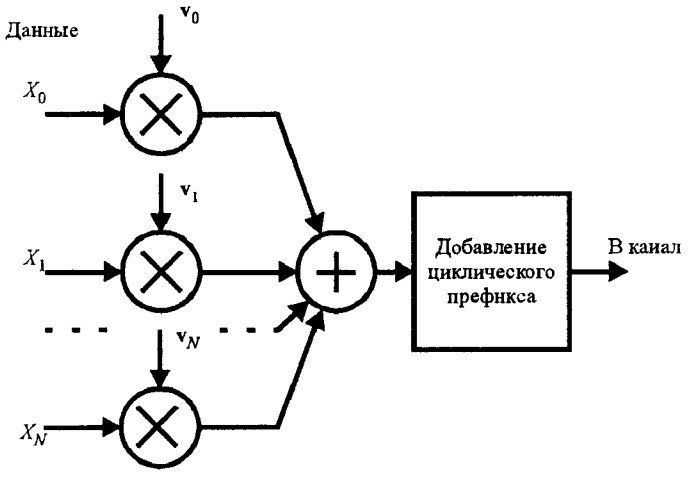


Рис. 12.2.4. Синтез сигнала для модуляции со многими несущими на основе обратного ДПФ

На практике отсчёты сигнала $\{x_n\}$ подаются на цифро-аналоговый преобразователь (ЦАП), а затем на ФНЧ, на выходе которого образуется канальный сигнал $x(t)$.

Проходя по каналу связи, этот сигнал искажается:

$$r(t) = x(t) \otimes h(t) + n(t), \qquad (12.2.15)$$

где $h(t)$ – импульсная характеристика канала, $\otimes$ – знак свёртки, а $n(t)$ – помеха.

При выборе полосы частот каждого подканала Δf достаточно малой величины длительность элементарного символа $T = 1/\Delta f$ становится большей, чем интервал временного рассеяния канала. Точнее говоря, мы вправе предположить, что это рассеяние охватывает $v+1$ отсчётов сигнала, причём $v << N$. При этом для устранения явления МСИ можно использовать защитный интервал длительностью vT/N, вводимый между смежными сигнальными блоками.

Другой метод борьбы с МСИ состоит в присоединении циклически повторяемого префикса между блоками, состоящими из N отсчётов сигнала $\{x_0, x_1, \ldots, x_{N-1}\}$. Этот

префикс состоит из отсчётов $x_{N-\nu}, x_{N-\nu+1}, \ldots, x_{N-1}$. Они присоединяются к началу каждого блока. Таким образом, длина каждого блока увеличивается до $N+\nu$ отсчётов, и их теперь можно индексировать величинами $n=-\nu, \ldots, N-1$, причём первые ν отсчётов образуют префикс. Тогда, если $\{h_n, 0 \le n \le \nu\}$ означают отсчёты импульсной характеристики канала, то их свёртка с $\{x_n, -\nu \le n \le N-1\}$ даёт $\{r_n\}$ – принимаемую последовательность. Нас интересуют лишь отсчёты $\{r_n\}$ для $0 \le n \le N-1$, по которым можно восстановить переданную информационную последовательность, используя для демодуляции N - точечное ДПФ. Следовательно, первые ν отсчётов $\{r_n\}$ отбрасываются за ненадобностью.

С частотной точки зрения, если дана импульсная характеристика канала $\{h_n, 0 \le n \le \nu\}$, то коэффициент передачи для -й поднесущей равен

$$H_k = H\left(\frac{2\pi k}{N}\right) = \sum_{n=0}^{\nu} h_n \mathrm{e}^{-j2\pi nk/N} . \qquad (12.2.16)$$

Благодаря префиксу смежные сигнальные блоки не интерферируют и, следовательно, демодулированная последовательность может быть представлена в виде

$$\hat{X}_k = H_k X_k + \eta_k, \quad k = 0, 1, ..., N-1, \qquad (12.2.17)$$

где $\{\hat{X}_k\}$ – выход демодулятора, т.е. N - точечного ДПФ , а η_k – ошибка, обусловленная аддитивным шумом. Заметим, что, выбирая $N >> \nu$, можно потерю скорости, вызванную префиксными вставками, сделать пренебрежимо малой.

Как показано на рис.12.2.3, сигнал демодулируется путём вычисления ДПФ после аналого-цифрового преобразования. При этом ДПФ может рассматриваться как перемножение отсчётов принятого сигнала $\{r_n\}$ с $\mathbf{v}_n^*$, где $\mathbf{v}_n$ определено в (12.2.13). Как и в случае с модулятором, вычисление ДПФ может быть эффективно выполнено с использованием алгоритма быстрого преобразования Фурье (БПФ).

При необходимости несложно оценить и скомпенсировать канальный множитель $\{H_k\}$ перед тем, как передать $\hat{X}_k$ в детектор и декодер. Для этого может быть использован «обучающий» сигнал либо в виде известной модулирующей последовательности на каждой поднесущей, либо в виде немодулированной несущей. Если параметры канала изменяются во времени медленно, то можно отследить их изменения во времени с использованием решений с выхода детектора или даже декодера. Таким образом, система передачи со многими поднесущими может быть сделана адаптивной.

Многочастотная КАМ, описанная выше, может быть использована во многих практических приложениях, в том числе в высокоскоростной передаче данных по телефонным линиям.

При выборе способа реализации ДПФ необходимо иметь в виду, что с точки зрения вычислительной сложности при $N \le 32$ выгоднее использовать обычное ДПФ, а при $N > 32$ – БПФ, причём выигрыш быстро растёт с увеличением N.

Есть одно ограничение на использование ДПФ в модуляторах и демодуляторах, обусловленное относительно большими боковыми лепестками частотной характеристики, присущими фильтрам ДПФ-типа. Первый лепесток лишь на 13 дБ ниже основного максимума, соответствующего выбранной поднесущей. Поэтому все приложения, основанные на использовании банков ДПФ-фильтров, уязвимы по отношению к межканальной интерференции (МКИ). Если вследствие канальных аномалий МКИ является проблемой, то следует использовать другие банки цифровых фильтров, которые

не имеют таких боковых лепестков частотной характеристики. В частности, несомненно привлекательным является класс многоскоростных банков цифровых фильтров, основанных на вейвлетных преобразованиях (см. Тзаннес и др., 1994; Ризос и др., 1994)

12.3. БИБЛИОГРАФИЧЕСКИЕ ЗАМЕЧАНИЯ И ССЫЛКИ

Многоканальная передача сигналов используется в меняющихся во времени каналах для преодоления влияния замираний сигналов.

Эта тема рассматривается более подробно в главе 14, где мы приводим большое число ссылок на опубликованные работы. Особо важно для трактовки многоканальной цифровой связи, данной в этой главе, являются две публикации Прайса (1962 a, b).

Имеется большое количество литературы по системам цифровой связи со многими несущими. Такие системы разработаны и используются около 30 лет. Одна из ранних систем, описанная Дольцем и др. (1957) и названная Kineplex, была использована для цифровой передачи в ВЧ диапазоне. Другая ранняя работа по синтезу систем со многими несущими была изложена в статьях Чанга (1966) и Зальцбурга (1967). Использование ДПФ для модуляции и демодуляции в системах со многими несущими было предложено Вайнштейном и Эбертом (1971).

Особый интерес за последние годы проявляется к использованию систем со многими несущими для цифровой передачи данных, факса и видеосигналов по набору каналов, включая узкополосные (4 кГц) коммутируемые телефонные сети, групповые (48 кГц) телефонные тракты, цифровые соединительные линии, сотовые радиоканалы и каналы звукового вещания. Интересующемуся читателю можно рекомендовать много статей в литературе. Мы цитируем, для примера, статьи Хиросаки и др. (1981, 1986), Чоу и др. (1991) и обзорную статью Бингама (1990). Статья Калета (1989) даёт процедуру синтеза для оптимизации скорости передачи в системе со многими несущими и КАМ при заданных ограничениях на канальные характеристики и мощность передатчика.

В заключение мы упомянем книгу Вайдьянатана(1993) и статьи Тзаннеса и др. (1994) и Ризоса и др. (1994) с трактовкой банков многоскоростных цифровых фильтров.

ЗАДАЧИ

12.1. $X_1, X_2, \ldots, X_N$ образуют набор из N статистических независимых и одинаково распределенных вещественных гауссовских случайных величин с моментами $E(X_n) = m$ и дисперсиями $E(X_n - m)^2 = \sigma^2$.

a) Определите $U = \sum_{n=1}^{N} X_n$.

Рассчитайте ОСШ для U, определённое так:

$$(\text{ОСШ})_U = \frac{[E(U)]^2}{2\sigma_U^2},$$

где σ_U^2 – дисперсия для U.

b) Определите

$$V = \sum_{n=1}^{N} X_n^2.$$

Рассчитайте ОСШ для V, определённое так:

$$(\text{ОСШ})_V = \frac{[E(V)]^2}{2\sigma_V^2},$$

где σ_V^2 – дисперсия для V.

c) Постройте зависимость $(\text{ОСШ})_U$ и $(\text{ОСШ})_V$ от m^2/σ^2 на одном графике и таким образом сравните эти отношения графически.

d) Какой выигрыш даст результат (c) при когерентном детектировании по сравнению с квадратичным детектированием и сложением многоканальных сигналов?

12.2. Двоичная система связи передает одну и ту же информацию по двум разнесённым каналам. Два принимаемых сигнала равны

$$r_1 = \pm\sqrt{E_b} + n_1,$$
$$r_2 = \pm\sqrt{E_b} + n_2,$$

где $E(n_1) = E(n_2) = 0$, $E(n_1^2) = \sigma_1^2$, $E(n_2^2) = \sigma_2^2$, и n_1, n_2 являются некоррелированными гауссовскими случайными величинами.

Детектор базирует своё решение на линейной комбинации r_1, r_2, т.е.

$$r = r_1 + kr_2.$$

a) Определите величину k, которая минимизирует вероятность ошибки.

b) Нарисуйте вероятность ошибки для $\sigma_1^2 = 1$, $\sigma_2^2 = 3$ и $k = 1$ и когда k принимает оптимальное значение, найденное в (a). Сравните результаты.

12.3. Определите цену циклически повторяемого префикса (используемого при модуляции со многими несущими для преодоления МСИ) через

a) увеличение полосы частот канала.

b) увеличение энергии сигнала.

12.4. Пусть $x(n)$ – дискретный во времени сигнал ограниченный длины N и пусть $X(k)$ – его N – точечное ДПФ. Предположим, что мы дополним $x(n)$ L нулями и вычислим $(N+L)$ – точечные ДПФ $X'(k)$. Какова связь между $X(0)$ и $X'(0)$? Нанося $|X(k)|$ и $|X'(k)|$ на один график, объясните связь между двумя кривыми.

12.5. Покажите, что последовательность $\{x_n\}$, определенное (12.2.11), соответствует отсчетам сигнала $x(t)$, определенного (12.1.12).

12.6. Покажите, что ОДПФ последовательности $\{X_k,\ 0 \le k \le N-1\}$ можно вычислить путем пропускания последовательности $\{X_k\}$ через набор N линейных фильтров дискретного времени с системными функциями

$$H_n(z) = \frac{1}{1 - e^{j2\pi n/N} z^{-1}}.$$

12.7. Нарисуйте $P_2(L)$ для $L = 1$ и $L = 2$ как функцию от $10\lg\gamma_b$ и определите потери в ОСШ, обусловленные потерями сложения для $\gamma_b = 10$.

13

ШИРОКОПОЛОСНЫЕ СИГНАЛЫ ДЛЯ ЦИФРОВОЙ СВЯЗИ (СИГНАЛЫ С РАССЕЯННЫМ СПЕКТРОМ)

Широкополосные сигналы (сигналы с рассеянным спектром), используемые для передачи цифровой информации, отличаются тем, что их полоса частот W намного больше, чем информационная скорость R бит/с. Это значит, что показатель расширения спектра $B_c = W/R$[1] для широкополосных сигналов намного больше единицы. Большая избыточность, свойственная широкополосным сигналам, требуется для преодоления высоких уровней интерференции, возникающая при передаче цифровой информации по некоторым радио- и спутниковым каналам. Поскольку кодированный сигнал также характеризуется показателем расширения спектра большим единицы и кодирование является эффективным методом введения избыточности, следует, что кодирование-важный элемент при синтезе широкополосных сигналов.

Второй важный элемент, используемый при синтезе широкополосных сигналов – это псевдослучайность, которая делает сигналы похожими на случайный шум и трудными для демодуляции «чужими» приемниками. Этот фактор тесно связан с применением таких сигналов.

Для конкретности укажем, что широкополосные сигналы используются для:

- борьбы или подавления вредного влияния мешающих сигналов (jamming), интерференции, возникающей от других пользователей канала, и собственной интерференции, обусловленной распространением сигналов,
- обеспечения скрытности сигнала путем его передачи с малой мощностью, что затрудняет его детектирование не предназначенными слушателями в присутствии основного шума,
- достижения защиты сообщения от других слушателей.

Кроме связи, широкополосные сигналы используются для получения точных дальностей (задержек сигнала во времени) и перемещений при измерениях в радиолокации и навигации.

Ради краткости мы ограничим наше обсуждение приложением широкополосных сигналов к цифровым системам связи.

Для борьбы с преднамеренной помехой (мешающими сигналами) для вступающих в связь важно, что источник мешающего сигнала, который пытается разрушить связь, не имеет априорной информации о характеристиках сигнала, исключая значения общей полосы частот и типа модуляции (ФМ, ЧМ и т.д.), которые используются. Если цифровая информация закодирована, как описано в главе 8, изощрённый постановщик помех (jammer) может с легкостью имитировать полезный сигнал, излученный передатчиком, и, таким образом, сильно навредить получателю. Чтобы это устранить, передатчик вводит элемент случайности (псевдослучайности) в каждом из передаваемых цифровых сигналов, который известен получателю, но неизвестен постановщику помех. Как следствие,

[1] B_c часто называют базой сигнала (прп.).

источник мешающего сигнала вынужден синтезировать и передавать свой сигнал без знания псевдослучайного образца.

Интерференция от других пользователей возникает в системах связи со множественным доступом, в которых определённое число пользователей владеют совместно общей полосой частот. Эти пользователи могут передавать информацию одновременно в общей полосе к соответствующим получателям. Предполагая, что все из этих пользователей используют один и тот же код для кодирования соответствующих информационных последовательностей, передаваемые сигналы в этой общей полосе можно отличить друг от друга при использовании для каждого переданного сигнала различного псевдослучайного образца, также называемых *кодом* или *адресом*. Таким образом, частный получатель может восстановить передаваемую информацию, если знает свой псевдослучайный образец, т.е. ключ, используемый соответствующим передатчиком. Этот тип техники связи, который позволяет многим пользователям совместно использовать общий канал для передачи информации, называется кодовым разделением при множественном доступе (МДКР или CDMA – CODE DIVISION MULTIPLE ACCESS). CDMA будет рассматриваться в разделах 13.2 и 13.3.

Возникающие многолучевые компоненты при распространении волн в диспергирующем канале с рассеянием можно рассматривать как вид собственной интерференции. Этот вид интерференции также можно подавить введением псевдослучайного образца в переданном сигнале, как будет описано ниже.

Сообщение может быть «спрятано» в основном шуме путем его рассеяния по полосе частот кодированием и передачей результирующего сигнала низким уровнем. Говорят, что из-за своего низкого уровня мощности переданный сигнал является "закрытым". Имеется малая вероятность перехватить такой сигнал (детектировать его случайным слушателем, поэтому его также называют сигналом с низкой вероятностью перехвата (НВП – LPI).

Наконец, закрытость сообщения можно получить путем введения псевдослучайного образца в передаваемом сообщении. Сообщение может детектировать получатель, который знает псевдослучайный образец или ключ, используемый при передаче, но не могут детектировать другие получатели, которые не знают ключ.

В следующих разделах мы опишем различные типы широкополосных сигналов, их характеристики и применение. Акцент будет сделан на использование широкополосных сигналов для радиопротиводействия (РП или глушения) или антирадиопротиводействия (АРП), для CDMA и для НВП. Вкратце опишем виды канальных характеристик, предполагаемых для применений, названных выше.

13.1. МОДЕЛЬ ЦИФРОВЫХ СИСТЕМ СВЯЗИ С ШИРОКОПОЛОСНЫМИ СИГНАЛАМИ

Блок-схема рис.13.1.1 иллюстрирует базовые элементы цифровой системы связи с широкополосными сигналами для передачи двоичных сообщений.

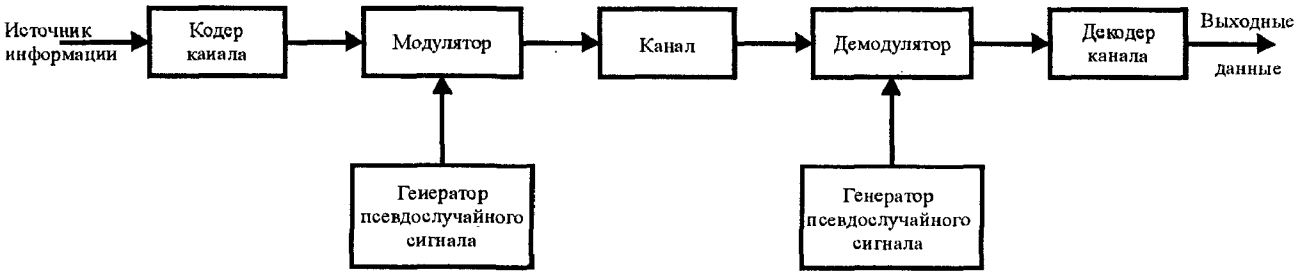


Рис. 13.1.1. Модель цифровой системы связи с широкополосным сигналом

Канальный кодер, декодер, модулятор и демодулятор являются базовыми элементами системы, которые были обсуждены в главах 5, 7 и 8. В дополнение к этим элементам мы имеем два идентичных генератора псевдослучайных образцов. Один коммутируется с модулятором на передающем конце, а второй коммутируется с демодулятором на приемном конце. Генератор генерирует псевдослучайные или псевдошумовые (ПШ) двоичные последовательности, которые вводятся в передаваемый сигнал модулятором и удаляются из принимаемого сигнала демодулятором.

Для демодуляции требуется синхронизация ПШ последовательности, генерируемой на приёме, с ПШ последовательностью, содержащейся в принимаемом сигнале. Первоначально до передачи информации, синхронизация может быть достигнута передачей фиксированного псевдослучайного образца, который приемник должен узнать в присутствии интерференции с высокой вероятностью. После того, как синхронизация во времени генераторов обеспечена, может начинаться передача информации.

Интерференцию в информационный сигнал вводит канал. Характеристики этой интерференции зависят в большой степени от ее происхождения. Она классифицируется как широкополосная или узкополосная относительно полосы частот информационного сигнала и как непрерывная или импульсная (не непрерывная) во времени. Для примера, сигнал глушения может состоять из одной или более синусоид в полосе, используемой для передачи информации. Частоты синусоид могут оставаться фиксированными или они могут меняться во времени согласно некоторому закону. Как второй пример, интерференция, создаваемая в CDMA другими пользователями канала, может быть широкополосной или узкополосной в зависимости от типа спектра сигнала, который используется для достижения множественного доступа. Если интерференция широкополосная, ее можно характеризировать как эквивалентный АБГШ. Мы хотим рассмотреть этот вид интерференции и некоторые другие в последующих разделах.

Наша трактовка широкополосных сигналов будет сконцентрирована на качестве цифровых систем связи в присутствии широкополосной и узкополосной интерференции. Рассматривается два вида модуляции ФМ и ЧМ. ФМ предполагается к применению, когда поддерживается фазовая когерентность между передаваемым и принимаемым сигналами на временном интервале, который относительно большой по сравнению с обратной величиной полосы частот сигнала. С другой стороны, предполагается к применению ЧМ, когда физически когерентность не может быть поддержана из-за изменений во времени характеристик линии. Это может быть случай линии связи между двумя высокоскоростными самолетами или между высокоскоростным самолетом и неподвижной станцией.

ПШ последовательность, генерируемая у модулятора, используется в соединении с ФМ модуляцией для псевдослучайного сдвига фазы ФМ сигнала, как описано в разделе 13.2. Результирующий модулированный сигнал назван прямой последовательностью (ПП, DS – direct sequence) или псевдошумовым (ПШ) широкополосным сигналом. Если ПШ последовательность использовать в соединении с двоичной или M-ичной (M>2) ЧМ, то псевдослучайные последовательности выбирают частоту передаваемого сигнала псевдослучайно. Результирующий сигнал называется в этом случае широкополосным сигналом со скачками частоты (СЧ, FH – frequency hopped). Хотя вкратце будут описаны некоторые другие типы широкополосных сигналов, акцент в нашем рассмотрении делается на широкополосный сигнал ПШ и ПЧ.

13.2. ШИРОКОПОЛОСНЫЕ СИГНАЛЫ С ПРЯМЫМИ ПСЕВДО-ШУМОВЫМИ (ПШ) ПОСЛЕДОВАТЕЛЬНОСТЯМИ

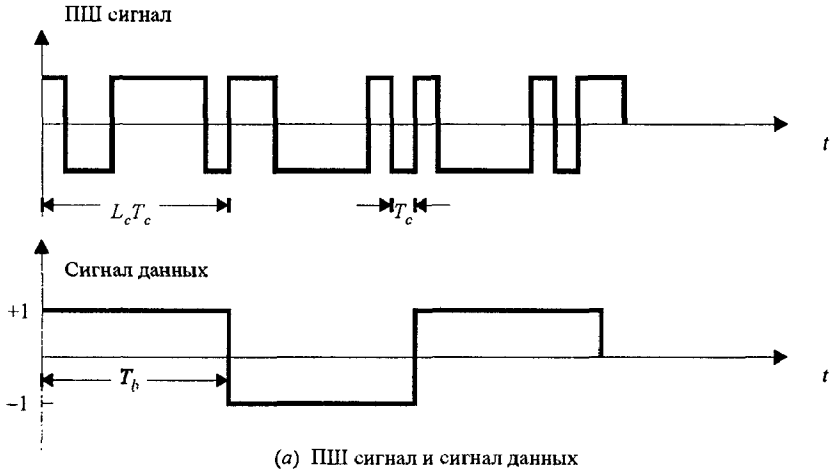
В модели, показанной на рис.13.2.1, мы предполагаем, что информационная скорость на входе кодера равна R бит/с, а доступная полоса частот канала равна W Гц. Модуляция сигнала – двоичная ФМ. Чтобы использовать всю доступную полосу частот, начальная фаза несущей сдвигается псевдослучайно, в соответствии с образцом ПШ генератора со скоростью W раз/с. Обратная величина W, обозначенная T_c, определяет длительность прямоугольного импульса, который называется чипом в то время как T_c называется чип интервалом. Этот импульс – базовый элемент широкополосного сигнала ПП.

Если мы определим $T_b = 1/R$ как длительность импульса, которая соответствует передаче одного информационного символа, показатель расширения полосы частот W/R можно выразить так

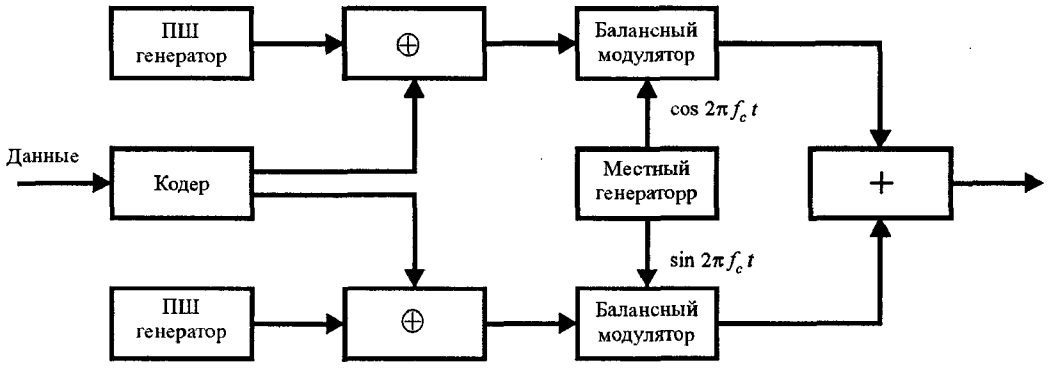
$$B_c = \frac{W}{R} = \frac{T_b}{T_c} \qquad (13.2.1)$$

В практических системах T_b / T_c – целое число

$$L_c = \frac{T_b}{T_c}, \qquad (13.2.2)$$



(*a*) ПШ сигнал и сигнал данных



(*b*) ПШ-КФМ модулятор

Рис.13.2.1. ПШ сигнал и сигнал данных (*a*)
и КФМ модулятор (*b*) для ПШ широкополосных систем

которое определяет число чипов на информационный символ. Это значит, что L_c – это число фазовых сдвигов, которые возникают в переданном сигнале на символьном интервале $T_b = 1/R$. Рис.13.2.1(а) иллюстрирует взаимосвязь между ПШ сигналом и сигналом данных.

Предположим, что кодер берет каждый раз на временном интервале kT_b k информационных символов и генерирует (n,k) двоичный линейный блоковый код. Интервал времени, выделенный для передачи n элементов кода, равен kT_b с, число чипов на этом временном интервале равно kL_c. Мы можем выбрать длину блока кода $n = kL_c$, тогда скорость кода k/n. Если кодер генерирует двоичный сверточный код со скоростью k/n, число чипов на временном интервале kT_b, также равно $n = kL_c$. Значит, нижеследующее обсуждение применимо как к блоковым, так и к свёрточным кодам.

Один метод для включения ПШ последовательности в передаваемый сигнал сводится к непосредственному изменению кодовых символов их суммированием по mod 2 с ПШ последовательностью.

Если используется четырёхфазная ФМ, то одна ПШ последовательность суммируется с информационной последовательностью и передается синфазной несущей, а вторая ПШ последовательность суммируется с другой информационной последовательностью и передается квадратурной несущей. Во многих широкополосных системах одна и та же двоичная информационная последовательность суммируется с двумя ПШ последовательностями для формирования двух квадратурных компонент. Таким образом, генерируется сигнал четырёхфазной ФМ с двоичным информационным потоком.

Итак, каждый информационный символ изменяется путём его суммирования с символом ПШ последовательности. Если b_i представляет i-й символ ПШ последовательности, а c_i – соответствующий символ на входе кодера, то сумма по mod 2 даёт

$$a_i = b_i \oplus c_i . \qquad (13.2.3)$$

Таким образом, $a_i = 1$ если $b_i = 1$ и $c_i = 0$ или $b_i = 0$ и $c_i = 1$; также $a_i = 0$ когда $b_i = 1$ и $c_i = 1$ или когда $b_i = 0$ и $c_i = 0$. Мы можем сказать, что $a_i = 0$, когда $b_i = c_i$ и $a_i = 1$, когда $b_i \neq c_i$. Последовательность $\{a_i\}$ отображается в двоичный ФМ сигнал в виде $s(t) = \pm \mathrm{Re}\left[g(t)^{j2\pi ft}\right]$ согласно правилу:

$$g_i(t) = \begin{cases} g(t - iT_c), & (a_i = 0) \\ -g(t - iT_c), & (a_i = 1), \end{cases} \qquad (13.2.4)$$

где $g(t)$ представляет импульс длительностью T_c с произвольной огибающей.

Суммирование по mod 2 кодовой последовательности $\{c_i\}$ и последовательности $\{b_i\}$ ПШ генератора можно также представить как умножение двух сигналов. Чтобы продемонстрировать эту точку зрения, предположим, что элементы кодовой последовательности отображаются в двоичный ФМ сигнал согласно отношению

$$c_i(t) = (2c_i - 1)g(t - iT_c) . \qquad (13.2.5)$$

Аналогично определяем сигнал $p_i(t)$ так

$$p_i(t) = (2b_i - 1)p(t - iT_c) , \qquad (13.2.6)$$

где $p(t)$ – прямоугольный импульс длительности T_c. Тогда эквивалентный низкочастотный передаваемый сигнал, соответствующий, i-му кодовому символу равен

$$g_i(t) = p_i(t)c_i(t) = (2b_i - 1)(2c_i - 1)g(t - iT_c). \qquad (13.2.7)$$

Этот сигнал идентичен тому, который определяется (13.2.4) и который получен от последовательности $\{a_i\}$. Следовательно, суммирование по mod 2 кодовых символов с ПШ последовательностью ведет к отображению в ФМ сигнале, который эквивалентен умножению двоичного ФМ сигнала, генерированного кодовыми символами, с последовательностью прямоугольных символов единичной амплитуды, каждый длительностью T_c, и полярностью, которая определяется ПШ последовательностью согласно (13.2.6). Хотя легче реализовать суммирование по mod 2 и затем ФМ вместо умножения сигналов, удобно для целей демодуляции рассматривать сигнал в мультипликативной форме, определяемой (13.2.7). Функциональная блок-схема четырехфазовой ФМ с ПП широкополосным сигналом показан на рис.13.2.1(b).

Принимаемый эквивалентный низкочастотный сигнал для i-го кодового элемента определяется так[1]:

$$r_i(t) = p_i(t)c_i(t) + z(t), \qquad iT_c \le t \le (i+1)T_c$$
$$r_i(t) = (2b_i - 1)(2c_i - 1)g(t - iT_c) + z(t), \qquad (13.2.8)$$

где $z(t)$ представляет сигнал интерференции или глушения, который искажает информацию в полезном сигнале. Интерференция считается стационарным случайным процессом с нулевым средним.

Если $z(t)$ является отсчётной функцией комплексного гауссовского процесса, оптимальный демодулятор можно реализовать или как фильтр, согласованный с сигналом $g(t)$, или как коррелятор, как показано на блок-схемах рис.13.2.2. При реализации согласованным фильтром выход согласованного фильтра умножается на величины $(2b_i - 1)$, которые получаются от генератора ПШ на приёме, когда он правильно синхронизирован. Поскольку $(2b_i - 1)^2 = 1$, когда $b_i = 0$ и $b_i = 1$, то влияние ПШ последовательности на принимаемый кодовый символ устраняется.

На рис.13.2.2 мы также видим, что взаимная корреляция может быть выполнена одним из двух путей. Первый, иллюстрируемый на рис.13.2.2(b), выполняет предварительное умножение $r_i(t)$ на сигнал $p_i(t)$, генерируемый ПШ генератором, а затем осуществляется взаимная корреляция с $g^*(t)$ и стробирование выхода на интервале каждого чипа. Второй метод, иллюстрированный на рис.13.2.2(c), включает сначала взаимную корреляцию принимаемого сигнала с $g^*(t)$, стробирование выхода коррелятора и затем умножение этого выхода на $(2b_i - 1)$, которое получается из ПШ генератора.

Если $z(t)$ не является гауссовским случайным процессом, методы демодуляции, иллюстрируемые на рис.13.2.2, не являются больше оптимальными, тем не менее, мы можем всё же использовать любую из этих трех структур демодулятора для демодуляции принимаемого сигнала. Если статистические характеристики интерференции $z(t)$ неизвестны априори, это, конечно, один из возможных подходов. Альтернативный метод, который описывается ниже, использует адаптивный фильтр до согласованного фильтра или коррелятора для подавления узкополосной интерференции. Целесообразность этого второго метода также излагается ниже.

---

[1] Для простоты мы считаем, что канальное ослабление $\alpha = 1$, а фазовый сдвиг в канале нулевой. Поскольку предполагается когерентное детектирование ФМ сигнала, то произвольный фазовый сдвиг в канале компенсируется до демодулятора.

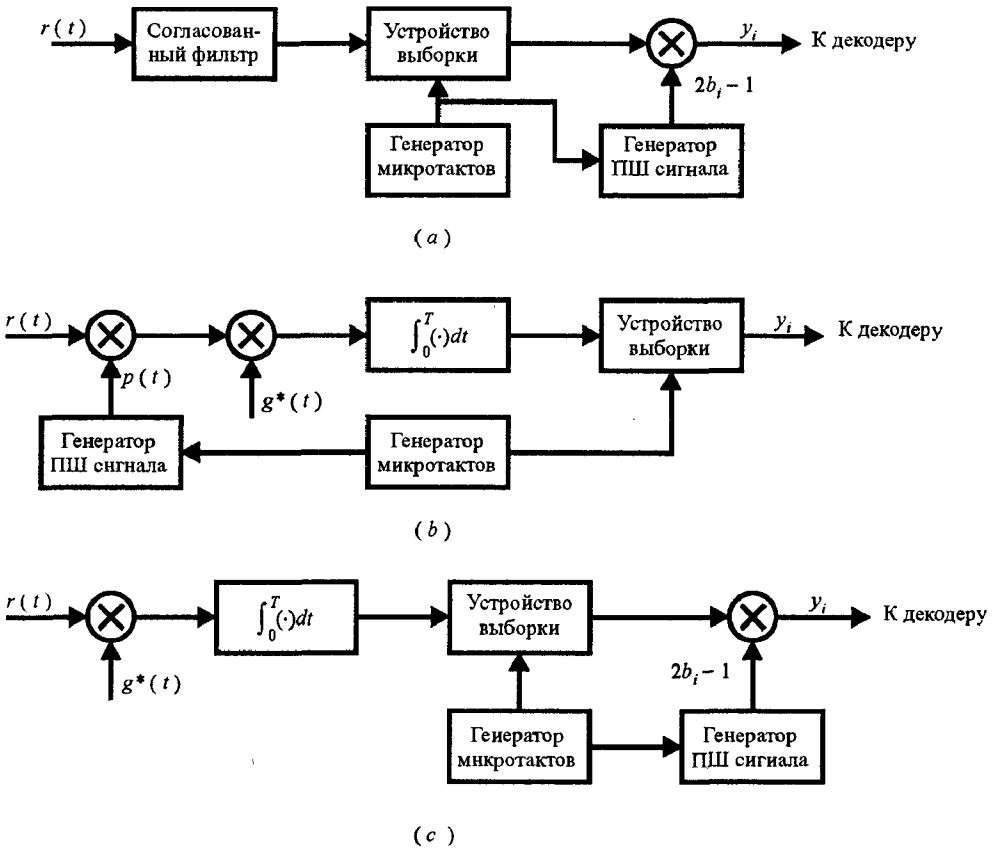


Рис. 13.2.2. Возможные структуры демодуляторов для псевдошумовых (ПШ) широкополосных сигналов

В разделе 13.2.2 мы определим вероятность ошибки для широкополосной системы с ПП (с рассеянным спектром) в присутствии широкополосной и узкополосной интерференции. Расчет базируется на предположении, что демодулятор имеет одну из трёх эквивалентных структур, показанных на рис.13.2.2.

13.2.1. Качество декодера

Обозначим неквантованный выход демодулятора через y_j, $1 \le j \le n$. Сначала рассмотрим линейный двоичный (n,k) блоковый код и, без потери общности, предположим, что передается кодовое слово из одних нулей.

Декодер, который выполняет детектирование мягких решений, вычисляет корреляционные метрики

$$CM_i = \sum_{j=1}^{n}(2c_{ij} - 1)y_j, \qquad i = 1, 2, ..., 2^k, \tag{13.2.9}$$

где c_{ij} означает j-й символ в i-м кодовом слове. Корреляционная метрика, соответствующая кодовому слову из одних нулей, равна

$$CM_1 = 2n\mathscr{E}_c + \sum_{j=1}^{n}(2c_{1j} - 1)(2b_j - 1)v_j =$$

$$= 2n\mathscr{E}_c - \sum_{j=1}^{n}(2b_j - 1)v_j, \tag{13.2.10}$$

где v_j, $1 \le j \le n$ – аддитивное шумовое слагаемые, которые искажает j-й кодовый символ, а $\mathscr{E}_c$ – энергия чипа. Величина v_j определяется так:

$$v_j = \mathrm{Re}\left\{\int_0^{T_c} g^*(t) z[t + (j-1)T_c]dt\right\}, \qquad j = 1, 2, \ldots, n. \quad (13.2.11)$$

Аналогично, корреляционная метрика, соответствующая кодовому слову $\mathbf{C}_m$ с весом w_m, равна

$$CM_M = 2\xi_c n\left(1 - \frac{2w_m}{n}\right) + \sum_{j=1}^{n}(2c_{mj} - 1)(2b_j - 1)v_j. \quad (13.2.12)$$

Следуя примеру, использованной в разделе 8.1.4, мы можем определить вероятность того, что $CM_m > CM_1$. Разница между CM_1 и CM_m равна

$$D = CM_1 - CM_m = 4\mathscr{E}_c w_m - 2\sum_{j=1}^{n} c_{mj}(2b_j - 1)v_j. \quad (13.2.13)$$

Поскольку кодовое слово $\mathbf{C}_m$ имеет вес w_m, имеется w_m ненулевых компонент при суммировании шумовых слагаемых в (13.2.13). Мы можем предположить, что минимальное расстояние кода достаточно велико, так что мы можем обратиться к центральной предельной теореме при суммировании шумовых компонент. Это предположение имеет силу для широкополосного сигнала ПШ, который имеет показатель расширения спектра 20 или больше[1]. Таким образом, сумма шумовых компонент моделируется гауссовской случайной величиной. Поскольку $E(2b_j - 1) = 0$ и $E(v_j) = 0$, среднее для второго слагаемого в (13.2.13) также равно нулю. Его дисперсия

$$\sigma_m^2 = 4\sum_{j=1}^{n}\sum_{i=1}^{n} c_{mi}c_{mj}E\left[(2b_j - 1)(2b_i - 1)\right]E(v_i v_j). \quad (13.2.14)$$

Последовательность двоичных символов ПШ предполагается некоррелированной. Следовательно,

$$E\left[(2b_j - 1)(2b_i - 1)\right] = \delta_{ij} \quad (13.2.15)$$

и

$$\sigma_m^2 = 4w_m E(v^2), \quad (13.2.16)$$

где $E(v^2)$ – второй момент любого элемента $\{v_j\}$. Этот момент легко вычислить:

$$E(v^2) = \int_0^{T_c}\int_0^{T_c} g^*(t)g(\tau)\phi_{zz}(t-\tau)\,dt\,d\tau = \int_{-\infty}^{\infty}|G(f)|^2\Phi_{zz}(f)df, \quad (13.2.17)$$

где $\phi_{zz}(\tau) = \frac{1}{2}E\left[z^*(t)z(t+\tau)\right]$ – автокорреляционная функция, а $\Phi_{zz}(f)$ – спектральная плотность мощности интерференции $z(t)$.

Мы видим, что если интерференция имеет одинаковую спектральную плотность внутри полосы частот[2] переданного сигнала, т.е.

$$\Phi_{zz}(f) = J_0, \qquad |f| \le \tfrac{1}{2}W, \quad (13.2.18)$$

тогда второй момент в (13.2.17) $E(v^2) = 2\mathscr{E}_c J_0$ и, следовательно, дисперсия интерференционного слагаемого в (13.2.16) равна

$$\sigma_m^2 = 8\mathscr{E}_c J_0 w_m. \quad (13.2.19)$$

В этом случае вероятность того, что D<0 равна

---

[1] Типичный показатель расширения спектра для широкополосных сигналов имеет порядок 100 и больше.

[2] Если полоса полосового канала W, то полоса эквивалентного низкочастотного канала $\frac{1}{2}W$.

$$P_2(m) = Q\left(\sqrt{\frac{2\mathscr{E}_c}{J_0} w_m}\right). \qquad (13.2.20)$$

Но энергия на кодовый символ $\mathscr{E}_c$ можно выразить через энергию информационного символа $\mathscr{E}_b$

$$\mathscr{E}_c = \frac{k}{n}\mathscr{E}_b = R_c \mathscr{E}_b. \qquad (13.2.21)$$

Представив это в (13.2.20), получаем

$$P_2(m) = \left(\sqrt{\frac{2\mathscr{E}_b}{J_0} R_c w_m}\right) = Q\left(\sqrt{2\gamma_b R_c w_m}\right), \qquad (13.2.22)$$

где $\gamma_b = \mathscr{E}_b / J_0$ – ОСШ на информационный бит. Наконец, вероятность ошибки кодового слова можно оценить сверху границей

$$P_m \le \frac{1}{k}\sum_{m=2}^{M} Q\left(\sqrt{2\gamma_b R_c w_m}\right), \qquad (13.2.23)$$

где $M=2^k$. Заметим, что это выражение идентично вероятности ошибки кодового слова при декодировании мягких решений линейного двоичного блокового кода в канале с АБГШ.

Хотя выше мы рассмотрели двоичный блоковый код описанная процедура аналогична и для (n,k) сверточного кода.

Результат такого рассмотрения дает следующую верхнюю границу для эквивалентной вероятности ошибки на бит

$$P_b \le \frac{1}{k}\sum_{d=d_{св}}^{\infty} \beta_d Q\left(\sqrt{2\gamma_b R_c d}\right). \qquad (13.2.24)$$

Набор коэффициентов $\{\beta_d\}$ получается из расширения производных передаточной функции $T(D,N)$, как описано в разделе (8.2.3).

Далее мы рассмотрим узкополосную интерференцию, концентрированную около несущей (около нуля для эквивалентного низкочастотного сигнала). Мы можем фиксировать суммарную (среднюю) мощность помехи $J_{ср} = J_0 W$, где J_0 – величина спектральной плотности мощности эквивалентной широкополосной интерференции (сигнал глушения). Узкополосная интерференция характеризуется спектральной плотностью мощности

$$\Phi_{zz}(f) = \begin{cases} \dfrac{J_{ср}}{W_1} = \dfrac{J_0 W}{W_1}, & \left(|f| \le \tfrac{1}{2} W_1\right) \\ 0, & \left(|f| > \tfrac{1}{2} W_1\right). \end{cases}, \qquad (13.2.25)$$

где $W >> W_1$

Подстановка (13.2.25) для $\Phi_{zz}(f)$ в (13.2.17) дает

$$E(v^2) = \frac{J_{ср}}{W_1}\int_{-W_1/2}^{W_1/2} |G(f)|^2 df. \qquad (13.2.26)$$

Величина $E(v^2)$ зависит от спектральных характеристик $g(t)$. В следующем примере мы рассмотрим два специальных случая.

Пример 13.2.1. Предположим, что $g(t)$ – прямоугольный импульс, показанный на рис.13.2.3(a), а $|G(f)|^2$ – соответствующая спектральная плотность энергии, показанная на рис.13.2.3(b).

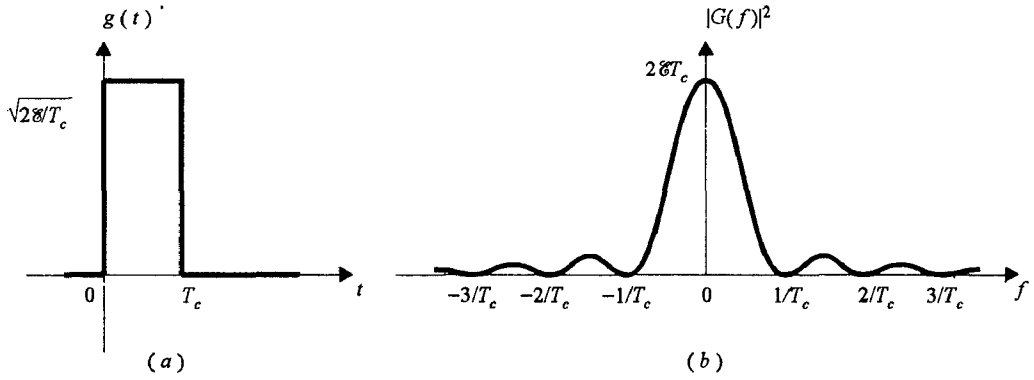


Рис. 13.2.3. Прямоугольный импульс и его энергетический спектр

Для узкополосной интерференции, определяемой (13.2.26), дисперсия общей интерференции равна

$$\sigma_m^2 4w_m E(v^2) = \frac{8\mathscr{E}_c w_m T_c J_{\text{ср}}}{W_1} \int_{-W_1/2}^{W_1/2} \left(\frac{\sin \pi f T_c}{\pi f T_c} \right) df = \frac{8\mathscr{E}_c w_m J_{\text{ср}}}{W_1} \int_{-\beta/2}^{\beta/2} \left(\frac{\sin \pi x}{\pi x} \right)^2 dx, \qquad (13.2.27)$$

где $\beta = W_1 T_c$. Рисунок 13.2.4 иллюстрирует величину этого интеграла для $0 \le \beta \le 1$.

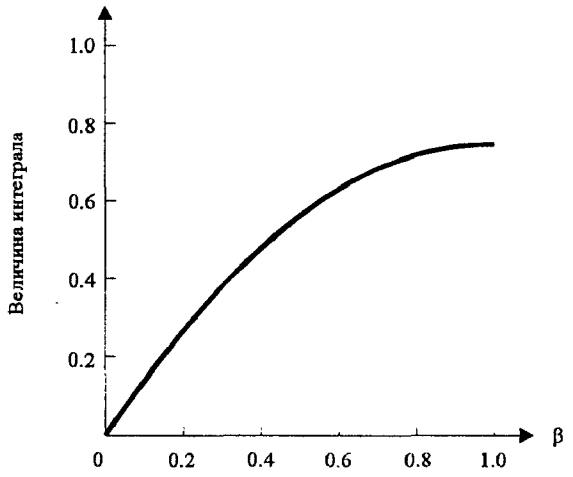


Рис. 13.2.4. График зависимости величины интеграла в (13.2.7) от интервала интегрирования β

Мы видим, что величина интеграла имеет верхнюю границу, $W_1 T_c$. Следовательно, $\sigma_m^2 \le 8\mathscr{E}_c w_m T_c J_{\text{ср}}$.

В пределе, когда W_1 становится нулём, интерференция определяется импульсом на несущей. В этом случае интерференция представляет чисто гармонический сигнал и она обычно называется гармонической помехой на несущей (ГПН, CW jamming signal). Спектральная плотность мощности равна

$$\Phi_{zz}(f) = J_{\text{ср}} \delta(f), \qquad (13.2.28)$$

и соответствующая дисперсия для величины $D = CM_1 - CM_m$ равна

$$\sigma_m^2 = 4w_m J_{\text{ср}} |G(0)|^2 = 8\mathscr{E}_c w_m T_c J_{\text{ср}}. \qquad (13.2.29)$$

Вероятность ошибки кодового слова для ГПН имеет верхнюю границу

$$P_M \le \sum_{m=2}^{M} Q\left(\sqrt{\frac{2\mathscr{E}_c}{J_{\text{ср}} T_c} w_m}\right). \qquad (13.2.30)$$

Но $\mathscr{E}_c = R_c \mathscr{E}_b$ Далее $T_c \approx 1/W$ и $J_{\text{ср}}/W = J_0$. Следовательно, (13.2.30) можно выразить так:

$$P_M \le \sum_{m=2}^{M} Q\left(\sqrt{\frac{2\mathscr{E}_b}{J_0} R_c w_m}\right), \qquad (13.2.31)$$

что является результатом, полученным раньше для широкополосной интерференции. Этот результат говорит о том, что ГПН имеет то же влияние на качество, что и эквивалентный широкополосный мешающий сигнал. Эта эквивалентность обсуждается ниже.

Пример 13.2.2. Определим качество широкополосной системы с ПП в присутствии CW jammer со средней мощностью $J_{\text{ср}}$, когда импульс переданного сигнала $g(t)$ определяется полупериодом синусоиды, как показано на рис.13.2.5, т.е.

$$g(t) = \sqrt{\frac{4\mathscr{E}_c}{T_c}} \sin\frac{\pi t}{T_c}, \qquad 0 \le t \le T_c. \qquad (13.2.32)$$

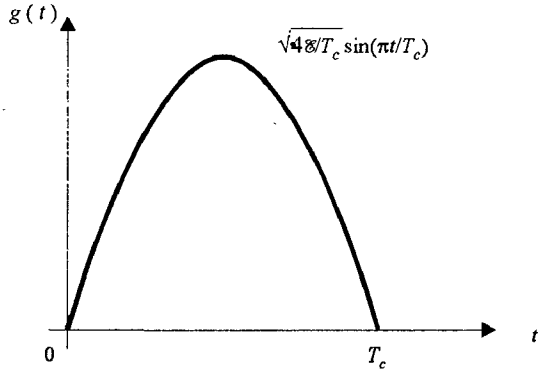


Рис. 13.2.5. Синусоидальный сигнальный импульс

Дисперсия интерференции для такого символа

$$\sigma_m^2 = 4w_m J_{\text{ср}} |G(0)|^2 = \frac{64}{\pi^2} \mathscr{E}_c T_c J_{\text{ср}} w_m. \qquad (13.2.33)$$

Таким образом, верхняя граница вероятности ошибки кодового слова

$$P_M \le \sum_{m=2}^{M} Q\left(\sqrt{\frac{\pi^2 \mathscr{E}_b}{4J_{\text{ср}} T_c} R_c w_m}\right). \qquad (13.2.34)$$

Мы видим, что качество, получаемое таким импульсом на 0,9 дБ лучше, чем то, которое получено при прямоугольном импульсе. Напомним, что такая огибающая, используемая в офсетной КФМ, ведёт к сигналу ММС. ММС модуляция часто используется в широкополосных системах с ПП.

Выигрыш обработки и помехозащищённость (jamming margin). Интересная интерпретация характеристики качества широкополосного сигнала с ПП можно получить, выражая энергию сигнала на бит $\mathscr{E}_b$ через среднюю мощность. Это значит, что $\mathscr{E}_b = P_{\text{ср}} T_b$,

где $P_{ср}$ – средняя мощность сигнала и T_b – символьный интервал. Рассмотрим качество, полученное в присутствии CW-jamming для прямоугольного импульса, обсуждённого в примере 13.2.1. Если подставить значения для $\mathscr{E}_b$ и J_0 в (13.2.31), мы получим

$$P_M \le \sum_{m=2}^{M} Q\left(\sqrt{\frac{2P_{ср}T_b}{J_{ср}T_c}R_c w_m}\right) = \sum_{m=2}^{M} Q\left(\sqrt{\frac{2P_{ср}}{J_{ср}}L_c R_c w_m}\right), \qquad (13.2.35)$$

где L_c – число чипов на информационный символ, а $P_{ср}/J_{ср}$ – отношение мощности сигнала к мощности помехи.

Аналогичный результат получен для широкополосного мешающего сигнала, для которого качество даётся (13.2.23). Для энергии сигнала на бит имеем

$$\mathscr{E}_b = P_{ср}T_b = \frac{P_{ср}}{R}, \qquad (13.2.36)$$

где R – информационная скорость в бит/с. Спектральная плотность мощности для мешающего сигнала можно выразить так:

$$J_0 = J_{ср}/W. \qquad (13.2.37)$$

Используя отношения (13.2.36) и (13.2.37), отношение $\mathscr{E}_b / J_0$ можно выразить так

$$\frac{\mathscr{E}_b}{J_0} = \frac{P_{ср}/R}{J_{ср}/W} = \frac{W/R}{J_{ср}/P_{ср}}. \qquad (13.2.38)$$

Отношение $J_{ср}/P_{ср}$ – это отношение средней мощности помехи к средней мощности сигнала, которое обычно больше единицы. Отношение $W/R = T_b/T_c = B_c = L_c$ как раз показатель расширения полосы частот или, что эквивалентно, число чипов на информационный бит. Это отношение обычно называется *выигрышем обработки* (ВО) *широкополосной системы* с ПП. Оно представляет преимущество, выигранное относительно помехи, которое получается благодаря расширению полосы частот передаваемого сигнала. Если будем интерпретировать $\mathscr{E}_b / J_0$ как ОСШ, требуемое для достижения заданной вероятности ошибки, а *W/R* как допустимый показатель расширения полосы частот, соотношение $J_{ср}/P_{ср}$ будет иметь смысл помехозащищённости (запаса по помехе) широкополосной системы с ПП. Другими словами, помехозащищённость – это наибольшая величина, которую может принять отношение $J_{ср}/P_{ср}$, при котором система передачи ещё удовлетворяет заданной вероятности ошибки.

Качество декодера мягких решений для линейного двоичного кода (n,k), выраженное через выигрыш обработки и помехозащищённость, определяется так:

$$P_m \le \sum_{m=2}^{M} Q\left(\sqrt{\frac{2W/R}{J_{ср}/P_{ср}}R_c w_m}\right) \le (M-1)Q\left(\sqrt{\frac{2W/R}{J_{ср}/P_{ср}}R_c d_{\min}}\right). \qquad (13.2.39)$$

В дополнение к зависимости от выигрыша обработки W/R и $J_{ср}/P_{ср}$ мы видим, что качество зависит от третьего множителя, именно $R_c w_m$. Этот множитель определяет выигрыш кода. Нижняя граница этого множителя равна $R_c d_{\min}$. Таким образом, помехозащищённость, достигаемая широкополосными сигналами с ПП, зависит от выигрыша обработки и выигрыша кодирования.

Некодированные широкополосные сигналы с ПП. Результаты качества, данные выше для широкополосных сигналов с ПП, генерируемые посредством (n,k) кода, могут быть конкретизированы для тривиального кода, именно для двоичного кода с

повторением. В этом случае $k=1$, а вес ненулевого кодового слова $w=n$. Таким образом, $R_c w=1$, следовательно, качество двоичной системы сигналов определяется так

$$P_2=Q\left(\sqrt{\frac{2\mathscr{E}_b}{J_0}}\right)=Q\left(\sqrt{\frac{2W/R}{J_{\text{ср}}/P_{\text{ср}}}}\right). \qquad (13.2.40)$$

Заметим, что тривиальный код не дает выигрыш кодирования. Он даёт выигрыш обработки W/R.

Пример 13.2.3. Предположим, что мы желаем достичь вероятность ошибки 10^{-6} или меньше при помощи широкополосной системы с ПП. Желательный показатель расширения полосы $W/R=1000$. Определим помехозащищённость.

Требуемая величина $\mathscr{E}_b/J_0$ для достижения вероятности ошибки на бит 10^{-6} при помощи некодированной двоичной ФМ равна 10,5 дБ. Выигрыш обработки равен $10\lg 1000=30$ дБ. Следовательно, максимальное допустимое значение отношения мощностей помехи и сигнала, т.е. помехозащищённость, равно

$$10\lg\frac{J_{\text{ср}}}{P_{\text{ср}}}=30-10{,}5=19{,}5\text{ дБ}.$$

Поскольку эта помехозащищённость достигается для некодированной широкополосной системы с ПП, её можно увеличить путём кодирования информационной последовательности.

Имеется другой путь для рассмотрения процессов модуляции и демодуляции для некодированной (код с повторением) широкополосной системы с ПП. У модулятора сигнал, генерируемый кодом с повторением, например при прямоугольном импульсе, идентичен прямоугольному импульсу $s(t)$ с единичной амплитудой длительностью T_b или его обратному значению, в зависимости от того является ли информационный символ соответственно 1 или 0. Это видно из (13.2.7), где кодовые чипы $\{c_i\}$ внутри информационного символа равны 1 или 0. ПШ последовательность умножается на $s(t)$ или $-s(t)$. Так, если информационный символ 1, то L_c чипов, генерируемых ПШ генератором передаются с той же полярностью. С другой стороны, если информационный символ 0, то L чипов при умножении на $-s(t)$ меняют полярность.

Демодулятор для кода с повторением, реализованный как коррелятор, иллюстрируется на рис.13.2.6.

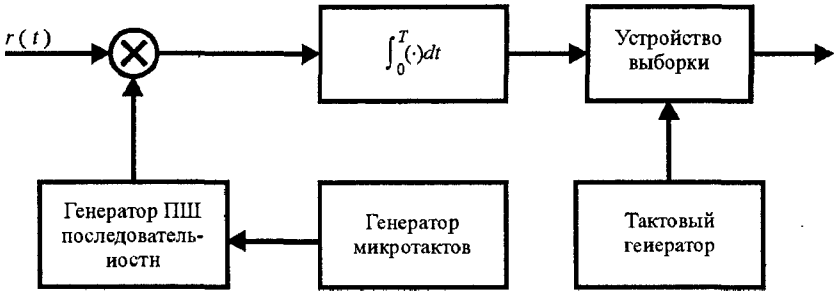


Рис. 13.2.6. Демодулятор корреляционного типа для кода с повторением

Видим, что интервал интегрирования в интеграторе равен символьному интервалу T_b. Таким образом, декодер для кода с повторением ограничен и его функция реализуется демодулятором.

Теперь качественно оценим процесс демодуляции на интерференцию $z(t)$. Умножение $z(t)$ на выход ПШ генератора, который выражается так

$$w(t)=\sum_i (2b_i-1)p(t-iT_c),$$

даёт

$$\upsilon(t)=w(t)z(t).$$

Сигналы $w(t)$ и $z(t)$ – статистически независимые случайные процессы, каждый с нулевым средним и автокорреляционными функциями $\phi_{ww}(\tau)$ и $\phi_{zz}(\tau)$ соответственно. Произведение $\upsilon(\tau)$ также случайный процесс, имеющий автокорреляционную функцию равную произведению $\phi_{ww}(\tau)$ и $\phi_{zz}(\tau)$. Таким образом, спектральная плотность мощности процесса $\upsilon(\tau)$ равно свертке спектральной плотности мощности процесса $w(t)$ и спектральной плотности мощности процесса $z(t)$.

Эффект свертки двух спектров сводится к рассеянию мощности по полосе. Поскольку полоса $w(t)$ занимает возможную полосу частот канала W, то результат свертки двух спектров сводится к рассеянию спектральной плотности мощности процесса $z(t)$ по полосе частот W. Если $z(t)$ – узкополосный процесс, т.е. его спектральная плотность мощности имеет полосу намного меньше W, спектральная плотность мощности процесса $\upsilon(t)$ будет охватывать полосу частот равную, по крайней мере, W.

Интегратор, использованный для взаимной корреляции и показанный на рис.13.2.6, имеет полосу частот примерно равную $1/T_b$. Поскольку $1/T_b << W$, только часть от общей мощности интерференции появится на выходе коррелятора. Эта часть примерно равна отношению полосы $1/T_b$ к W. То есть,

$$\frac{1/T_b}{W}=\frac{1}{WT_b}=\frac{T_c}{T_b}=\frac{1}{L_c}.$$

Другими словами, умножение сигнала интерференции на сигнал от ПШ генератора рассеивает интерференцию до полосы частот сигнала W, а узкополосное интегрирование, следующее за умножением, выделяет только $1/L_c$ часть от общей интерференции. Таким образом, качество некодированной широкополосной системы с ПП увеличивается на величину выигрыша обработки L_c.

Каскадное объединение произвольного линейного кода с двоичным кодом с повторением. Как показано выше, двоичный код с повторением увеличивает помехозащищённость по отношению к мешающему сигналу, но не даёт выигрыша кодирования. Чтобы получить улучшение в качестве, мы можем использовать линейный (n_1,k) блоковый или свёрточный код, где $n_1 \le n = kL_c$. Одна возможность заключается в выборе $n_1 < n$ и к повторению каждого кодового символа n_2 раз так, что $n=n_1n_2$. Так мы можем конструировать линейный (n,k) код путем каскадного объединения кода (n_1,k) с двоичным кодом $(n_2,1)$ с повторением. Это можно рассматривать как тривиальную форму каскадного кода, где внешний код – это (n_1,k), а внутренний код – это код с повторением.

Поскольку код с повторением не даёт выигрыша кодирования, выигрыш кодирования, достигаемый объединением кодов, должен уменьшиться до величины, достигаемой внешним (n_1,k) кодом. Покажем, что это на самом деле так. Выигрыш кодирования объединенного кода равен

$$R_c w_m = \frac{k}{n} w_m, \qquad m = 2, 3, \ldots, 2^k .$$

но веса $\{w_m$ объединенного кода можно выразить так:

$$w_m = n_2 w_m^0,$$

где $\{w_m^0\}$ – веса внешнего кода. Следовательно, выигрыш кодирования объединенного кода

$$R_c w_m = \frac{k}{n_1 n_2} n_2 w_m^0 = \frac{k}{n_1} w_m^0 = R_c^0 w_m^0, \qquad (13.2.41)$$

что как раз равно выигрышу кодирования, получаемого от внешнего кода.

Выигрыш кодирования также достигается, если внешний код (n_1, k) декодирует жёсткие решения. Вероятность ошибки на бит, получаемый с $(n_2, 1)$ кодом с повторением (при декодировании мягких решений) равна

$$P = Q\left(\sqrt{\frac{2n_2 \mathscr{E}_c}{J_0}}\right) = Q\left(\sqrt{\frac{2\mathscr{E}_b}{J_0} R_c^0}\right) = Q\left(\sqrt{\frac{2W/R}{J_{\text{ср}}/P_{\text{ср}}} R_c^0}\right). \qquad (13.2.42)$$

Тогда вероятность ошибки кодового слова для линейного (n_1, k) блокового кода имеет верхнюю границу

$$P_M \le \sum_{m=t+1}^{n_1} \binom{n_1}{m} p^m (1-p)^{n_1 - m}, \qquad (13.2.43)$$

где $t = \left\lfloor \frac{1}{2}(d_{\min} - 1) \right\rfloor$ или

$$P_M \le \sum_{m=2}^{M} [4p(1-p)]^{w_m^0/2}, \qquad (13.2.44)$$

где последнее отношение определяется границей Чернова. Для (n_1, k) двоичного сверточного кода верхняя граница для вероятности ошибки на бит равна

$$P_b \le \sum_{d=d_{\text{св}}}^{\infty} \beta_d P_2(d), \qquad (13.2.45)$$

где $P_2(d)$ определяется (8-2-28) для нечётных d и (8-2-29) для четных d.

Каскадное кодирование для широкополосных систем с ПП. Из приведённого выше обсуждения очевидно, что можно достичь улучшения в качестве путем замены кода с повторением более мощным кодом, который даст выигрыш кодирования в дополнение к выигрышу обработки. В принципе, цель широкополосной системы с ПП – создать длинный, низкоскоростной код, имеющий большое минимальное расстояние. Это можно выполнить наилучшим образом, используя каскадное кодирование. Если двоичная ФМ используется в соединении с широкополосной системой с ПП, элементы каскадированных кодовых слов можно выразить в двоичной форме.

Лучшее качество можно получить, когда декодирование мягких решений используется для внутренних и внешних кодов. Однако альтернативно, что обычно ведет к уменьшению сложности для декодера, используется декодирование мягких решений для внутреннего кода и декодирование жестких решений для внешнего кода. Выражение для вероятности ошибки этих схем декодирования зависит частично от типа кодов (блоковых или сверточных), выбираемых для внутреннего и внешнего кодов. Для примера, кодирование двух блоковых кодов можно рассматривать как общий длинный двоичный (n, k) блоковый код, имеющий качество, даваемое (13.2.39). Качество других каскадных кодов также можно в принципе проанализировать. Из соображений сложности мы не будем рассматривать такое кодовое каскадирование.

13.2.2. Некоторые приложения широкополосных сигналов с ПП

В этом подразделе мы хотим рассмотреть использование кодированных широкополосных сигналов с ПП для трех специальных применений.

Одно связано с обеспечением невосприимчивости к jamming сигналу. Во втором сигнал связи скрывается в основном шуме путем передачи сигнала на очень низком уровне мощности. Третье приложение связано с обеспечением передачи определенного числа сигналов в общей полосе частот, т.е. с CDMA.

Антипомеховое (АП) приложение. В разделе 13.2.1 мы получили вероятность ошибки для широкополосных систем с ПП в присутствии узкополосного или широкополосного мешающего сигнала. В качестве примеров, иллюстрирующих качество цифровых систем связи при наличии jamming сигналов, мы выберем три кода. Один – это код Голея (24, 12), который характеризуется распределением весов, данных в таблице 8.1.1 и имеет минимальное расстояние $d_{\min} = 8$. Второй код является укороченным кодом Голея (24, 11), полученный путем выбора 2048 кодовых слов постоянного веса 12. Конечно, этот укороченный код нелинеен. Эти два кода будут использованы в соединении с кодом с повторением. Третий код, который будет рассматриваться, это код максимальной длины регистра сдвига.

Вероятность ошибки кода Голея (24, 12) при декодировании мягких решений

$$P_M \le \left[759Q\left(\sqrt{\frac{8W/R}{J_{ср}/P_{ср}}}\right) + 2576Q\left(\sqrt{\frac{12W/R}{J_{ср}/P_{ср}}}\right) + 759Q\left(\sqrt{\frac{16W/R}{J_{ср}/P_{ср}}}\right) + Q\left(\sqrt{\frac{24W/R}{J_{ср}/P_{ср}}}\right)\right], \quad (13.2.46)$$

где W/R – выигрыш обработки, а $J_{ср}/P_{ср}$ – помехозащищённость. Поскольку $n = n_1 n_2 = 12W/R$ и $n_1 = 24$, каждый кодовый бит повторяется фактически $n_2 = W/2R$ раз. Для примера, если $W/R = 100$ (выигрыш обработки равен 20 дБ) длина блока кода с повторением $n_2 = 50$.

Если используется декодирование жестких решений, вероятность ошибки для кодового символа равна

$$P = Q\left(\sqrt{\frac{W/R}{J_{ср}/P_{ср}}}\right), \quad (13.2.47)$$

а соответствующая вероятность ошибки кодового слова имеет верхнюю границу

$$P_M \le \sum_{m=4}^{24} \binom{24}{m} p^m (1-p)^{24-m}. \quad (13.2.48)$$

В качестве альтернативы мы можем использовать границу Чернова при декодировании жестких решений, которая дает

$$P_M \le 759[4p(1-p)]^4 + 2576[4p(1-p)]^6 + 759[4p(1-p)]^8 + [4p(1-p)]^{12}. \quad (13.2.49)$$

Рис.13.2.7 иллюстрирует качество кода Голея (24, 12), как функцию от запаса глушения $J_{ср}/P_{ср}$ и с выигрышем обработки как параметра. Для подсчета вероятности ошибки при декодировании жестких решений была использована граница Чернова.

Вероятность ошибки при декодировании мягких решений в основном определяется слагаемым

$$759Q\left(\sqrt{\frac{8W/R}{J_{ср}/P_{ср}}}\right),$$

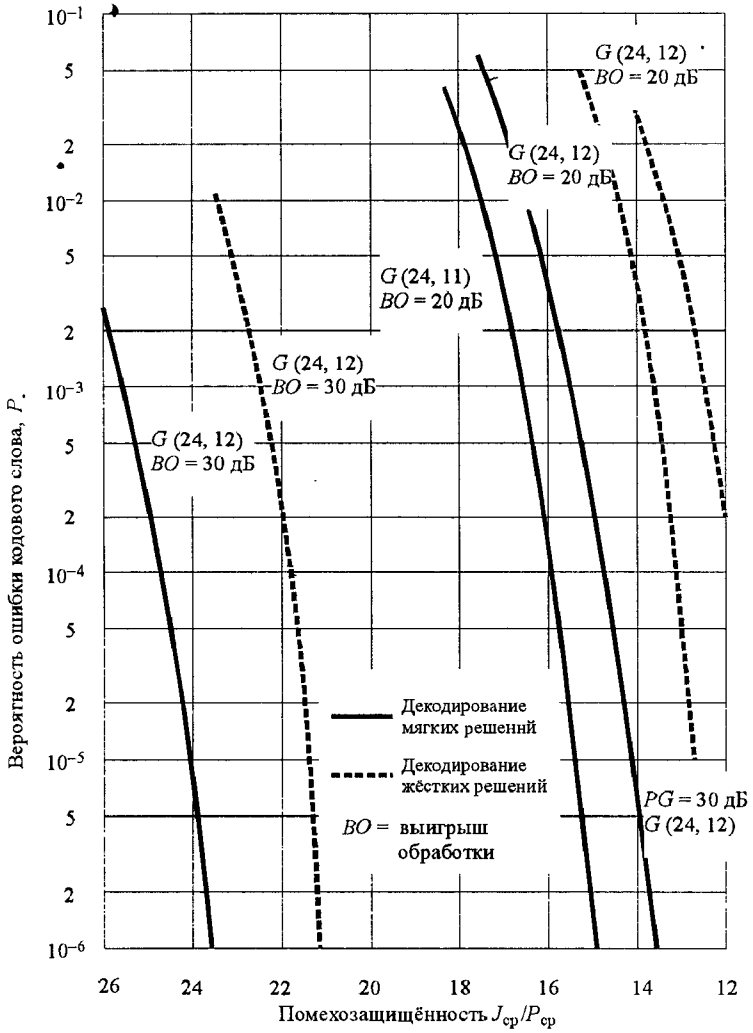


Рис. 13.2.7. Характеристики кода Голея, использующего ПШ широкополосный сигнал

а при декодировании жёстких решений в основном определяется слагаемым $759[4p(1-p)]^4$. Таким образом, выигрыш кодирования при декодировании мягких решений[1] самое большее $10\lg 4 = 6$ дБ. Заметим, что две кривые, соответствующие W/R= 1000 (30 дБ) и $W/R = 100$ (20 дБ) идентичны по огибающей, исключая того, что последняя сдвинута на 10 дБ вправо по отношению к первой. Этот сдвиг это просто разница в выигрыше обработки между этими двумя широкополосными сигналами с ПШ.

Вероятность ошибки для укороченного кода Голея (24, 11) ограничена сверху

$$P_M \le 2047\,Q\left(\sqrt{\frac{11W/R}{J_{ср}/P_{ср}}}\right) \qquad (13.2.50)$$

для декодирования мягких решений[2] и

$$P_M \le 2047[4p(1-p)]^6 \qquad (13.2.51)$$

при декодировании жёстких решений, где p определяется так:

$$p = Q\left(\sqrt{\frac{11W/R}{J_{ср}/P_{ср}}}\right). \qquad (13.2.52)$$

Характеристики качества этого кода также даны на рис.13.2.7 для $W/R = 100$. Видим, что укороченный код Голея (24, 11) примерно на 1 дБ лучше, чем код Голея (24, 12).

Вместо использования блокового кода каскадно с низкоскоростным $(1/n_2)$ кодом с повторением, рассмотрим использование простого низкоскоростного кода. Подходящий набор низкоскоростных кодов – это набор кодов регистра сдвига максимальной длины, описанного в разделе 8.1.3.

Напомним, что для этого набора кодов

$$(n, k) = (2^m - 1,\ m),$$

[1] Выигрыш кодирования меньше 6 дБ из-за множителя 759, который увеличивает вероятность ошибки по сравнению с качеством двоичной некодированной системой.

[2] Мы напомним читателю, что объединённая граница не очень плотная для таких сигнальных ансамблей.

$$d_{\min} = 2^{m-1}. \qquad (13.2.53)$$

Все кодовые слова, исключая слово из одних нулей, имеет одинаковый вес 2^{m-1}. Следовательно, вероятность ошибочного декодирования мягких решений ограничена сверху[1]. Так,

$$P_M \le (M-1)Q\left(\sqrt{\frac{2W/R}{J_{\text{ср}}/P_{\text{ср}}} R_c d_{\min}}\right) \le 2^m Q\left(\sqrt{\frac{2W/R}{J_{\text{ср}}/P_{\text{ср}}} \frac{m2^{m-1}}{2^m-1}}\right)$$
$$\le 2^m \exp\left(-\frac{W/R}{J_{\text{ср}}/P_{\text{ср}}} \frac{m2^{m-1}}{2^m-1}\right). \qquad (13.2.54)$$

Для умеренных значений m, $R_c d_{\min} \approx \frac{1}{2}m$ и, следовательно, (13.2.54) можно выразить так:

$$P_m \le 2^m Q\left(\sqrt{\frac{W/R}{J_{\text{ср}}/P_{\text{ср}}} m}\right) \le 2^m \exp\left(-\frac{mW/R}{2J_{\text{ср}}/P_{\text{ср}}}\right). \qquad (13.2.55)$$

Таким образом, выигрыш кодирования не более $10\lg\frac{1}{2}m$.

Для примера, если выберем $m = 10$, тогда $n = 2^{10} - 1 = 1023$. Поскольку $n=kW/R=mW/R$, то следует, что $W/R \approx 102$. Таким образом мы имеем выигрыш обработки около 20 дБ и выигрыш кодирования порядка 7 дБ. Такое качество сравнимо с тем, которое получается укороченным кодом Голея (24, 11). Больший выигрыш кодирования можно достичь при больших значениях m.

Если используется декодирование жестких решений для кода максимальной длины, то вероятность ошибки регистра сдвига ограничена сверху границей Чернова так

$$P_M \le (M-1)[4p(1-p)]^{d_{\min}/2} = (2^m-1)[4p(1-p)]^{2^{m-2}}, \qquad (13.2.56)$$

где

$$p = Q\left(\sqrt{\frac{2W/R}{J_{\text{ср}}/P_{\text{ср}}} R_c}\right) = Q\left(\sqrt{\frac{2W/R}{J_{\text{ср}}/P_{\text{ср}}} \frac{m}{2^m-1}}\right). \qquad (13.2.57)$$

Для $m = 10$ вероятность ошибки кодового слова P_m сравнима с той, которая получается с укороченным кодом Голея (24, 11) при декодировании жёстких решений.

Данные выше результаты иллюстрируют качество, которое можно получить при обычном кодировании. Больший выигрыш кодирования можно достичь каскадными кодами.

Скрытная передача сигналов. В этом приложении сигнал целевым образом передается с очень низким уровнем мощности относительно основного шума канала и теплового шума, который создается на приемной стороне. Если широкополосные сигналы с ПП занимают полосу частот W, а спектральная плотность аддитивного шума равна N_0, то средняя мощность шума в полосе W равна $N_{\text{ср}} = WN_0$.

Средняя мощность принимаемых сигналов на приемной стороне обозначим $P_{\text{ср}}$. Если мы хотим скрыть присутствие сигнала от приемников, которые находятся вблизи заданного приемника, сигнал следует передавать на низком уровне мощности $P_{\text{ср}}/N_{\text{ср}} << 1$. Заданный приемник может восстановить информацию в сигнале при помощи выигрыша обработки и выигрыша кодирования. Любой другой приёмник, который не имеет

[1] $M=2^m$ сигналов, генерируемых регистром сдвига максимальной длины, образует симплексный ансамбль сигналов (смотри задачу 8.13). Точное выражение для вероятности ошибки, данное в разделе 5.2.4, можно использовать для больших значений M, когда объединенная граница получается очень свободной.

априорной информации о ПШ последовательности, не в состоянии извлечь преимущества от выигрыша обработки и выигрыша кодирования. Следовательно, присутствие информационного сигнала трудно обнаружить. Мы говорим, что сигнал имеет низкую вероятность быть перехваченным (НВП) и он называется НВП-сигналом.

Результаты вероятности ошибки, данные в разделе 13.2.7, также применимы для демодуляции и декодирования сигналов НВП на заданном приемнике.

Кодовое разделение при множественном доступе. Увеличение качества, получаемое от широкополосных сигналов с ПП через выигрыш обработки и выигрыш кодирования, можно использовать для размещения многих широкополосных ПП сигналов в той же полосе канала при условии, что каждый сигнал имеет свою собственную отличную ПШ последовательность. Таким образом, имеется возможность для нескольких пользователей передавать сообщения одновременно по той же полосе частот. Этот вид цифровой связи, в которой каждый пользователь (пара приемник-передатчик) имеет различный ПШ код для передачи в общей полосе канала называют кодовым разделением со множественным доступом (CDMA) или многостанционным доступом с рассеянным спектром (SSMA).

При демодуляции каждого ПШ сигнала, сигналы остальных одновременных пользователей канала выглядят, как аддитивная интерференция. Уровень интерференции меняется и зависит от числа пользователей в заданное время. Главное преимущество CDMA в том, что может размещаться большое число пользователей в системе, если каждый передает сообщения в коротком интервале времени. В такой системе со множественным доступом относительно просто или прибавить новых пользователей или уменьшить число пользователей без разрушения системы.

Определим число одновременных сигналов, которые можно разместить в CDMA-системе[1]. Для простоты предположим, что все сигналы имеют одинаковые средние мощности. Так, если имеется $N_п$ одновременных пользователей, отношение средних мощностей сигнала и шума интерференции для данного приёмника равно

$$\frac{P_{ср}}{J_{ср}} = \frac{P_{ср}}{(N_п - 1)P_{ср}} = \frac{1}{N_п - 1}. \qquad (13.2.58)$$

Таким образом, вероятность ошибки при декодировании жёстких решений для данного приёмника ограничена сверху так:

$$P_M \le \sum_{m=2}^{M} Q\left(\sqrt{\frac{2W/R}{N_п - 1} R_c w_m}\right) \le (M-1)Q\left(\sqrt{\frac{2W/R}{N_п - 1} R_c d_{\min}}\right). \qquad (13.2.59)$$

Здесь мы предположили, что интерференция от других пользователей – гауссовский случайный процесс, а флуктуационным шумом мы пренебрегли.

В этом разделе интерференция от других пользователей трактуется как случайный процесс. Это так, если нет корреляции между пользователями. В главе 15 мы рассмотрим CDMA передачу, в которой интерференция от других пользователей известна и она подавляется на приеме.

В качестве примера предположим, что требуемое качество передачи (вероятность ошибки 10^{-6}) достигается при

$$\frac{W/R}{N_u - 1} R_c d_{\min} = 20.$$

[1] В этом разделе помехи от других пользователей рассматриваются как случайные процессы. Это случай, когда между пользователями нет взаимодействия. В гл.15 мы рассматриваем CDMA, при которой интерференция от других пользователей известна и подавляется в приёмнике.

Тогда максимальное число пользователей, которое может содержать CDMA система, равно

$$N_{\text{п}} = \frac{W/R}{20} R_c d_{\min} + 1. \qquad (13.2.60)$$

Если $W/R = 100$ и $R_c d_{\min} = 4$, как это было получено кодом Голея (24, 12), то максимальное число $N_{\text{п}} = 21$. Если $W/R = 1000$ и $W/R = 1000$, это число получится равным $N_{\text{п}} = 201$.

При определении максимального числа одновременных пользователей канала мы безоговорочно предположили, что ПШ кодовые последовательности взаимно ортогональны и интерференция от других пользователей суммируется только по мощности. Однако, ортогональность среди определенного числа ПШ кодовых последовательностей не легко достичь, особенно если требуемое число ПШ кодовых последовательностей велико. Действительно, выбор хорошего ансамбля ПШ последовательностей для системы CDMA – важная проблема, которая привлекла значительное внимание в технической литературе. Мы хотим обсудить эту проблему в разделе 13.2.3.

13.2.3. Влияние импульсной интерференции на широкополосные ПП системы

До сих пор мы рассмотрели влияние непрерывной интерференции или непрерывной прицельной помехи на широкополосные ПП сигналы. Мы видели, что выигрыш обработки и выигрыш кодирования обеспечивают средство для преодоления вредных влияний этих видов интерференции. Однако существует мешающий сигнал, который имеет весьма существенное влияние на качество широкополосных ПП систем. Такой мешающий сигнал состоит из импульсов с равномерным частотным спектром, который покрывает полностью полосу частот сигнала W. Их обычно называют импульсной интерференцией или парциально-временным мешающим сигналом (ПВМС).

Предположим, что мешающий сигнал имеет среднюю мощность $J_{\text{ср}}$ в полосе частот сигнала W. Тогда $J_0 = J_{\text{ср}}/W$. Вместо непрерывной передачи источник мешающего сигнала передаёт импульсы повышенной мощности $J_{\text{ср}}/\alpha$ за долю времени α, т.е. вероятность того, что источник мешающего сигнала создаёт в данный момент времени помеху, равна α. Для простоты мы предположим, что импульсы интерференции простираются на целое число сигнальных интервалов и таким образом влияют на целое число символов. Если источник мешающего сигнала не выдаёт помеху, то переданный сигнальный бит предполагается принятым без ошибки, а когда он работает, вероятность ошибки для некодированной широкополосной ПП системы равна $Q\left(\sqrt{2\alpha \mathscr{E}_b / J_0}\right)$. Таким образом, средняя вероятность ошибки на бит равна

$$P_2(\alpha) = \alpha Q\left(\sqrt{2\alpha \mathscr{E}_b / J_0}\right) = \alpha Q\left(\sqrt{\frac{2\alpha W/R}{J_{\text{ср}}/P_{\text{ср}}}}\right). \qquad (13.2.61)$$

Источник мешающего сигнала выбирает параметр α (называемый иногда дежурным циклом – duty cycle), чтобы максимизировать вероятность ошибки. Дифференцируя (13.2.61) по α, находим, что наихудший случай ПВМС имеет место, когда

$$\alpha^* = \begin{cases} \dfrac{0{,}71}{\mathscr{E}_b / J_0}, & (\mathscr{E}_b / J_0 \geq 0{,}71) \\ 1, & (\mathscr{E}_b / J_0 < 0{,}71) \end{cases}, \qquad (13.2.62)$$

а соответствующая вероятность ошибки равна

$$P_2 = \begin{cases} \dfrac{0{,}083}{\mathscr{E}_b / J_0} = \dfrac{0{,}083 J_{ср} / P_{ср}}{W / R}, & (\mathscr{E}_b / J_0 > 0.71) \\ Q\left(\sqrt{\dfrac{2W/R}{J_{ср}/P_{ср}}}\right), & (\mathscr{E}_b / J_0 < 0.71) \end{cases} \qquad (13.2.63)$$

Вероятность ошибки, определяемая (13.2.61) для $\alpha = 1{,}0$, 0,1 и 0,01, а также в наихудшем случае α^* даны на рис.13.2.8. Сравнивая вероятность ошибки при мешающем непрерывном гауссовском шуме (α=1) и наихудшем случае ПВМС, видим большую разницу в качестве, которая примерно равна 40 дБ[1] при вероятности ошибки 10^{-6}.

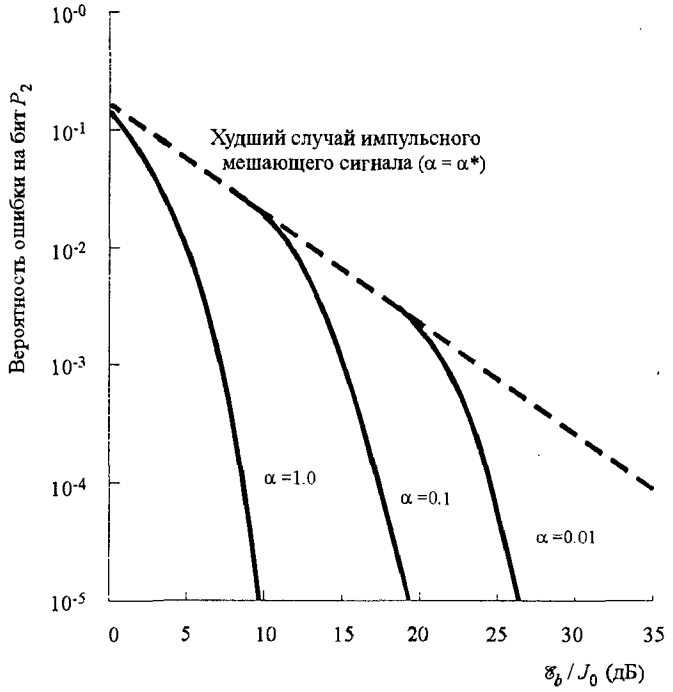


Рис. 13.2.8. Характеристики псевдошумовой двоичной ФМ с импульсным мешающим сигналом

Мы хотим подчеркнуть, что проведенный анализ приложим, когда длительность мешающего импульса равна или больше длительности символа. Дополнительно мы хотим указать, что практические соображения могут запретить источнику мешающего сигнала достигать больших пиковых значений мощности (малых значений α). Всё же, вероятность ошибки, даваемая (13.2.63), служит верхней границей для качества некодированной двоичной ФМ в наихудшем случае ПВМС. Ясно, что качество широкополосной ПП системы при наличии такой помехи наиболее низкое.

Если мы просто прибавим кодирование к широкополосной ПП системе, улучшение относительно некодированной системы равно выигрышу кодирования. Таким образом, требуемое $\mathscr{E}_b / J_0$ уменьшится за счет выигрыша кодирования, который в большинстве случаев ограничен величиной меньшей 10 дБ. Причина плохого качества заключается в том, что длительность импульса мешающего сигнала можно так выбрать, чтобы влиять на многие соседние кодовые символы, когда источник помехи включён. Следовательно,

[1] Это означает, что источник мешающего сигнала при использовании оптимальной стратегии может добиться своей цели, расходуя мощность на 40 дБ меньше (прп).

вероятность ошибки кодового слова велика из-за импульсного характера мешающего сигнала.

Чтобы улучшить качество, мы можем перемежать кодовые символы до их передачи по каналу. Влияние перемежения, как говорилось в разделе 8.1.9, сводится к тому, чтобы сделать кодовые символы, поражённые глушителем, независимыми.

Блок-схема цифровой системы связи, которая включает перемежение-деперемежение, показана на рис.13.2.9. Показана также возможность того, что приёмник знает состояние источника помехи, т.е. знает, включен он или нет.

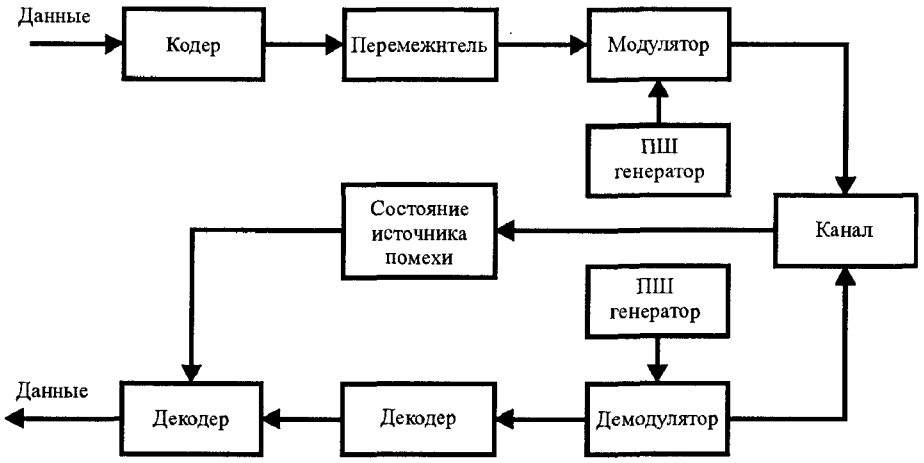


Рис. 13.2.9. Блок-схема АП системы связи

Знание состояния источника помехи (называемое *сторонней информацией*) иногда имеется в распоряжении при измерении уровней шума в канале в соседних частотных полосах. В нашей трактовке мы рассмотрим два экстремальных случая, именно, нет никакой информации о состоянии источника помехи или имеется полное знание о его состоянии. В любом случае, случайная величина ζ, представляющая состояние источника помехи, характеризуется вероятностями

$$P(\zeta = 1) = \alpha, \qquad P(\zeta = 0) = 1 - \alpha.$$

Когда источник помехи включён, канал моделируется как имеющий АБГШ со спектральной плотностью мощности $N_0 = J_0 / \alpha = J_{ср} / \alpha W$, а когда источник помехи выключен – в канале нет шума. Знание состояния источника помехи подразумевает, что декодер знает, когда $\zeta = 1$ и когда $\zeta = 0$, и используют эту информацию при вычислении корреляционных метрик. Для примера, декодер может взвешивать выход демодулятора для каждого кодового символа величиной обратной уровня мощности шума на данном интервале. Альтернативно декодер может дать нулевой вес для пораженного помехой символа.

Сначала рассмотрим влияние помехи без знания состояния источника помехи. Предполагается, что пара перемежитель-деперемежитель приводит к независимым попаданиям помехи на кодовые символы. Как пример качества, достигаемого кодированием, мы приводим результат качества из статьи Мартина и Мак-Адама (1980). Здесь качество двоичных свёрточных кодов вычислена для наихудшего случая ПВМС. Рассматривается декодирование Витерби жестких и мягких решений. Мягкие решения получены квантованием выхода демодулятора на восемь уровней. Для этой цели используется равномерный квантователь для которого пороговые уровни оптимизированы применительно к уровням импульсного шума глушителя. Квантователь играет важную

роль для ограничения размера выходов демодулятора, когда источник помехи включен. Эффект ограничения гарантирует, что любой «удар» на кодовый символ не будет существенно влиять на соответствующие метрики пути.

Оптимальный «дежурный цикл» для ПВМС в кодированной системе обычно обратно пропорционален ОСШ, но его величина отличается от той, которая определяется (13.2.62) для некодированной системы. Рис.13.2.10 иллюстрирует графически оптимальный дежурный цикл глушителя при декодировании жёстких и мягких решений для сверточного кода со скоростью 1/2. Соответствующие результаты для вероятности ошибки для этого наихудшего случая ПВМС иллюстрируются рис.13.2.11 и 13.2.12 для скорости кода 1/2 и кодового ограничения $3 \le K \le 9$, соответственно при декодировании жёстких и мягких решений АВ.

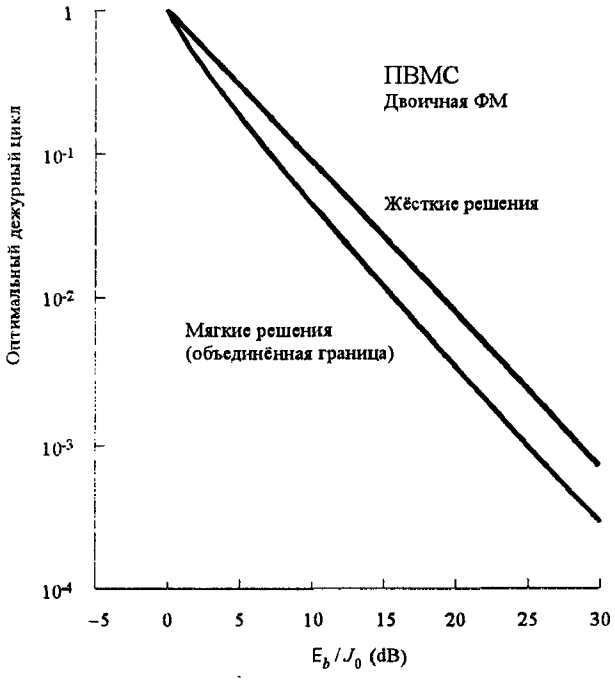


Рис. 13.2.10. Оптимальный дежурный цикл для ПВМС

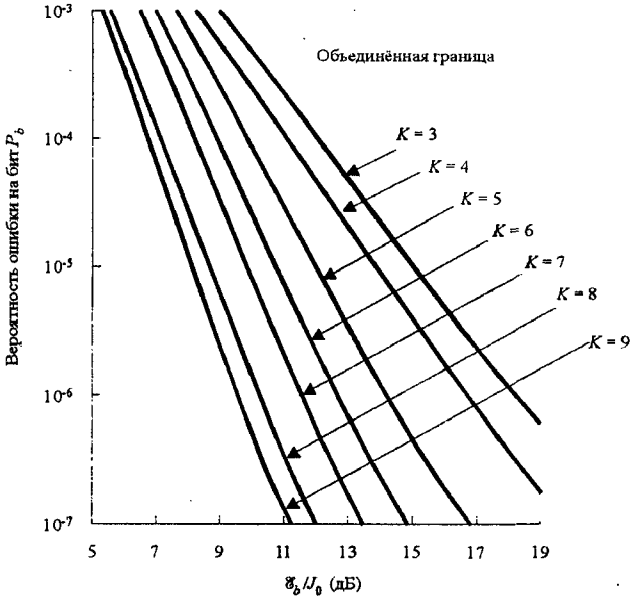


Рис. 13.2.11. Характеристики свёрточного кода со скоростью 1/2 с декодером жёстких решений Витерби, двоичной ФМ с оптимальным ПВМС
[*Martin* и *McAdam* (1980), © 1980 *IEEE*]

Для примера заметим, что при $P_2 = 10^{-6}$ сверточный код с K=7 при декодировании мягких решений требует $\mathcal{E}_b / J_0 = 7{,}6$ в то время декодирование жёстких решений требует $\mathcal{E}_b / J_0 = 11{,}7$ дБ. Эта разница в 4,1 дБ в ОСШ относительно велика. При непрерывном гауссовском шуме соответствующие ОСШ для вероятности ошибки 10^{-6} равна 5 дБ для декодирования мягких решений и 7 дБ для декодирования жестких решений. Таким образом, случай наихудшего ПВМС уменьшает качество на 2,6 дБ для декодирования мягких решений и на 4,7 дБ для декодирования жестких решений. Эти уровни деградации увеличиваются по мере уменьшения кодового ограничения. Важнейшее обстоятельство, однако, то, что потери в ОСШ, обусловленные помехой, уменьшаются от 40 дБ для некодированной системы до менее чем 5 дБ для кодированной системы при использовании сверточного кода с К=7 и скорости 1/2.

Простой метод для вычисления качества (кодированной) системы в условиях действия ПВМС сводится к вычислению параметра предельной скорости R_0, как предложено Омурой и Левитом (1982). Например, для двоичной кодовой модуляции предельную скорость можно выразить так

$$R_0 = 1 - \log(1 + D_\alpha), \qquad (13.2.64)$$

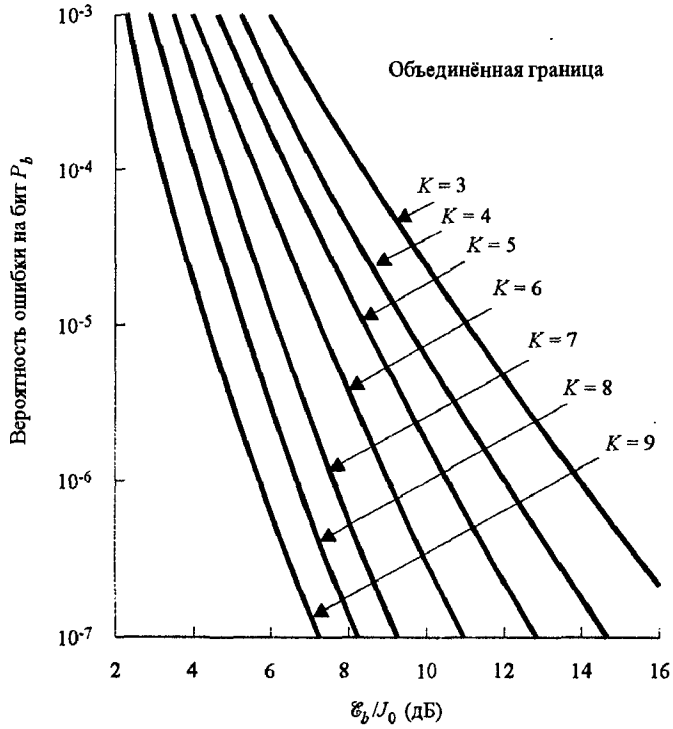


Рис. 13.2.12. Характеристики свёрточного кода со скоростью 1/2 с декодером мягких решений Витерби, двоичной ФМ с оптимальным ПВМС [*Martin* и *McAdam* (1980), © 1980 *IEEE*]

где D_α зависит от шумовых характеристик канала и обработки декодером. Напомним, что для двоичной ФМ в канале с АБГШ и декодировании мягких решений

$$D_\alpha = e^{-\mathscr{E}_c/N_0}, \tag{13.2.65}$$

где $\mathscr{E}_c$ – энергия на кодовый символ. При декодировании жестких решений

$$D_\alpha = \sqrt{4p(1-p)} \tag{13.2.66}$$

где p – вероятность ошибочного приема кодового символа. Здесь мы имеем $N_0 \equiv J_0$.

Для кодированной двоичной ФМ при импульсной помехе Омура и Левит (1982) показали, что

$$D_\alpha = \alpha e^{-\alpha\mathscr{E}_c/N_0} \tag{13.2.67}$$

для декодировании мягких решений при знании состояния глушителя,

$$D_\alpha = \min_{\lambda\geq 0}\left\{\left[\alpha\exp\left(\lambda^2\mathscr{E}_c N_0/\alpha\right)+1-\alpha\right]\exp\left(-2\lambda\mathscr{E}_c\right)\right\} \tag{13.2.68}$$

для декодирования мягких решений при незнании состояния глушителя,

$$D_\alpha = \sqrt{4p(1-p)} \tag{13.2.69}$$

для декодировании жестких решений при знании состояния глушителя,

$$D_\alpha = \sqrt{4\alpha p(1-\alpha p)} \tag{13.2.70}$$

для декодирования жестких решений при знании состояния глушителя, где вероятность ошибки декодирования жёстких решений для двоичной ФМ равна

$$p = Q\left(\sqrt{\frac{2\alpha\mathscr{E}_c}{N_0}}\right).$$

Графики R_0, как функции от $\mathscr{E}_c/N_0$, иллюстрируются на рис.13.2.13 для случаев, оговоренных выше. Заметим, что эти графики представляют предельно-достижимую

скорость для наихудшей величины $\alpha = \alpha^*$, которая максимизирует D_α (минимизирует R_0 при каждом значении $\mathscr{E}_c / N_0$. Далее заметим, что при декодировании мягких решений и отсутствии знаний о состоянии глушителя $R_0 = 0$. Такая ситуация возникает потому, что выход демодулятора не квантован.

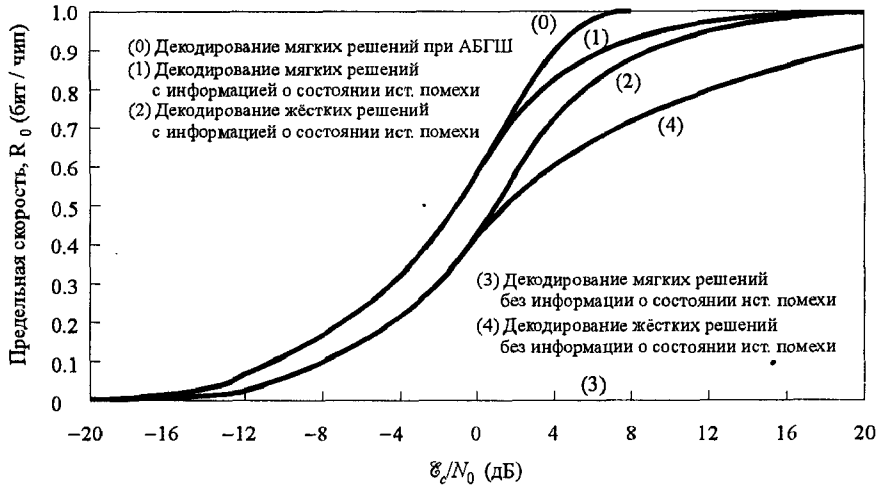


Рис. 13.2.13. Предельная скорость для ПП двоичной ФМ
[*Omura* и *Levitt* (1982), © 1982 *IEEE*]

Графики рис.13.2.13 можно использовать для расчета качества кодированных систем Чтобы продемонстрировать процедуру, предположим, что мы желаем определить требуемое ОСШ для достижения вероятности ошибки 10^{-6} с кодированной двоичной ФМ при наихудшем случае импульсного глушения. Для конкретности предположим, что мы имеем сверточный код с К=7 и скоростью кода 1/2. Начинаем с расчёта качества сверточного кода с К=7, скоростью 1/2 при декодировании мягких решений в канале с АБГШ. При $P_2 = 10^{-6}$ требуемое ОСШ находим из рис. 8.2.21.

$$\mathscr{E}_c / N_0 = 5\,\text{дБ}.$$

Поскольку скорость кода равна $\frac{1}{2}$, имеем

$$\mathscr{E}_c / N_0 = 2\,\text{дБ}.$$

Теперь переходим к графикам рис.13.2.13 и находим, что для канала с АБГШ (классический канал) с $\mathscr{E}_c / N_0 = 2\,\text{дБ}$ соответствующая величина предельной скорости равна

$$R_0 = 0{,}74 \text{ бит/символ}.$$

Если мы имеем другой канал с другими характеристиками шума (наихудший случай канала с импульсным шумом), но с той же величиной предельной скорости R_0, тогда верхняя граница вероятности ошибки на бит та же, т.е. 10^{-6} в нашем случае Следовательно, мы можем использовать эту скорость для определения ОСШ, требуемого для случая канала с наихудшим импульсным глушением. Из графика рис.13.2.13 находим:

$$\frac{\mathscr{E}_c}{J_0} = \begin{cases} 10\,\text{дБ} & \text{для декодирования жёстких решений при отсутствии знания о состоянии источника помехи.} \\ 5\,\text{дБ} & \text{для декодирования жёстких решений при наличии знаний о состоянии источника помехи.} \\ 3\,\text{дБ} & \text{для декодирования мягких решений при наличии знания о состоянии источника помехи.} \end{cases}$$

Следовательно, соответствующие значения $\mathscr{E}_b / J_0$ для сверточного кода с $K = 7$ и скоростью $1/2$ равно 13,8 и 6 дБ, соответственно.

Этот общий подход можно использовать для создания графиков вероятности ошибки для кодированных двоичных сигналов в случае канала с наихудшим ПВМС путём использования соответствующих графиков вероятности ошибки для канала с АБГШ. Подход, который мы описали выше, легко обобщить на M-позиционную систему сигналов, как указано Омурой и Левитом (1982).

Сравнивая предельную скорость для кодированной широкополосной системы с ПП при двоичной ФМ, показанную на рис.13.2.13, мы видим, что для скоростей ниже 0,7 нет ухудшения в ОСШ при декодировании мягких решений и при знании состояния источника помехи по сравнению с качеством канала с АБГШ $(\alpha = 1)$. С другой стороны, при $R_0 = 0{,}7$ имеется разница в качестве на 6 дБ для ОСШ в канале с АБГШ и требуемый для декодирования жестких решений при отсутствии информации о состоянии источника помехи. При скоростях ниже 0,4 нет потерь в ОСШ при декодировании с жёстким решением при неизвестном состоянии источника помехи. Однако имеются и ожидаемые потери на 2 дБ при декодировании жестких решений по сравнению декодированием мягких решений в канале с АБГШ.

13.2.4. Генерирование ПШ последовательностей

Генерирование ПШ последовательностей для применения широкополосных сигналов является темой, которая привлекла особое внимание в технической литературе. Мы вкратце обсудим конструкцию некоторых ПШ последовательностей и представим важные свойства автокорреляционной и взаимокорреляционной функций таких последовательностей. Для исчерпывающей трактовки этого вопроса интересующемуся читателю рекомендуется книга Голомба (1967).

Пожалуй, наиболее широко известными двоичными ПШ последовательностями являются последовательности максимальной длины сдвигового регистра, введенные в разделы 8.1.3 в контексте кодирования, и которые снова предлагались для использования как низкоскоростные коды. Последовательность максимальной длины сдвигового регистра, или m-последовательность для краткости, имеет длину $n = 2^m - 1$ символов и генерируется m ячеечным регистром сдвига с линейной обратной связью, как иллюстрирует рис.13.2.14. Последовательность периодическая с периодом n. Каждый период последовательности содержит 2^{m-1} единиц и $2^{m-1} - 1$ нулей.

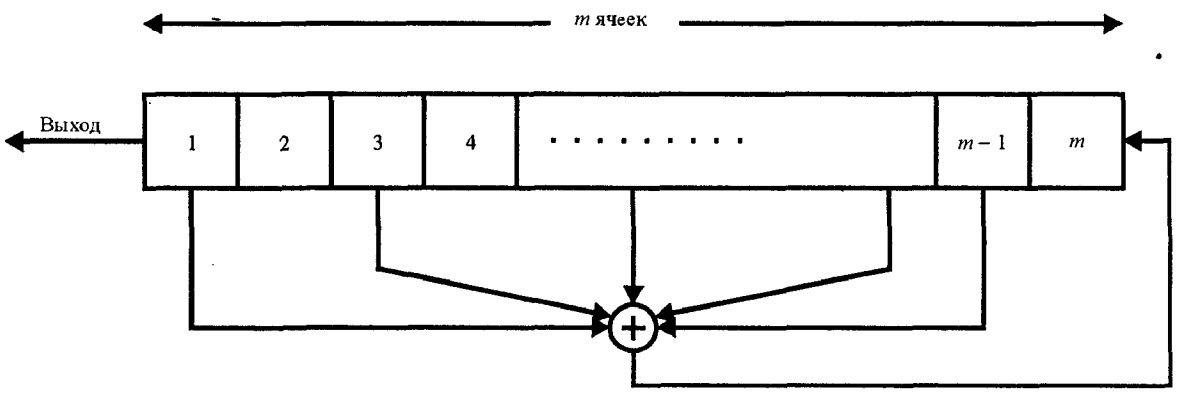


Рис. 13.2.14. Схема m-каскадного регистра сдвига с линейной обратной связью

Для применения в качестве широкополосных ПП сигналов двоичная последовательность с элементами $\{0,1\}$ отображается соответствующей последовательностью из положительных и отрицательных импульсов согласно соотношению

$$p_i(t) = (2b_i - 1)p(t - iT),$$

где $p_i(t)$ – импульс, соответствующий элементу b_i в последовательности с элементами $\{0,1\}$. Эквивалентно, мы можем сказать, что двоичная последовательность с элементами $\{0,1\}$ обращается в соответствующую двоичную последовательность с элементами $\{-1,1\}$. Мы будем называть эквивалентную последовательность с элементами $\{-1,1\}$ *биполярной последовательностью*, поскольку она определяется импульсами с положительными и отрицательными амплитудами.

Важной характеристикой периодической ПШ последовательности – её периодическая автокорреляционная функция, которая обычно определяется через слагаемые биполярной последовательности так

$$\phi(j) = \sum_{i=1}^{n} (2b_i - 1)(2b_{i+j} - 1), \qquad 0 \le j \le n-1, \qquad (13.2.71)$$

где n – период. Ясно, что $\phi(j + rn) = \phi(j)$ для любого целого r.

В идеале псевдослучайная последовательность должна иметь автокорреляционную функцию со свойством $\phi(0) = n$ и $\phi(j) = 0$ для $1 \le j \le n-1$. В случае m-последовательности периодическая функция автокорреляции равна

$$\phi(j) = \begin{cases} n, & (j = 0) \\ -1, & (1 \le j \le n-1) \end{cases} \qquad (13.2.72)$$

Для больших значений n, т.е. для длинных m-последовательностей, величина отношения пиковых значений боковых лепестков $\phi(j)$ к пиковой величине функции корреляции $\phi(j)/\phi(0) = -1/n$ мала и, с точки зрения практики, пренебрежима. Следовательно, m-последовательность почти идеальная, если посмотреть с точки зрения корреляционной функции.

С точки зрения антипомеховых (АП) приложений ПШ широкополосных сигналов, период последовательности должен быть большим для того, чтобы затруднить источнику помехи изучить схему соединений в цепи обратной связи ПШ генератора. Однако, это требование непрактично в большинстве случаев, поскольку постановщик мешающего сигнала может определить соединения обратной связи путём наблюдения только $2m$ чипов ПШ последовательности. Эта уязвимость ПШ последовательности обусловлена линейными свойствами генератора. Чтобы снизить уязвимость, выходные последовательности с отдельных ячеек регистра сдвига или выходы отдельных различных m последовательностей соединяются нелинейным путем, чтобы образовать нелинейную последовательность, которая значительно более трудна для изучения постановщиком помехи. Дальнейшее снижение уязвимости достигается путем частой смены соединений обратной связи и (или) числа ячеек регистра сдвига, в соответствие с заранее согласованным планом между передатчиком и заданным приемником.

В некоторых приложениях взаимокорреляционные свойства ПШ последовательностей столь же важны, как корреляционные свойства. Для примера, в CDMA каждому пользователю присваивается индивидуальная ПШ последовательность. В идеале ПШ последовательности отдельных пользователей должны быть взаимно ортогональны так, чтобы уровень интерференции, испытываемый одним пользователем от передачи других

пользователей, был бы равен нулю. Однако ПШ последовательности, используемые различными пользователями, на практике не обладают строгой ортогональностью.

Для конкретности, рассмотрим класс m-последовательностей. Известно (Сарвейт и Пурслей, 1980), что периодическая взаимокорреляционная функция между парой m последовательностей на том же периоде может иметь относительно большие пики. Таблица 13.2.1 дает пиковые амплитуды ϕ_{max} для периодической функции взаимной корреляции между парами m-последовательностей для $3 \le m \le 12$. Таблица также показывает число m последовательностей длины $n = 2^m - 1$ для $3 \le m \le 12$. Как мы можем видеть, число m последовательностей длины n быстро растет с m. Мы также видим, что для большинства последовательностей пиковые амплитуды ϕ_{max} взаимокорреляционной функции составляют большой процент от пикового значения автокорреляционной функции.

Таблица 13.2.1. Пиковые значения взаимной корреляции m-последовательностей и последовательностей Голда

m	$n = 2^m - 1$	**Число m-последовательностей**	**Пиковые значения взаимной корреляции ϕ_{max}**	$\phi_{max}/\phi(0)$	$t(m)$	$t(m)/\phi(0)$
3	7	2	5	0,71	5	0,71
4	15	2	9	0,60	9	0,60
5	31	6	11	0,35	9	0,29
6	63	6	23	0,36	17	0,27
7	127	18	41	0,32	17	0,13
8	255	16	95	0,37	33	0,13
9	511	48	113	0,22	33	0,06
10	1023	60	383	0,37	65	0,06
11	2047	176	287	0,14	65	0,03
12	4095	144	1407	0,34	129	0,03

Такие большие значения для взаимных корреляций нежелательны в CDMA. Хотя возможно выбрать малое подмножество m последовательностей, которые имеют относительно малые значения пиков взаимной корреляции, число последовательностей в этом подмножестве слишком мало для применений CDMA.

ПШ последовательности с лучшими свойствами периодической функции взаимной корреляции, чем m последовательности, даны Голдом (1967, 1968) и Касами (1966). Они образуются из m последовательностей, как описано ниже.

Голд и Касами доказали, что некоторые пары m последовательностей длины n имеют взаимно корреляционную функцию с тремя уровнями $\{-1, -t(m), t(m)-2\}$, где

$$t(m) = \begin{cases} 2^{(m+1)/2} + 1 & (\text{нечётные } m), \\ 2^{(m+2)/2} + 1 & (\text{чётные } m). \end{cases} \qquad (13.2.73)$$

Для примера, если $m = 10$, тогда $t(10) = 2^6 + 1 = 65$ и три возможные значения периодической взаимокорреляционной функции равны $\{-7, -65, 63\}$. Таким образом, максимальное значения (по модулю) взаимной корреляции пары m-последовательностей равно 65, в то время как пик для семейства 60 возможных последовательностей,

генерируемых 10-разрядным регистром сдвига с различными соединениями обратной связи равен $\phi_{max} = 383$ – примерно шестикратная разница в пиковых значениях. Две m-последовательности длины n с периодической взаимокорреляционной функцией, которая принимает значения $\{-1, -t(m), t(m)-2\}$ называют *предпочтительными последовательностями*.

Из пары предпочтительных последовательностей, скажем, $\mathbf{a} = [a_1, a_2, \ldots a_n]$ и $\mathbf{b} = [b_1, b_2, \ldots b_n]$, мы конструируем ансамбль последовательностей длины n, взяв сумму по mod 2 последовательности **a** и n циклически сдвинутых версий **b** или наоборот. Таким образом, мы получаем n новых периодических последовательностей[1] с периодом $n = 2^m - 1$. Мы можем также включить в ансамбль исходные последовательности **a** и **b** и, таким образом, имеем n+2 последовательностей. n+2 последовательностей, сконструированных таким образом, называют *последовательностями Голда*.

Пример 13.2.4. Рассмотрим генерацию последовательностей Голда длины $n = 2^5 - 1 = 31$. Как указано выше, для m=5 пик взаимной корреляции равен

$$t(5) = 2^3 + 1 = 9.$$

Две предпочтительные последовательности, которые найдены Питерсоном и Уэлдоном (1972), описываются полиномами

$$g_1(p) = p^5 + p^2 + 1,$$
$$g_2(p) = p^5 + p^4 + p^2 + p + 1.$$

Регистры сдвига для генерирования двух m-последовательностей и соответствующих последовательностей Голда показаны на рис.13.2.15. В этом случае имеется 33 различных последовательностей, соответствующие 33 различным взаимным сдвигам двух m-последовательностей. Из них 31 последовательность не является последовательностями максимальной длины.

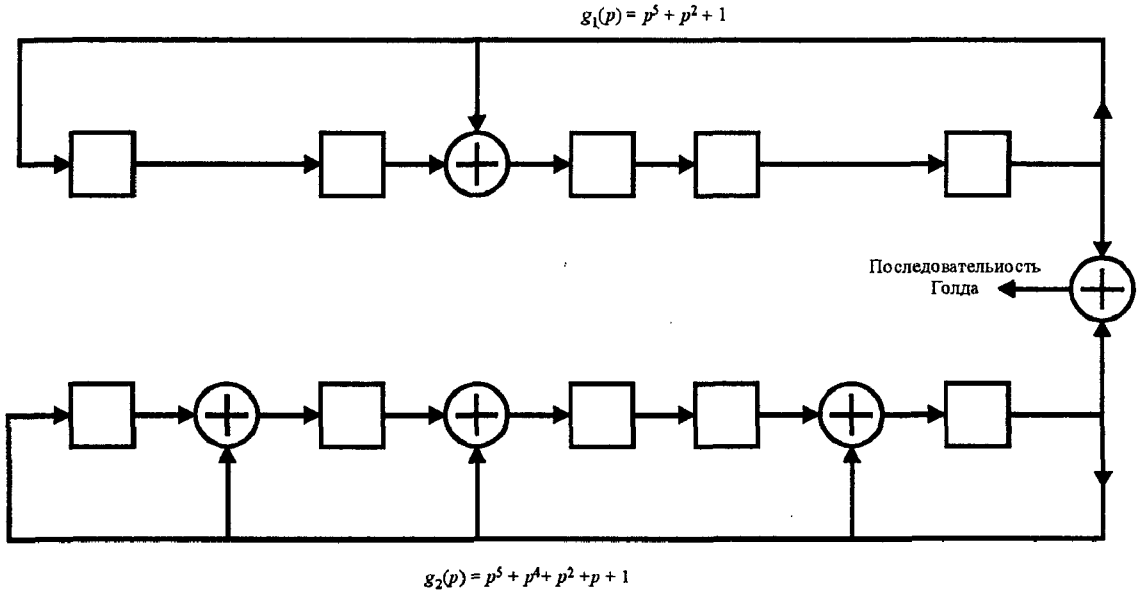


Рис. 13.2.15. Генерирование последовательности Голда длиной 31

[1] Эквивалентный метод генерирования n новых последовательностей сводится к использованию регистра сдвига длины $2m$ с соединениями обратной связи, определяемые полиномами $h(p) = g_1(p)g_2(p)$, где $g_1(p)$ и $g_2(p)$ – это полиномы, которые определяют соединения обратной связи для m-ячеечных регистров сдвига, которые генерируют m-последовательности **a** и **b**.

Исключая последовательности **a** и **b**, ансамбль последовательностей Голда не включает в себя последовательности максимальной длины n регистра сдвига. Следовательно, их автокорреляционные функции не являются двоичными. Голд (1968) показал, что взаимокорреляционная функция любой пары последовательностей ансамбля n+2 последовательностей Голда является троичной с возможными значениями

$$\{-1, -t(m), t(m)-2\},$$

где $t(m)$ определяется (13.2.73). Аналогично, пиковые значения автокорреляционной функции для последовательностей Голда принимают значения из множества $\{-1, -t(m), t(m)-2\}$. Таким образом, пиковые значения автокорреляционной функции ограничена сверху $t(m)$.

Величины пиков автокорреляционной функции и пиков взаимокорреляционной функции, т.е. $t(m)$, для последовательностей Голда даны в табл.13.2.7. Также даны значения, нормированные к $\phi(0)$.

Интересно сравнить пиковые значения взаимной корреляции последовательностей Голда с известной нижней границей взаимной корреляции между произвольной парой двоичных последовательностей периода n в ансамбле из M последовательностей. Нижняя граница, найденная Уолшем (1974), для ϕ_{max} равна

$$\phi_{max} \geq n\sqrt{\frac{M-1}{Mn-1}}, \qquad (13.2.74)$$

которая хорошо аппроксимируется, для больших n и M, как $\sqrt{n}$. Для последовательностей Голда $n = 2^m - 1$ и, следовательно, нижняя граница $\phi_{max} \approx 2^{m/2}$. Эта граница ниже на $\sqrt{2}$ для нечётных m и на 2 для чётных m относительно $\phi_{max} = t(m)$ для последовательностей Голда.

Процедура, похожая на использованную при генерировании последовательностей Голда, может генерировать более узкий ансамбль из $M = 2^{m/2}$ двоичных последовательностей периода $n = 2^m - 1$, когда m чётно. В этой процедуре мы начинаем с m последовательности **a** и формируем двоичную последовательность **b**, взяв каждый $2^{m/2}+1$ символ из **a**. Таким образом, последовательность **b** формируется путём децимации **a** через $2^{m/2}+1$. Можно показать, что полученная последовательность **b** периодическая с периодом $2^{m/2}-1$. Для примера, если $m = 10$, то период **a** равен n=1023, а период **b** равен 31. Следовательно, если мы наблюдаем 1023 символа последовательности **b** мы можем видеть 33 повторений 31 символьных последовательностей. Теперь, взяв $n = 2^m - 1$ символа из последовательностей **a** и **b**, мы формируем новый ансамбль последовательностей путем суммирования по mod 2 символов из **a** и символов из **b** и всех $2^{m/2}-2$ циклических сдвигов символов из **b**. Включая **a** в ансамбль, мы получаем ансамбль из $2^{m/2}$ двоичных последовательностей длины $n = 2^m - 1$. Их называют *последовательностями Касами*. Автокорреляционная и взаимокорреляционная функции этих последовательностей принимает значения из ряда $\{-1, -(2^{m/2}+1), 2^{m/2}-1\}$. Следовательно, значение максимума взаимной корреляции для любой пары последовательностей этого ансамбля равно

$$\phi_{max} = 2^{m/2} + 1 \qquad (13.2.75)$$

Эта величина ϕ_{max} удовлетворяет нижней границе Уолша для ансамбля из $2^{m/2}$ последовательностей длины $n=2^m-1$. Таким образом, последовательности Касами оптимальны.

Кроме хорошо известных последовательностей Голда и Касами, имеются другие двоичные последовательности, подходящие для применения в CDMA. Интересующемуся читателю рекомендуем работы Шольца (1979), Олсена (1972) и Сарвейта и Пурслея (1980).

В заключение хотим отметить, что хотя мы обсудили периодические взаимокорреляционные функции между парами периодических последовательностей много практических систем CDMA могут использовать длительности информационных символов, которые составляют только части периодических последовательностей. В таких случаях важным является частично-периодическая взаимная корреляция между двумя последовательностями. Определенное число статей обсуждает эти проблемы, включая статьи Линдхольма (1968), Вайнберга и Вольфа (1970), Фридриксона (1975), Бекира и др. (1978) и Пурслея (1979).

13.3. ШИРОКОПОЛОСНЫЕ СИГНАЛЫ СО СКАЧКАМИ ЧАСТОТЫ

В широкополосных системах связи со скачками частоты (СЧ) предоставленная полоса частот канала подразделяется на большое число прилегающих частотных полосок. В любом сигнальном интервале передаваемый сигнал занимает одну или больше возможных частотных полосок. Выбор частотной полоски в каждом сигнальном интервале делается псевдослучайно, согласно выходу ПШ генератора. Рис.13.3.1 иллюстрирует частный образец системы со скачками частоты в частотно-временной области.

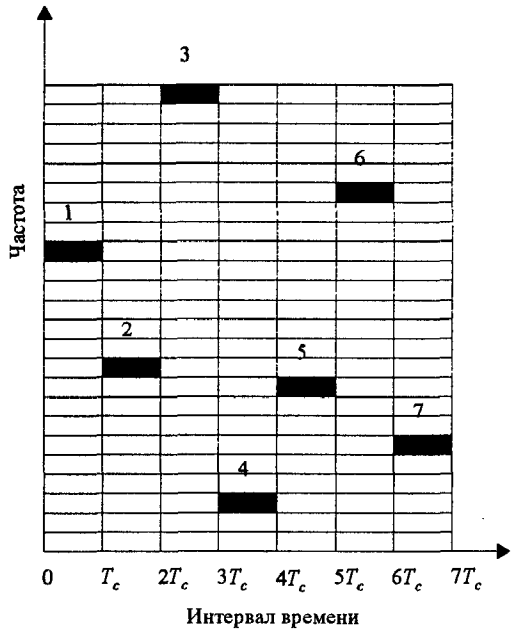


Рис.13.3.1. Пример расположения рабочих участков частотио-временного поля в системе со скачками частоты (СЧ)

Блок-схема передатчика и приёмника для широкополосной системы со скачками частоты показан на рис.13.3.2. Модуляция обычно двоичная или M-ичная ЧМ. Для примера, если используется двоичная ЧМ, модулятор выбирает одну из двух частот, соответствующих передаче 0 или 1. Результирующий ЧМ сигнал передается на частоте, величина которой определяется выходом ПШ генератора, который обратно используется

для выбора частоты, которая синтезируется синтезатором частоты. Сигнал на этой частоте смешивается с выходом модулятора и результирующий частотно-транслируемый сигнал предается по каналу. Для примера, m символов ПШ генератора можно использовать для определения $2^m - 1$ возможных частот трансляции.

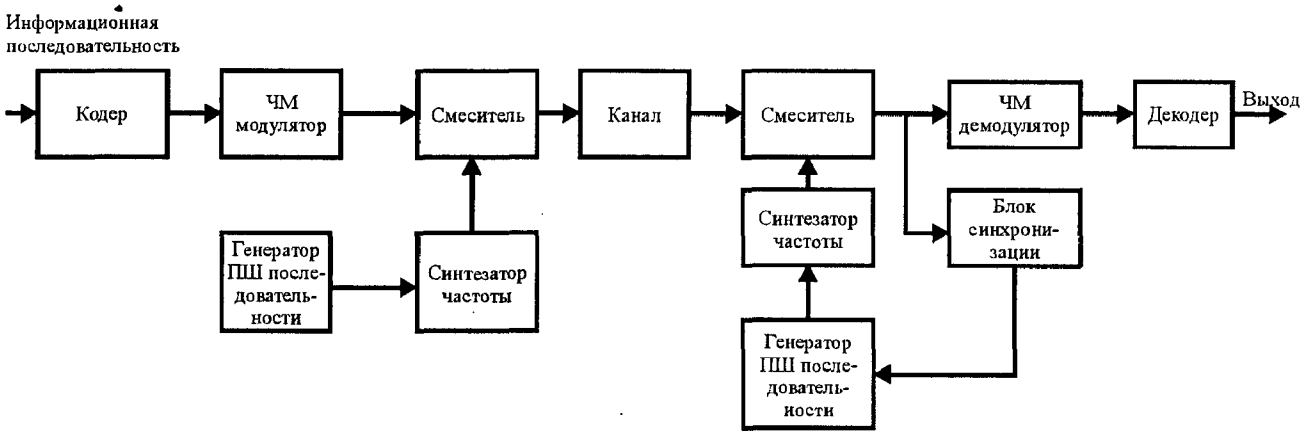


Рис. 13.3.2. Блок-схема широкополосной системы с СЧ

На приёме мы имеем идентичный ПШ генератор, синхронизированный с принимаемым сигналом, который используется для управления выходом синтезатора частоты. Таким образом, псевдослучайная частота трансляции, введенная на передаче, восстанавливается на приеме путем смешивания выхода синтезатора с принимаемым сигналом. Результирующий сигнал демодулируется посредством демодулятора ЧМ сигнала. Сигнал для поддержания синхронизма ПШ генератора с принимаемым сигналом обычно извлекается из принимаемого сигнала.

Хотя ФМ даёт лучшее качество, чем ЧМ в канале с АБГШ, трудно поддерживать фазовую когерентность при синтезе частот, используемых в системах со скачками частоты. Дело объясняется также условиями распространения сигнала различных частот по каналу, так как сигнал «прыгает» от одной частоты к другой в пределах широкой полосы. Поэтому в широкополосных системах с СЧ обычно используется ЧМ с некогерентным детектированием.

В системе со скачками частоты, показанной на рис.13.3.2, частота несущей скачет псевдослучайно на каждом сигнальном интервале. M частот, несущие информацию примыкают друг к другу и выбираются с интервалом $1/T_c$, где T_c – сигнальный интервал. Этот вид скачка частоты называется *блоковым скачком*.

Другой вид скачка частоты, который меньше уязвим для некоторых стратегий постановки помех – это независимые скачки частоты внутри блока. В этой схеме M возможных частот модулятора выбирается с широким разбросом частотных полосок. Один метод такой реализации иллюстрируется на рис.13.3.3. Здесь m символов ПШ генератора и k информационных символов используется для определения частотных полосок передаваемого сигнала.

Скорость скачков частоты обычно выбирается равной или больше скорости передачи (кодированных или не кодированных) символов. Если имеется много скачков частоты на символ, мы имеем сигнал с быстрыми скачками. С другой стороны, если скачки совершаются со скоростью передачи символов, мы имеем сигнал с медленными скачками.

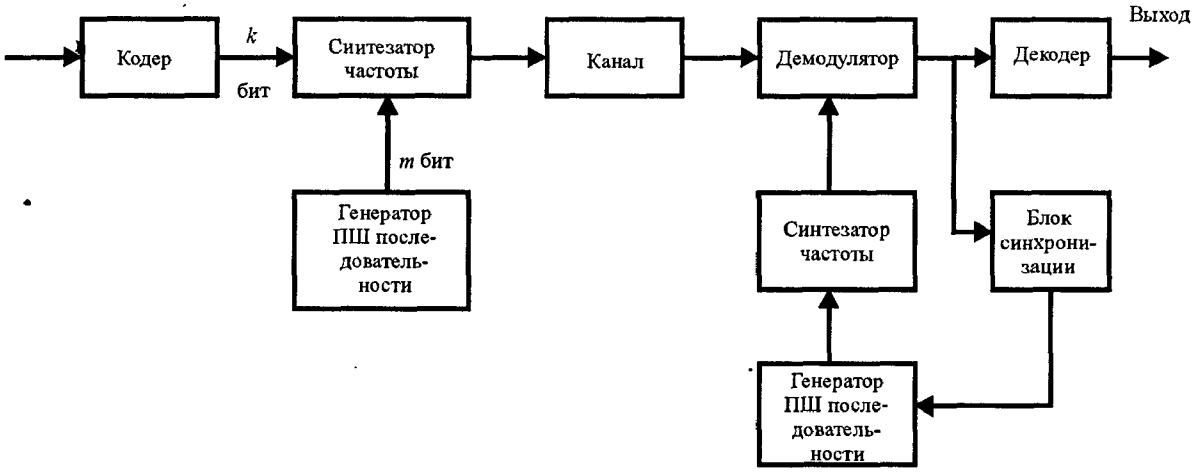


Рис. 13.3.3. Блок-схема широкополосной системы с СЧ с независимым тоном

Быстрые скачки частоты используются в антипомеховых приложениях, когда необходимо помешать определенному типу помех, называемых *«преследующим глушителем»*, чтобы он не имел достаточно времени для перехвата частоты и для её ретрансляции так, чтобы создать интерференционные сигнальные компоненты. Однако здесь имеется ухудшение, вызванное делением сигнала на различные элементы со скачками частоты, поскольку из-за некогерентной обработки энергия отдельных элементов используется не полностью. Следовательно, демодулятор испытывает ухудшение из-за некогерентного сложения, как описано в разделе 12.1.

Широкополосные сигналы с СЧ были впервые использованы в цифровых системах связи, которые требовали АП проектирования и в системах CDMA, где многие пользователи работают в общей полосе частот. В большинстве случаев СЧ сигнал является более предпочтительным, чем широкополосный ПП сигнал из-за свойственных ему более высоких требований к синхронизации. В частности, в ПП системе синхронизация должна быть установлена внутри части интервала чипа $T_c \approx 1/W$. С другой стороны, в системе СЧ чип-интервал – это интервал, выделенный передаваемому символу в отдельной частотной полоске с полосой $B << W$. Этот интервал примерно равен $1/B$, что намного больше, чем $1/W$. Таким образом, требования к синхронизации в системе СЧ не так строги, как в ПП системе.

В разделах 13.3.2 и 13.3.3 мы сконцентрируем наше внимание на AJ и CDMA применениях широкополосных сигналов с СЧ. Сначала мы определим вероятности ошибки для кодированных и некодированных сигналов СЧ в присутствии широкополосной интерференции в виде АБГШ. Затем мы рассмотрим более серьезный вид интерференции, которая возникает в AJ и CDMA приложениях, называемый *парциально-полосовой интерференцией*. Определяются преимущества кодирования, получаемые при таком типе интерференции. Мы включили в обсуждение в разделе 13.3.3 пример системы CDMA с СЧ, которая была спроектирована для нужд подвижных пользователей спутникового канала.

13.3.1. Качество широкополосных сигналов со скачками частоты (СЧ) в канале с АБГШ

Рассмотрим качество широкополосных сигналов с СЧ в присутствии широкополосной интерференции, характеризуемой статистически как АБГШ со спектральной плотностью мощности J_0. Для двоичной ортогональной ЧМ с некогерентным детектированием при

медленных скачках частоты (1 скачок на символ) вероятность ошибки, полученная в разделе 5.4.1, равна

$$P_2 = \tfrac{1}{2} e^{-\gamma_b/2}, \qquad (13.3.1)$$

где $\gamma_b = \mathscr{E}_b / J_0$. С другой стороны, если символьный интервал разделить на L подынтервалов и в каждом подынтервале передаётся двоичный сигнал ЧМ со СЧ; мы имеем сигнал СЧ с быстрыми скачками. При квадратичном сложении выходных сигналов от соответствующих согласованных фильтров для L подынтервалов, вероятность ошибки сигнала с СЧ, следующий из результатов раздела 12.1, равна

$$P_2(L) = \frac{1}{2^{2L-1}} e^{-\gamma_b/2} \sum_{i=0}^{L-1} K_i \left(\tfrac{1}{2}\gamma_b\right)^i, \qquad (13.3.2)$$

где ОСШ на бит $\gamma_b = \mathscr{E}_b / J_0 = L\gamma_c$, γ_c – ОСШ на чип в L-чиповом символе и

$$K_i = \frac{1}{i!} \sum_{r=0}^{L-1-i} \binom{2L-1}{r}. \qquad (13.3.3)$$

Напомним, что для заданного ОСШ на бит γ_b вероятность ошибки, следующая из (13.3.2) больше, чем получаемая из (13.3.1). Разница в ОСШ для заданной вероятности ошибки и заданном L названы *потерями некогерентного сложения*, которые были описаны и проиллюстрированы в разделе 12.1.

Кодирование улучшает качество широкополосного сигнала со СЧ на величину, которую мы назвали *выигрышем кодирования*, зависящим от параметров кода.

Допустим, что мы используем линейный (n,k) блоковый код и двоичную ЧМ с одним скачком на кодированный символ для передачи символов. При декодировании мягких решений, когда осуществляется квадратичная демодуляция ЧМ сигнала, вероятность ошибки кодового слова

$$P_M \le \sum_{m=2}^{M} P_2(m), \qquad (13.3.4)$$

где $P_2(m)$ – вероятность ошибки при различении m-го кодового слова и кодового слова из одних нулей, когда действительно передается последний. Выражение для $P_2(m)$ было получено в разделе 8.1.4 и оно имеет ту же форму, что (13.3.2) и (13.3.3) с заменой L на w_m и γ_b на $\gamma_b R_c w_m$, где w_m – вес m-го кодового слова, а R_c – скорость кода. Произведение $R_c w_m$, которое не меньше, чем $R_c d_{\min}$, представляет выигрыш кодирования. Таким образом, мы имеем оценку качества системы с СЧ с блоковым кодом при медленных скачках частоты и широкополосной интерференции.

Вероятность ошибки при быстрых скачках частоты с n_2 скачками на кодовый символ можно получить, если переинтерпретировать вероятность двоичного ошибочного события $P_2(m)$ в (13.3.4). Наличие n_2 скачков на символ можно интерпретировать как код с повторением, который, будучи соединён с нетривиальным (n_1,k) двоичным линейным кодом, имеющим распределение весов $\{w_m\}$, дает $(n_1 n_2, k)$ двоичный линейный код с распределением весов $\{n_2 w_m\}$. Следовательно, $P_2(m)$ имеет форму (13.3.2) с заменой L на $\{n_2 w_m\}$ γ_b на $\gamma_b R_c n_2 w_m$, где $R_c = k/n_1 n_2$. Заметим, что $\gamma_b R_c n_2 w_m = \gamma_b w_m k/n_1$, что как раз определяет выигрыш кодирования, полученный нетривиальным (n_1,k)-кодом. Следовательно, использование кода с повторением ведет к увеличению потерь от некогерентного сложения.

При декодировании жестких решений и при медленных скачках частоты, вероятность ошибки на бит на выходе демодулятора при некогерентном детектировании равна

$$P = \tfrac{1}{2} e^{-\gamma_b R_c/2}. \qquad (13.3.5)$$

Вероятность ошибочного детектирования кодового слова легко ограничить сверху, если использовать границу Чернова, так

$$P_M \le \sum_{m=2}^{M} [4p(1-p)]^{w_m/2}. \qquad (13.3.6)$$

Однако, при быстрых скачках частоты с n_2 скачками на кодовый символ и квадратичным сложением выходов от соответствующих согласованных фильтров для n_2 скачков, как это делается при декодировании мягких решений, при формирования двух величин для решения о кодовом символе вероятность ошибки на бит p теперь определяется (13.3.2) с заменой L на n_2 и γ_b на $\gamma_b R_c n_2$, где R_c – скорость нетривиального кода (n_1, k). Следовательно, качество системы с быстрыми скачками частоты и широкополосной интерференцией ухудшается относительно СЧ систем с медленными скачками на величину, равную потерям некогерентного сложения принимаемых сигналов от n_2 скачков.

Мы видели, что как при декодировании жестких, так и мягких решений, использование кода с повторением в системе с быстрыми скачками частоты не дает выигрыша кодирования. Единственный выигрыш кодирования получается от (n_1, k) блокового кода. Таким образом, код с повторением неэффективен в системе с СЧ с быстрыми скачками при некогерентном сложении. Более эффективен такой метод кодирования, в котором используется либо простой низкоскоростной двоичный код, либо каскадный код. Дополнительное улучшение качества можно получить, используя недвоичные коды в соединении с M-ичной ЧМ. Границы для вероятности ошибки для этого случая можно получить из результатов, данных в разделе 12.1.

Хотя мы рассчитали выше только качество линейных блоковых кодов, легко получить результаты качества для двоичных сверточных кодов. Мы оставим в качестве упражнения для читателя расчет вероятности ошибки на бит для декодирования по Витерби мягких и жестких решений сигналов с СЧ, поражённых широкополосной интерференцией.

В заключение мы заметим, что энергию на бит $\mathscr{E}_b$ можно выразить так: $\mathscr{E}_b = P_{ср}/R$, где R – информационная скорость в битах в секунду, а $J_0 = J_{ср}/W$. Следовательно, γ_b можно выразить так:

$$\gamma_b = \frac{\mathscr{E}_b}{J_0} = \frac{W/R}{J_{ср}/P_{ср}}. \qquad (13.3.7)$$

В этом выражении мы узнаём W/R как выигрыш обработки и $J_{ср}/P_{ср}$ как помехозащищённость (запас помехоустойчивости) для широкополосного сигнала СЧ.

13.3.2. Качество широкополосных сигналов с СЧ при парционально-полосовой интерференции

Парционально-полосовая интерференция, рассматриваемая в этом подразделе, моделируется как гауссовский случайный процесс с нулевым средним с равномерной спектральной плотностью мощности на доли α от общей полосы W и равный нулю в остальной части полосы. В области или областях, где спектральная плотность мощности ненулевая, ее величина равна $\Phi_{zz}(f) = J_0/\alpha$, $0 < \alpha \le 1$. Эту модель можно применить к jamming сигналу или интерференции от других пользователей в системе с CDMA со СЧ.

Предположим, что парционально-полосовая интерференция возникает от глушителя, который может выбрать α для оптимизации влияния на систему связи. При некодированной псевдослучайной системе со СЧ (с медленными скачками) при использовании двоичной ЧМ и некогерентного детектирования принимаемый сигнал подвергается глушению с вероятностью α и не подвергается глушению с вероятностью $1-\alpha$. Когда он подвергается глушению, вероятность ошибки равна $\frac{1}{2}\exp(-\alpha\mathscr{E}_b/2J_0)$ и когда он не подвергается глушению, ошибок нет. Следовательно, средняя вероятность ошибки равна

$$P_2(\alpha)=\frac{1}{2}\alpha\exp\left(-\frac{\alpha\mathscr{E}_b}{2J_0}\right), \tag{13.3.8}$$

где $\mathscr{E}_b/J_0$ можно также выразить как $(W/R)/(J_{ср}/P_{ср})$.

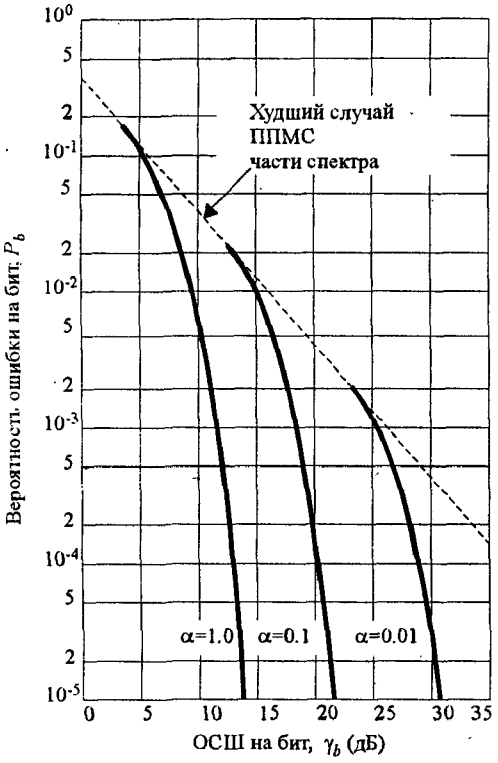


Рис. 13.3.4. Характеристики двоичной ЧМ с парциально-полосовым мешающим сигналом (ППМС)

Рис.13.3.4 иллюстрирует вероятность ошибки, как функцию $\mathscr{E}_b/J_0$, для нескольких значений α. Оптимальная стратегия глушителя сводится к выбору значения α, которое максимизирует вероятность ошибки Дифференцируя $P_2(\alpha)$ и найдя решение для экстремума с учетом ограничения $0\le\alpha\le1$, мы находим, что

$$\alpha^*=\begin{cases}\dfrac{1}{\mathscr{E}_b/2J_0}=2\dfrac{J_{ср}/P_{ср}}{W/R}, & (\mathscr{E}_b/J_0\ge2)\\ 1, & (\mathscr{E}_b/J_0<2)\end{cases} \tag{13.3.9}$$

Соответствующая вероятность ошибки для наихудшего случая парционально-полосового глушителя равна

$$P_2=\frac{e^{-1}}{\mathscr{E}_b/J_0}=\left[e\left(\frac{W/R}{J_{ср}/P_{ср}}\right)\right]^{-1}. \tag{13.3.10}$$

В то время как вероятность ошибки уменьшается экспоненциально при полнокомпонентном глушении, теперь мы нашли, что вероятность ошибки уменьшается только обратно пропорционально $\mathscr{E}_b/J_0$ при наихудшем случае парционально-полосового глушения. Этот результат похож на вероятность ошибки двоичной ЧМ в релеевском канале с замираниями (см. раздел 14.3) и на случай некодированной ПП системы с рассеянным сигналом, пораженным наихудшим вариантом импульса глушения (см. раздел 13.2.3).

Как мы покажем ниже, разнесение сигнала, получаемое посредством кодирования, обеспечивает достаточное улучшение качества относительно некодированных сигналов. Тот же подход к синтезу сигнала также эффективен для передачи сигналов по каналу с замираниями, как мы покажем в главе 14.

Чтобы показать выгоду разнесения для СЧ сигнала с рассеянным спектром при парционально-полосовой интерференции, предположим, что один и тот же информационный символ передан посредством двоичной ЧМ по L независимым скачкам частоты. Это можно выполнить путем деления тактового интервала передачи на L

подынтервалов, как описано раньше при быстрых скачках частоты. После того, как скачки частоты восстановлены в месте приёма, сигнал демодулируется путем пропускания его через ряд согласованных фильтров, чьи выходы подвергаются квадратичному детектированию и стробированию в конце каждого подынтервала. Продетектированные сигналы, соответствующие L скачкам частоты, взвешиваются и суммируются для образования двух величин для решения (метрик), которые обозначим U_1 и U_2.

Если U_1 содержит компоненты сигнала, величины U_1 и U_2 можно выразить так:

$$U_1 = \sum_{k=1}^{L} \beta_k \left| 2\mathscr{E}_c + N_{1k} \right|^2, \qquad U_2 = \sum_{k=1}^{L} \beta_k \left| N_{2k} \right|^2, \tag{13.3.11}$$

где $\{\beta_k\}$ представляет коэффициенты взвешивания, $\mathscr{E}_c$ – энергия сигнала на чип в L-чиповом символе, а $\{N_{jk}\}$ представляет слагаемые аддитивных гауссовских шумов на выходе согласованных фильтров.

Коэффициенты выбираются оптимальным образом, чтобы препятствовать мешающему сигналу полностью подавить полезный сигнал на интервале одного или большего числа скачков. В идеале β_k выбирается обратно пропорционально дисперсии соответствующих шумовых слагаемых $\{N_k\}$. Таким образом, дисперсия шума для каждого чипа нормируется этим взвешиванием к единице, а соответствующий сигнал также соответственно масштабируется. Это означает, что когда сигнальные частоты при некотором скачке поражены помехой, соответствующий вес очень мал. В отсутствие помехи для данного чипа вес относительно большой. На практике при парционально-полосовой помехе взвешивание можно выполнить, используя АРУ, дающую выигрыш, который основан на измерениях мощности шума, полученных на соседних частотах. Это эквивалентно тому, что имеется точная информация о состоянии помехи у декодера.

Предположим, что мы имеем широкополосной гауссовский шум со спектральной плотностью мощности N_0 и парционально-полосовую интерференцию на αW части полосы частот, которая также гауссовская со спектральной плотностью мощности J_0/α. В присутствии парционально-полосовой интерференции вторые моменты слагаемого шума N_{1k} и N_{2k} равны

$$\sigma_b^2 = \frac{1}{2} E\left(\left|N_{1k}\right|^2\right) = \frac{1}{2} E\left(\left|N_{2k}\right|^2\right) = 2\xi_c \left(N_0 + \frac{J_0}{\alpha} \right). \tag{13.3.12}$$

В этом случае мы выберем $\beta_k = 1/\sigma_k^2 = \left[2\mathscr{E}_c\left(N_0 + J_0/\alpha\right)\right]^{-1}$. При отсутствии парционально-полосовой интерференции $\sigma_k^2 = 2\mathscr{E}_c N_0$ и, следовательно, $\beta_k = \left(2\mathscr{E}_c N_0\right)^{-1}$. Заметим, что β_k является случайной величиной.

При демодуляции возникают ошибки, если $U_2 > U_1$. Хотя возможно определить точную вероятность ошибки, мы хотим обратиться к границе Чернова, которая содержит результат, который легче вычислить и интерпретировать. Конкретнее, верхние границы Чернова для вероятности ошибки

$$P_2 = P(U_2 - U_1 > 0) \le E\{\exp[\nu(U_2 - U_1)]\} = E\left\{ \exp\left[-\nu \sum_{k=1}^{L} \beta_k \left(\left|2\mathscr{E}_c + N_{1k}\right|^2 - \left|N_{2k}\right|^2 \right) \right] \right\}, \tag{13.3.13}$$

где ν – величина, которая оптимизируется для получения наиболее плотной границы.

Усреднение (13.3.13) выполняется с учетом статистики шумовых компонент и статистики взвешенных коэффициентов $\{\beta_k\}$, которые являются случайными вследствие статистической природы интерференции. Сохраняя $\{\beta_k\}$ фиксированными и усредняя

сначала по статистике шума, мы получим

$$P_2(\boldsymbol{\beta}) = E\left[\exp\left(-\nu\sum_{k=1}^{L}\beta_k|2\mathscr{E}_c + N_{1k}|^2 + \nu\sum_{k=1}^{L}\beta_k|N_{2k}|^2\right)\right] =$$
$$= \prod_{k=1}^{L} E\left[\exp\left(-\nu\beta_k|2\mathscr{E}_c + N_{1k}|^2\right)\right] E\left[\left(\exp\nu\beta_k|N_{2k}|^2\right)\right] = \prod_{k=1}^{L}\frac{1}{1-4\nu^2}\exp\left(\frac{-4\mathscr{E}_c^2\beta_k\nu}{1+2\nu}\right) \qquad (13.3.14)$$

Так как частоты ЧМ поражаются помехой с вероятностью α, то $\beta_k = [2\mathscr{E}_c(N_0 + J_0/\alpha)]^{-1}$ с вероятностью α и $\beta_k = (2\mathscr{E}_c N_0)^{-1}$ с вероятностью $1-\alpha$. Следовательно, граница Чернова даст

$$P_2 \le \prod_{k=1}^{L}\left\{\frac{\alpha}{1-4\nu^2}\exp\left[\frac{-2\mathscr{E}_c\nu}{(N_0 + J_0/\alpha)(1+2\nu)}\right] + \frac{1-\alpha}{1-4\nu^2}\exp\left[\frac{-2\mathscr{E}_c\nu}{N_0(1+2\nu)}\right]\right\} =$$
$$= \left\{\frac{\alpha}{1-4\nu^2}\exp\left[\frac{-2\mathscr{E}_c\nu}{(N_0 + J_0/\alpha)(1+2\nu)}\right] + \frac{1-\alpha}{1-4\nu^2}\exp\left[\frac{-2\mathscr{E}_c\nu}{N_0(1+2\nu)}\right]\right\}^L. \qquad (13.3.15)$$

Следующий шаг заключается в оптимизации границы в (13.3.15) по величине ν. Однако в настоящей форме с границей манипулировать сложно. Достаточное упрощение возникает, если предположить, что $J_0/\alpha >> N_0$, что делает второе слагаемое в (13.3.15) пренебрежимо малым по сравнению с первым. Альтернативно мы предполагаем $N_0 = 0$, так что граница P_2 сокращается до

$$P_2 \le \left\{\frac{\alpha}{1-4\nu^2}\exp\left[\frac{-2\alpha\nu\mathscr{E}_c}{J_0(1+2\nu)}\right]\right\}^L. \qquad (13.3.16)$$

Легко видеть, что минимальное значение этой границы по ν и минимальное по α (исходный случай парциально-полосовой интерференции) возникает, когда $\alpha = 3J_0/\mathscr{E}_c \le 1$ и $\nu = \frac{1}{4}$. Для этих значений параметров (13.3.16) приводит к

$$P_2 \le P_2(L) = \left(\frac{4}{e\gamma_c}\right)^L = \left(\frac{1.47}{\gamma_c}\right)^L, \qquad \gamma_c = \frac{\mathscr{E}_c}{J_0} = \frac{\mathscr{E}_b}{LJ_0} \ge 3, \qquad (13.3.17)$$

где γ_c – ОСШ на чип в символьном чипе. Эквивалентно

$$P_2 \le \left[\frac{1{,}47(J_{ср}/P_{ср})}{W/R}\right]^L, \qquad \frac{W/R}{L(J_{ср}/P_{ср})} \ge 3. \qquad (13.3.18)$$

Результат (13.3.17) впервые был получен Витерби и Джекобсом (1975). Мы видели, что вероятность ошибки в наихудшем случае парциально-полосовой интерференции уменьшается экспоненциально с увеличением ОСШ на чип γ_c. Этот результат очень похож на характеристики качества техники разнесённого приема для каналов с релеевскими замираниями (см. раздел 14.4). Мы можем выразить правую часть (13.3.17) в виде

$$P_2(L) = \exp[-\gamma_b h(\gamma_c)], \qquad (13.3.19)$$

где функция $h(\gamma_c)$ определяется так:

$$h(\gamma_c) = -\frac{1}{\gamma_c}\left[\ln\left(\frac{4}{\gamma_c}\right) - 1\right]. \qquad (13.3.20)$$

Кривая $h(\gamma_c)$ дана на рис.13.3.5.

Видим, что функция имеет максимум $\frac{1}{4}$ в точке $\gamma_c = 4$. Следовательно, имеется

оптимальное ОСШ на чип, равное $10\lg\gamma_c = 6$дБ. При оптимальном ОСШ вероятность ошибки ограничена сверху так:

$$P_2 \le P_2(L_{\text{опт}}) = \mathrm{e}^{-\gamma_b/4}. \qquad (13.3.21)$$

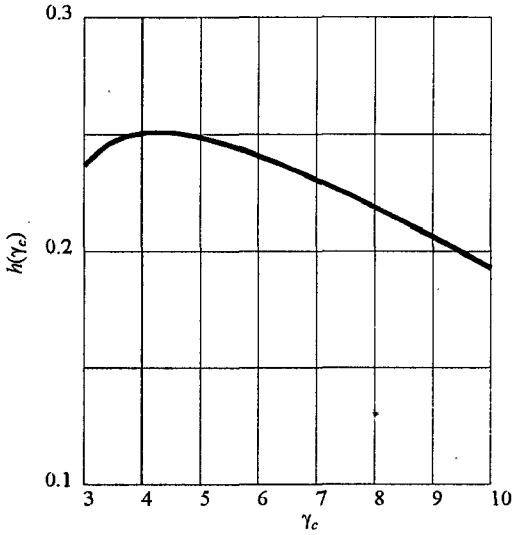


Рис. 13.3.5. График функции $h(\gamma_c)$

Если сравним границу вероятности ошибки (13.3.21) с вероятностью ошибки для двоичной ЧМ при равномерном спектре шума, определяемый (13.3.1), то видим, что эффект объединения наихудшего случая парциально – полосовой интерференции и потери некогерентного сложения при квадратичном сложении L чипов равно 3 дБ. Подчеркнем, однако, что для данного $\mathscr{E}_b / J_0$ потери больше, когда порядок разнесения выбран не оптимально. Кодирование дает средство улучшения качества системы по скачкам частоты, пораженные парциально- полосовой интерференцией. В частности, если используется блоковый ортогональный код с $M = 2^k$ кодовыми словами и разнесение L-го порядка кодового слова, вероятность ошибки кодового слова ограничена сверху так

$$P_M \le (2^k - 1)P_2(L) = (2^k - 1)\left(\frac{1{,}47}{\gamma_c}\right)^L = (2^k - 1)\left(\frac{1{,}47}{k\gamma_b / L}\right)^L, \qquad (13.3.22)$$

и эквивалентная вероятность ошибки на бит ограничена сверху так:

$$P_b \le 2^{k-1}\left(\frac{1{,}47}{k\gamma_b / L}\right)^L \qquad (13.3.23)$$

Рис.13.3.6 иллюстрирует вероятность ошибки на бит для $L = 1, 2, 4, 8$ и $k = 1, 3$.

При оптимальном выборе разнесения, верхнюю границу можно выразить так

$$P_b \le 2^{k-1}\exp\left(-\tfrac{1}{4}k\gamma_b\right) = \tfrac{1}{2}\exp\left[-k\left(\tfrac{1}{4}\gamma_b - \ln 2\right)\right]. \qquad (13.3.24)$$

Таким образом, мы имеем улучшение качества на величину, равную $10\lg\left[k\left(1 - 2{,}77/\gamma_b\right)\right]$. Для примера, если $\gamma_b = 10$ и $k = 3$ (восьмеричная модуляция) выигрыш равен 3,4 дБ, если же $k = 5$, то выигрыш 5,6 дБ.

Дополнительный выигрыш можно достичь использованием каскадных кодов в соединении с декодированием мягких решений. В нижеследующем примере мы используем k-дуальный свёрточный код как внешний код и код Адамара как внутренний код в канале с парциально-полосовой интерференцией.

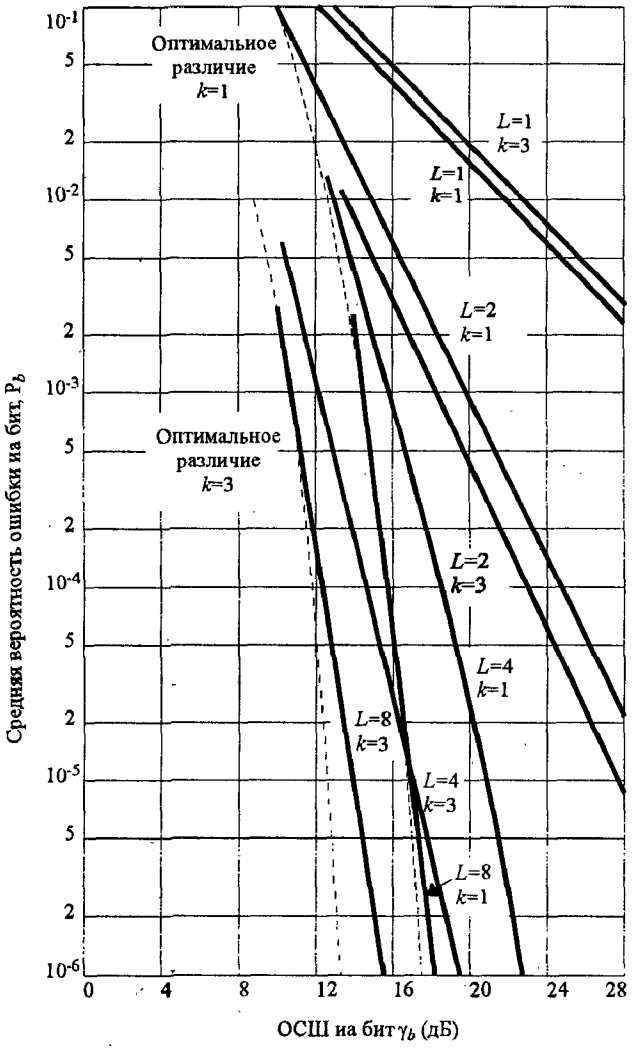


Рис. 13.3.6. Характеристики двоичной и восьмеричной ФМ для канала с худшим случаем интерференции

Пример 13.3.1. Допустим, что мы используем код Адамара $H(n,k)$ с постоянным весом вместе с амплитудной манипуляцией (OOK – on-off keying) для каждого кодового символа. Минимальное расстояние кода $d_{\min} = \frac{1}{2}n$ и, следовательно, эффективный порядок разнесения с OOK модуляцией $\frac{1}{2}d_{\min} = \frac{1}{4}n$. Имеется $\frac{1}{2}n$ тонов со скачками частоты на кодовое слово. Следовательно,

$$\gamma_c = \frac{k}{\frac{1}{2}n}\gamma_b = 2R_c\gamma_b, \tag{13.3.25}$$

если этот код используется один. Вероятность ошибки на бит для этого кода при декодировании мягких решений и наличии парциально-полосовой интерференции имеет верхнюю границу

$$P_b \le 2^{k-1}P_2(\tfrac{1}{2}d_{\min}) = 2^{k-1}\left(\frac{1{,}47}{2R_c\gamma_b}\right)^{n/4}. \tag{13.3.26}$$

Теперь, если используется код Адамара (n,k) как внутренний код и k-дуальный свёрточный код со скоростью 1/2 (см. раздел 8.2.6) как внешний код, вероятность ошибки на бит при наихудшем случае парциально-полосовой интерференции равна (см. (8.2.40)):

$$P_b \le \frac{2^{k-1}}{2^k-1}\sum_{m=4}^{\infty}\beta_m P_2\left(\tfrac{1}{2}md_{\min}\right) = \frac{2^{k-1}}{2^k-1}\sum_{m=4}^{\infty}\beta_m P_2\left(\tfrac{1}{4}mn\right), \qquad (13.3.27)$$

где $P_2(L)$ определяется (13.3.17) с

$$\gamma_c = \frac{k}{n}\gamma_b = R_c\gamma_b . \qquad (13.3.28)$$

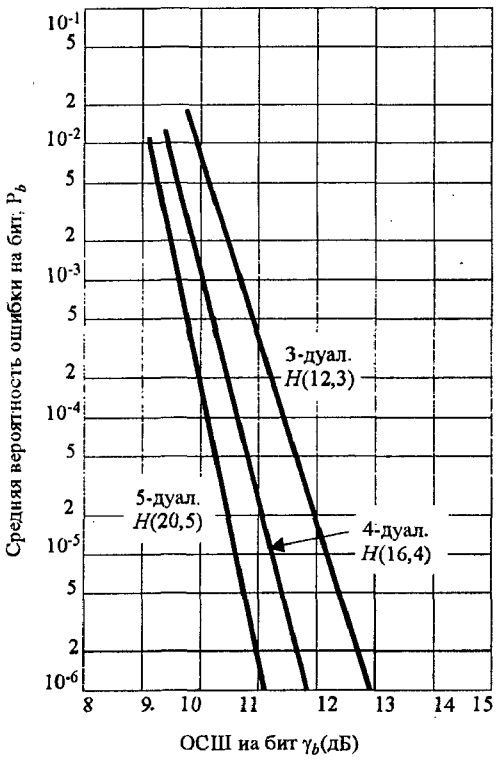


Рис. 13.3.7. Качество k-дуального кода для k=5, 4 и 3, каскадно соединённого с кодом Адамара, для канала с худшим случаем интерференции

Рис.13.3.7 иллюстрирует качество k-дуального кода для k=5, 4 и 3, каскадно соединённого с кодом Адамара $H(20,5)$, $H(16,4)$ и $H(12,3)$ соответственно.

В приведенном обобщении, мы сосредоточились на декодировании мягких решений. С другой стороны, качество, достигаемое декодированием жёстких решений, существенно (на несколько децибел) хуже, чем то, которое достигается при декодировании мягких решений. В каскадной схеме, однако, декодирование мягких решений для внутреннего кода и декодирование жёстких решений для внешнего представляет разумный компромисс между сложностью декодирования и качеством.

В заключение мы хотим указать, что другой серьёзной помехой в СЧ системе с рассеянным спектром является парциально-полосовая многотоновая помеха.

По своему воздействию этот вид интерференции подобен парциально-полосовой помехе с неизменной спектральной плотностью. Разнесение, обеспечиваемое кодированием, является эффективным средством для улучшения качества СЧ систем.

Дополнительное улучшение достигается путем надлежащего взвешивания выхода демодулятора так, чтобы подавить влияние мешающего сигнала.

13.3.3. CDMA система, основанная на широкополосных сигналах с СЧ

В разделе 13.2.2 мы рассмотрели CDMA систему, основанную на использовании широкополосных ПП сигналов. Как указано выше, можно также иметь CDMA систему, основанную на широкополосных СЧ сигналах с рассеянным спектром. Каждая пара передатчик–приёмник в такой системе рассчитана на свой собственный шаблон псевдослучайных скачков частоты. Исключая эти различия, передатчики и приёмники всех пользователей могут иметь идентичные кодеры, декодеры, модуляторы и демодуляторы. CDMA системы, основанные на широкополосных СЧ сигналах привлекательны, в частности, для подвижных (земных, воздушных, морских) пользователей, поскольку требования синхронизации не так строги, как в широкополосных ПП системах. Кроме того, техника синтеза частот и соответствующее оборудование СЧ систем детально проработаны. Следовательно, при СЧ возможны большие выигрыши обработки. Пропускная способность CDMA с СЧ также относительно велика. Витерби (1978) показал, что с k-дуальными кодами и многоступенчатой ЧМ возможно обеспечить до $\frac{3}{8}W/R$

одновременно работающих пользователей, которые передают информацию со скоростью R бит/с по каналу с полосой W.

Одна из ранних CDMA систем, основанная на использовании кодированных широкополосных сигналов с СЧ, была построена так, чтобы обеспечить множественный доступ по тактическим спутникам связи для малоподвижных (земных, воздушных, морских) терминалов, каждый из которых передаёт относительно короткие сообщения по каналу с перемежением. Систему назвали тактической системой передачи (TATS), и она описана в статье Дроуилхета и Бернштейна (1969). В системе TATS использован восьмеричный код Рида–Соломона (7, 2). Таким образом, два трехбитовых информационных символа на входе кодера используется для генерирования семисимвольных кодовых слова. Каждый трехбитовый закодированный символ передается посредством восьмеричной ЧМ. Восемь различных частот передаются с разносом $1/T_c$ Гц, где T_c – длительность передачи на одной частоте. В дополнение к семи символам в кодовое слово включён восьмой символ. Этот символ и его соответствующая частота фиксированы. Он передаётся в начале каждого кодового слова для обеспечения тактовой и частотной синхронизации[1] на приеме. Следовательно, каждое кодовое слово передается за $8T_c$ секунд.

Система TATS была спроектирована для информационной скорости передачи 75 и 2400 бит/с. Значит, $T_c = 10$ мс и 312,5 мкс соответственно. Каждая тональная частота, соответствующая кодовому символу, подвергается скачкам. Таким образом, скорость скачков равна $100\ \text{с}^{-1}$ при информационной скорости 75 бит/с и $3200\ \text{с}^{-1}$ при информационной скорости 2400 бит/с.

В коде Рида–Соломона (7, 2) имеется $M = 2^6 = 64$ кодовых слов, а минимальное расстояние этого кода $d_{\min} = 6$. Это значит, что код обеспечивает эффективный порядок разнесения, равный 6.

На приёме сигнал сначала освобождается от скачков и затем демодулируется пропусканием через параллельный блок из восьми согласованных фильтров, где каждый фильтр настроен на одну из восьми возможных частот – выход каждого фильтра детектируется по огибающей, квантуется 4 битами (один из 16 уровней) и подаётся на декодер. Декодер берёт 56 выходов фильтров, соответствующие приему каждого семисимвольного кодового слова и формирует 64 величин для решения, соответствующих 64 возможным кодовым словам кода (7, 2) путём линейного сложения соответствующих огибающих на выходах детекторов. Решение выносится в пользу кодового слова, имеющего наибольшую величину для решения.

При квантовании выходов согласованных фильтров 16 уровнями интерференция от других пользователей канала вызывает относительно малые потери в качестве (0,75 дБ при сильной интерференции на один чип и 1,5 дБ при сильной интерференции на два чипа из семи) АРУ, используемая в системе TATS, имеет постоянную времени большую, чем интервал чипа T_c, так что нетрудно реализовать оптимальное взвешивания по выходам демодуляторов, как описано в разделе 13.3.2.

Расчёт вероятности ошибки сигналов TATS в канале с АБГШ при наихудшем случае парциально – полосовой интерференции оставляем для упражнения читателю (задачи 13.23 и 13.24).

---

[1] Поскольку вовлечены подвижные пользователи, имеются доплеровские сдвиги частоты, связанные с передачей. Эти частотные сдвиги должны быть отслежены и скомпенсированы для демодуляции сигнала. Синхронизирующий символ используется и для этих целей.

13.4. ДРУГИЕ ТИПЫ ШИРОКОПОЛОСНЫХ СИГНАЛОВ

ПП и СЧ являются наиболее общими формами широкополосных сигналов, используемых на практике. Однако можно использовать другие методы для введения псевдослучайности в широкополосных сигналах. Один метод, который дуален по отношению к СЧ – скачки времени (СВ). В системах с СВ интервал времени, который выбирается намного больше, чем величина, обратная информационной скорости, делится на большое число временных участков (слотов).

Кодированные информационные символы передаются при псевдослучайном выборе временных участков в виде блока одного или больше кодовых слов. Для передачи кодовых символов используется ФМ.

Для примера предположим, что интервал времени T подразделяется на 1000 временных участков шириной $T/1000$. При информационной скорости R бит/с число бит, которые надо предать за время T, равно RT. Кодирование увеличивает это число до RT/R_c бит, где R_c – скорость кода. Следовательно, на временном интервале $T/1000$ с мы должны передать RT/R_c бит. Если используется двоичная ФМ, битовая скорость равна $1000R/R_c$ и требуемая полоса примерно равна $W = 1000R/R_c$.

Блок-схема передатчика и приёмника для широкополосной системы со скачками времени (СВ) показана на рис.13.4.1.

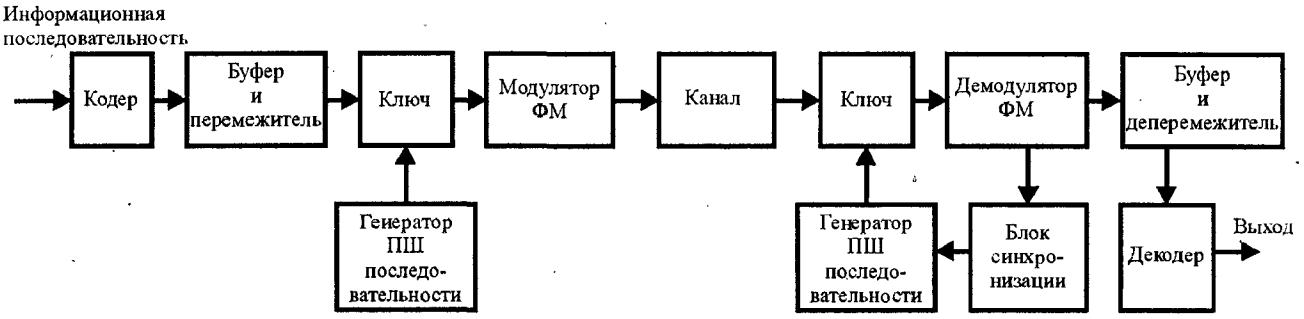


Рис. 13.4.1. Блок-схема широкополосной системы с о скачками времени (СВ)

С учетом разрывных характеристик передаваемого сигнала, в СВ системе должна быть на передаче буферная память. Буфер может также использоваться на приёме для обеспечения равномерного потока данных к пользователю.

Так же, как парциально-полосовая интерференция искажает не кодированную широкополосную систему с СЧ, парциально-временная интерференция оказывает схожее воздействие на широкополосную систему с СВ. Кодирование и перемежение являются эффективными средствами борьбы с этим видом интерференции, как мы уже показали для ПП системы с СЧ. Наибольшая неприятность для СВ системы – это жёсткие требования к синхронизации не только по сравнению с СЧ, но и с ПП.

Другие типы широкополосных сигналов можно получить комбинированием ПП, СЧ и СВ. Например, мы можем иметь гибрид ПП/СЧ, что означает, что ПШ последовательность используется вместе со скачками частоты. Сигнал, посылаемый при отдельном скачке, состоит из широкополосных ПП сигналов, которые демодулируются когерентно. Однако принимаемые сигналы от различных скачков суммируются некогерентно (по огибающим или квадратично). Поскольку когерентное детектирование формируется внутри скачка, имеется выигрыш относительно чистой системы ПП. Однако цена этого выигрыша – увеличение сложности, большая стоимость и более строгие требования к синхронизации.

13.5. СИНХРОНИЗАЦИЯ ШИРОКОПОЛОСНЫХ СИСТЕМ

Временную синхронизацию на приёме относительно принимаемого широкополосного сигнала можно разделить на две фазы. Имеется начальная фаза обнаружения сигнала и фаза отслеживания после того, как сигнал начально обнаружен.

Обнаружение. В широкополосных системах с ПП ПШ код должен быть синхронизирован во времени с точностью до небольшой части интервала чипа $T_c \approx 1/W$. Проблему начальной синхронизации можно рассматривать как задачу синхронизации часов приёмника относительно часов передатчика. Обычно в широкополосных системах используются точные и стабильные во времени часы. Следовательно, точное сопряжение во времени часов сводится к снижению временной неопределённости между передатчиком и приёмником. Однако всегда имеется начальная неопределённость времени, обусловленная неопределённостью расстояния между передатчиком и приёмником. Эта проблема становится особенно острой, когда устанавливается связь между подвижными пользователями. Обычная процедура для обеспечения начальной синхронизации – это посылка передатчиком известной псевдослучайной последовательности данных к приёмнику. Приёмник непрерывно находится в режиме поиска, ожидая эту последовательность для обеспечения начальной синхронизации.

Предположим, что начальная неопределённость во времени равна T_u, а длительность чипа T_c. Если начальная синхронизация имеет место в присутствии аддитивного шума и другой интерференции, необходимо исследовать интервал $T_d = NT_c$ для того, чтобы выбрать в нём момент синхронизации. При грубом поиске с шагом $\frac{1}{2}T_c$ время, требуемое для достижения начальной синхронизации, равно

$$T_{\text{синхр}} = \frac{T_u}{\frac{1}{2}T_c} NT_c = 2NT_u. \qquad (13.5.1)$$

Ясно, что синхронизационная последовательность, переданная к приёмнику, должна быть по крайней мере длительностью $2NT_c$ для того, чтобы приёмник имел достаточно времени для обеспечения необходимого поиска в последовательном режиме.

В принципе, оптимальными методами для обеспечения начальной синхронизации являются согласованная фильтрация или взаимная корреляция. Фильтр согласуется с известной последовательностью данных, генерируемой известной ПШ последовательностью, и непрерывно следит за превышением предварительно установленного порога. Когда это происходит, начальная синхронизация достигнута и демодулятор входит в режим «приёма данных».

Альтернативно, мы можем использовать коррелятор, как показано на рис.13.5.1.

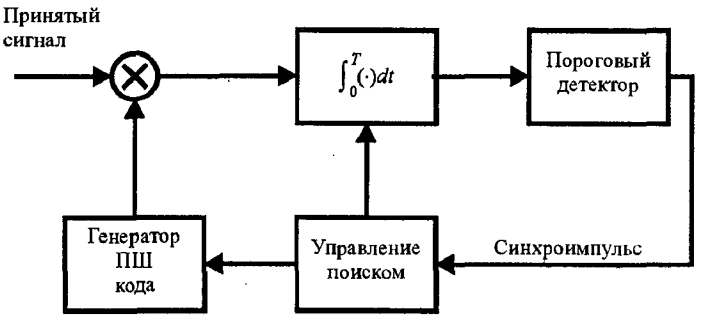


Рис. 13.5.1. Скользящий коррелятор для обнаружения ПШ сигнала

Коррелирование производится по интервалу времени NT_c (N чипов) и выход коррелятора сравнивается с порогом, чтобы определить, присутствует ли в принимаемом сигнале известная сигнальная последовательность. Если порог не превышается, известная опорная последовательность продвигается во времени на $\frac{1}{2}T_c$ секунд и процесс коррелирования повторяется. Эти операции выполняются до тех пор, пока сигнал не обнаружится, или до тех пор, пока поиск не выйдет за пределы интервала неопределенности T_u. В последнем случае процесс поиска возобновляется.

Аналогичный процесс можно также использовать для СЧ сигналов. В этом случае проблема сводится к синхронизации ПШ кода, который управляет скачками частоты. Чтобы выполнить эту начальную синхронизацию, известный сигнал СЧ передаётся приёмнику. Система начального обнаружения на приёме следит за этим известным сигналом СЧ. Для примера, можно использовать набор согласованных фильтров, настроенных на известные образцы передаваемых частот. Их выходы должны быть правильно сдвинуты, продетектированы по огибающей (или по закону квадратов), взвешены, если необходимо, и просуммированы (некогерентное интегрирование) для образования выходного сигнала, который сравнивается с порогом. Сигнал считается присутствующим, если превышается порог. Процесс поиска обычно производится непрерывно во времени, пока порог не будет превышен. Блок-схема, иллюстрирующая обнаружение сигнала, дана на рис.13.5.2.

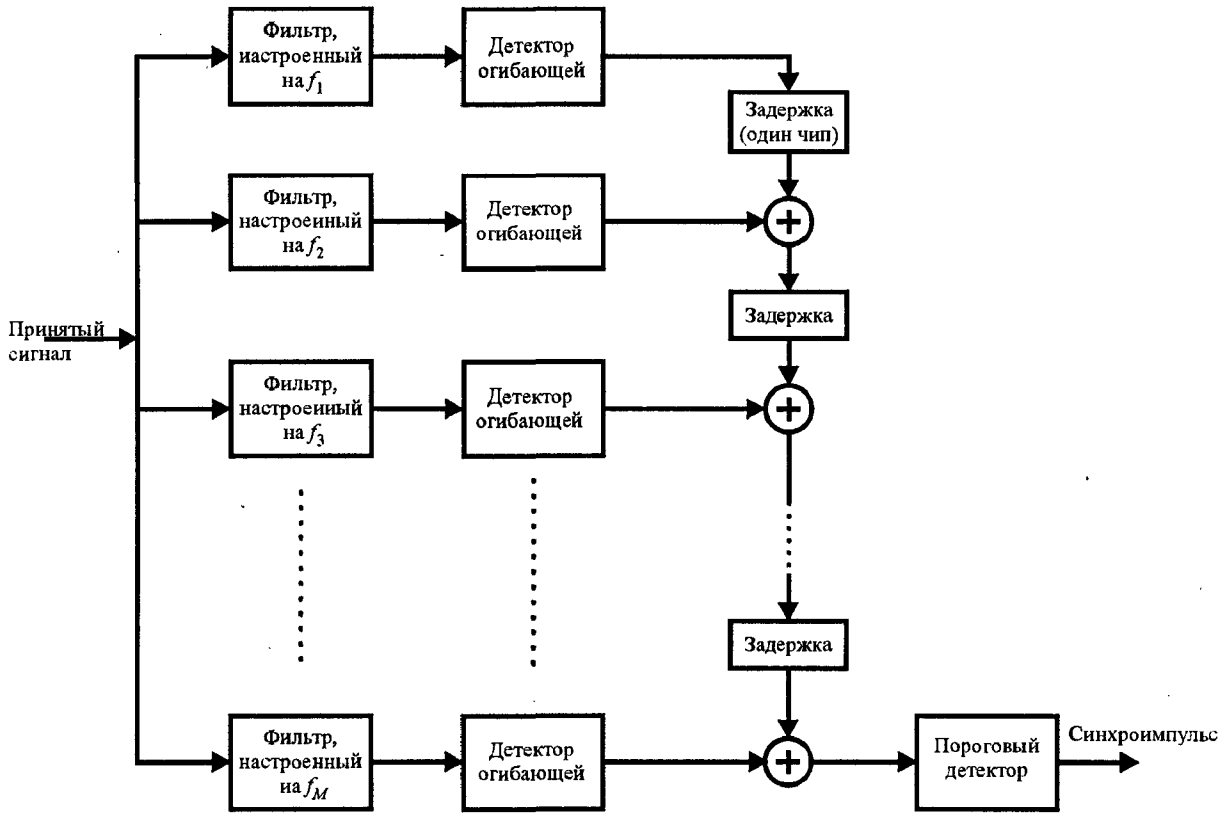


Рис. 13.5.2. Система для обнаружения СЧ сигнала

В качестве альтернативы, можно использовать простую пару блоков согласованный фильтр–детектор огибающей. Перед ней генератор образца скачка частоты, а после неё – последовательный интегратор и пороговый детектор. Эта конфигурация, показанная на

рис.13.5.3, основана на последовательном поиске и подобна скользящему коррелятору для широкополосных ПП сигналов.

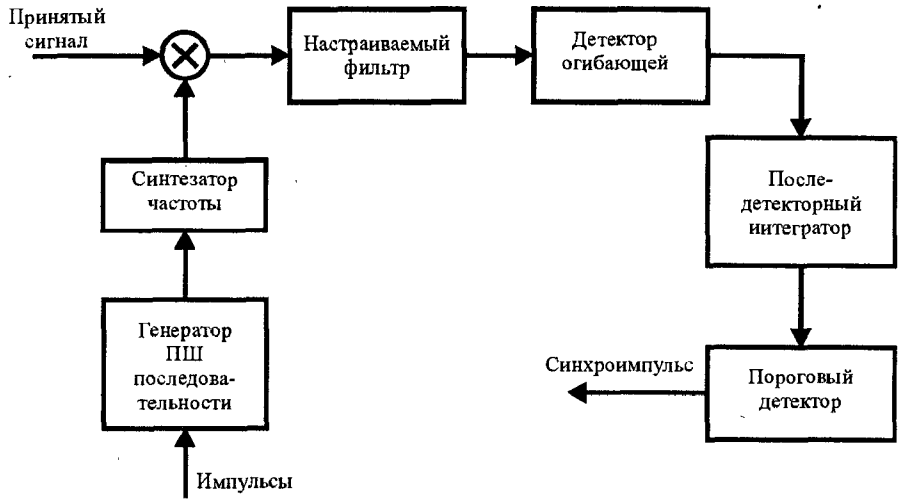


Рис. 13.5.3. Альтернативная система для обнаружения сигнала с СЧ

Скользящий коррелятор для ПП сигналов и схема по рис. 13.5.3 для СЧ сигналов осуществляют последовательный поиск, который является расточителем времени. В качестве альтернативы, можно ввести определённую степень параллелизма, имея два или больше таких корреляторов, работающих параллельно, и каждый из них ищет по неперекрывающимся участкам времени. В таком случае сокращается время поиска в обмен на большую сложность и стоимость внедрения. Рисунок 13.5.2 представляет такую параллельную реализацию для СЧ сигналов.

На протяжении поиска могут быть ложные тревоги, которые возникают с заданной вероятностью ложной тревоги, предусмотренной при проектировании. Чтобы справиться с ними, необходимо иметь дополнительный механизм, подтверждающий, что принятый сигнал на выходе коррелятора остаётся выше порога. С такой стратегией детектирования большой импульс шума, который вызовет ложную тревогу, будет вызывать только временное превышение порога. С другой стороны, если пороговое устройство сработало из-за обнаруженного сигнала, сигнал на выходе коррелятора или согласованного фильтра будет оставаться выше порога на протяжении всего принятого сигнала. Таким образом, если подтверждение отсутствует, поиск возобновляется.

Другая стратегия начального поиска, названная *последовательным поиском*, была разработана Вардом (1965–1977). В этом методе время пребывания по каждой задержке в процессе поиска делается переменным путем использования коррелятора с переменным временем интегрирования, выход которого сравнивается с двумя порогами. Таким образом, здесь имеются три возможных решения:

1. Если верхний уровень превышается выходом коррелятора, начальная синхронизация объявляется установленной;
2. Если выход коррелятора находится ниже нижнего порога, сигнал объявляется отсутствующим при этой задержке и процесс поиска продолжается при другой задержке;
3. Если выход коррелятора находится между двумя порогами, время интегрирования увеличивается на один чип и результирующий выход снова сравнивается с двумя порогами.

Далее, шаги 1, 2 и 3 повторяются на каждом чиповом интервале до тех пор, пока выход коррелятора или превысит верхний порог или окажется ниже нижнего порога.

Метод последовательного поиска относится к классу методов последовательного оценивания, предложенного Вальдом (1947), который, как известно, даёт более эффективный поиск в том смысле, что среднее время поиска минимизируется. Таким образом, время поиска при последовательном поиске меньше, чем при фиксированном времени пребывания интегратора на каждом чипе.

В приведённом выше обсуждении мы предполагали только неопределённость во времени при осуществлении начальной синхронизации. Однако другой аспект начальной синхронизации – это неопределённость по частоте. Если передатчик или приёмник подвижны, относительная скорость между ними порождает доплеровский сдвиг частоты в принимаемом сигнале относительно переданного сигнала. Поскольку приёмник обычно не знает относительную скорость априори, доплеровский сдвиг частоты неизвестен и должен быть определен посредством метода поиска частоты. Такой поиск обычно выполняется параллельно по подходящим квантованным частотам на интервале неопределённости и последовательно по интервалу временной неопределённости. Блок-схема соответствующего устройства показана на рис.13.5.4. Подходящие методы поиска доплеровских сдвигов частоты можно также разработать для СЧ сигналов.

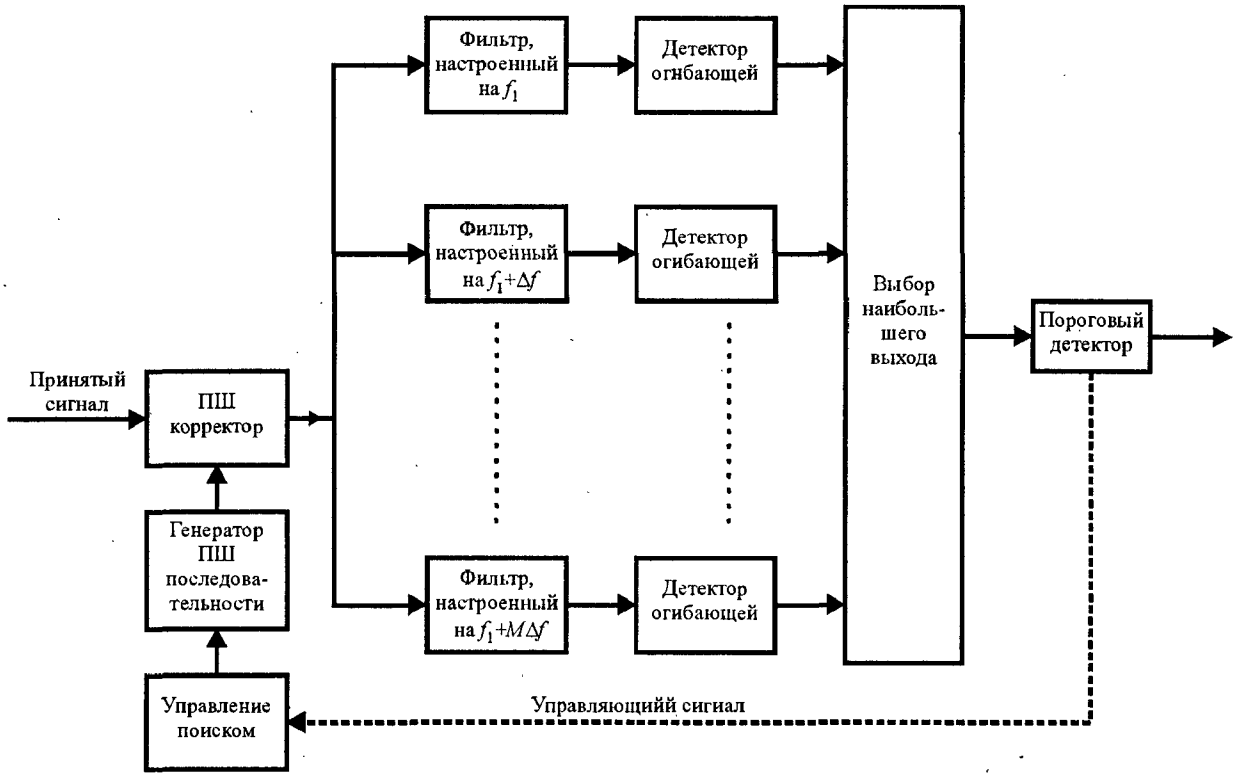


Рис. 13.5.4. Начальный поиск сигнала в широкополосных состемах с доплеровским сдвигом частоты

Отслеживание. Когда сигнал обнаружен, первоначальный процесс поиска останавливается и начинается процесс точной синхронизации и отслеживания. Отслеживание поддерживает ПШ кодовый генератор приёмника в синхронизме с приходящим сигналом. Отслеживание включает точную синхронизацию чипа и, при когерентной демодуляции, отслеживание фазы несущей.

Обычно используемая схема отслеживания для широкополосных ПП сигналов – петля с задержкой (ПЗ, DLL), показанная на рис.13.5.5.

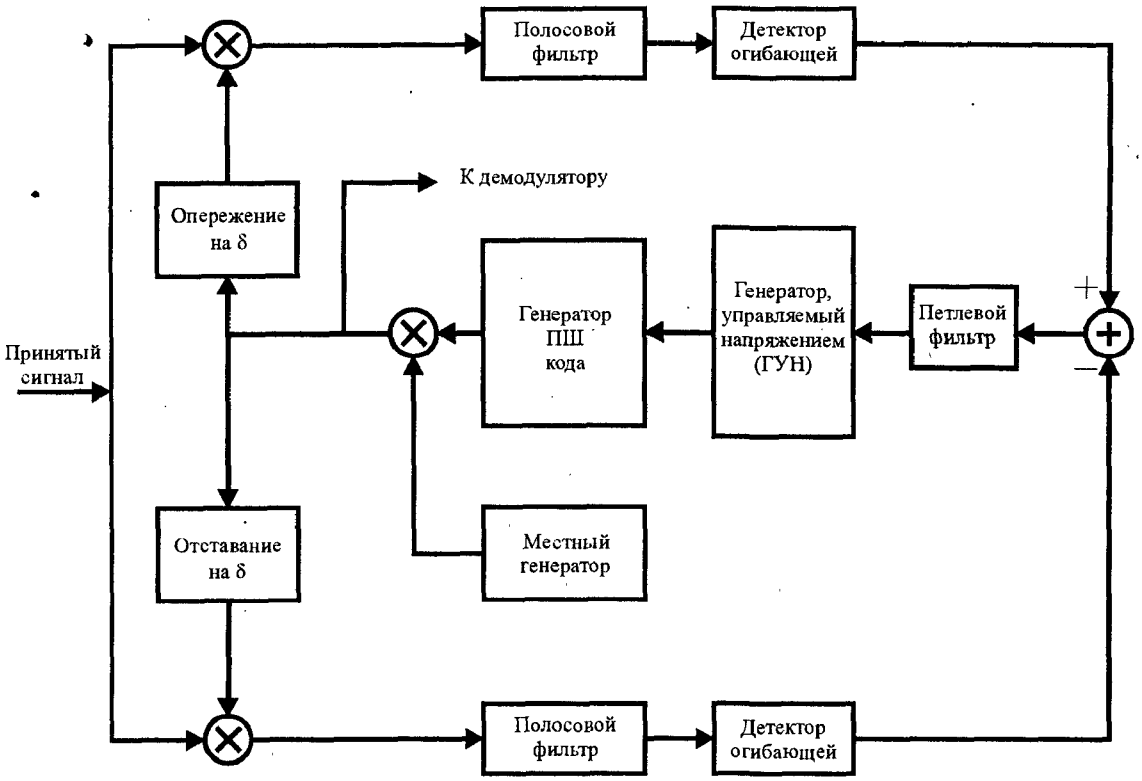


Рис. 13.5.5. Петля с задержкой для слежения за ПШ кодом

В этой петле отслеживания принимаемый сигнал подаётся на два умножителя, где он умножается на два выхода местного ПШ кодового генератора, которые сдвинуты во времени друг относительно друга на величину $2\delta \le T_c$. Сигналы произведения определяют взаимную корреляцию между принимаемым сигналом и ПШ последовательностью при двух значениях задержки. Эти произведения проходят полосовой фильтр и затем детектируются по огибающей и затем вычитаются.

Разностный сигнал подаётся к петлевому фильтру, который управляет ГУН (генератором, управляемым напряжением). ГУН служит в качестве часов для ПШ кодового генератора сигнала.

Если синхронизация неточная, отфильтрованный выход одного коррелятора будет больше другого, и ГУН будет соответственно ускорен или приторможен.

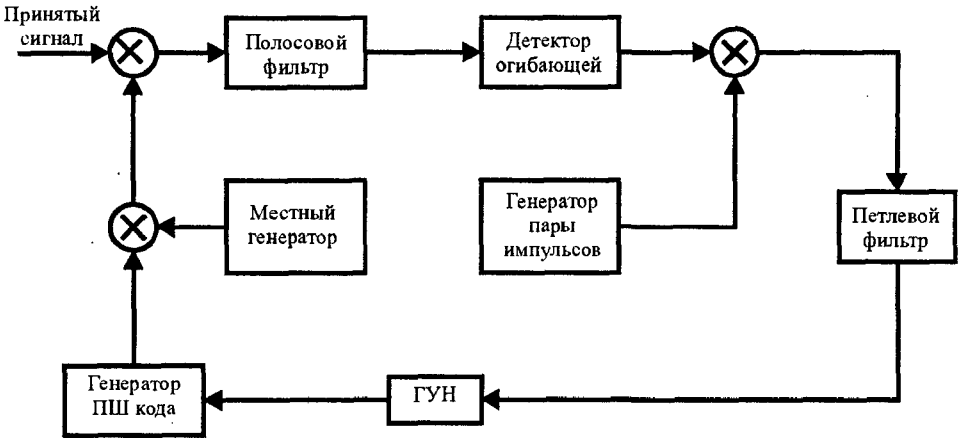
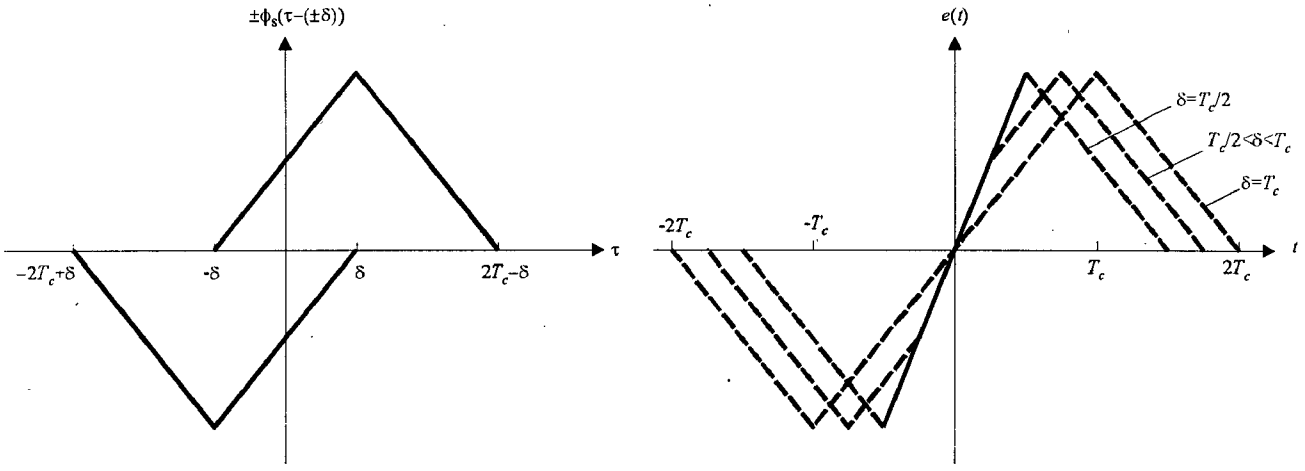


Рис. 13.5.6. Петля с временной "вилкой" для слежения за ПШ кодом

В точке равновесия выходы двух фильтрующих корреляторов будут одинаково смещены относительно пикового значения, и выход ПШ кодового генератора будет точно засинхронизирован с принимаемым сигналом, который поступает на демодулятор. Мы видим, что такая реализация ПЗ для отслеживания ПП сигнала эквивалентна синхронизатору с окнами на задержку-опережение, ранее обсуждавшемуся в разделе 6.3.2 и показанному на рис. 6.3.5.

Альтернативный метод для отслеживания во времени ПП сигнала сводится к использованию петли с временной вилкой (ПВВ), показанной на рис.13.5.6. ПВВ использует одно «плечо» вместо двух «плеч», показанных на рис.13.5.5.



(*a*) Функция автокорреляции (*b*) Сигнал ошибки слежения

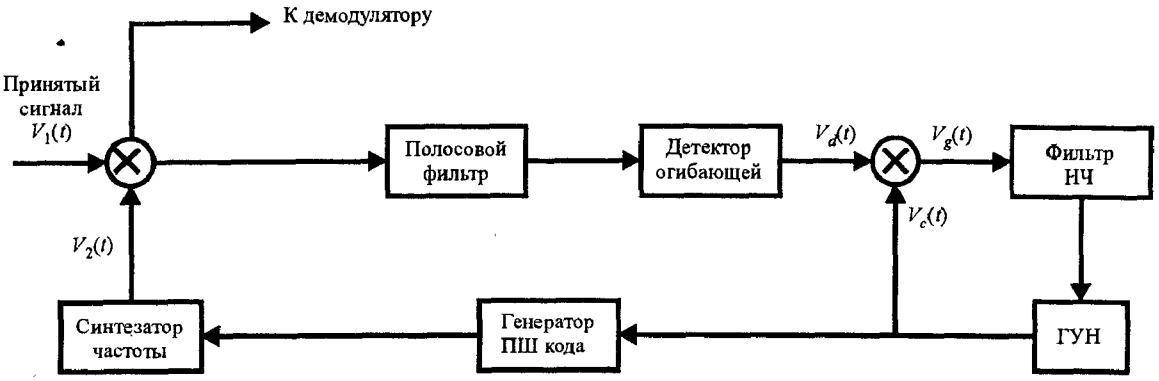
Рис. 13.5.7. Функция автокорреляции и сигнал ошибки слежения для петли с задержкой

Обеспечивая подходящий сигнал «вилки», возможно сделать эту «одноплечевую» реализацию эквивалентной «двухплечевой». В этом случае берутся два отсчёта взаимной корреляции: один с опережением, другой с отставанием – и вычитаются один из другого. В результате вырабатывается сигнал ошибки отслеживания. Важное достоинство ПВВ – простота реализации, обусловленная использованием одного «плеча» вместо двух, как в обычной ПЗ. Второе достоинство – это то, что оба вычитаемых отсчёта одинаково масштабированы и схема ПВВ точно сбалансирована. ПЗ (и её эквивалент ПВВ) выдают сигнал ошибки при стробировании функции корреляции сигналов с отклонением $\pm\delta$ от пика, как показано на рис.13.5.7(a). Сигнал ошибки показан на рис.13.5.7(b). Анализ качества ПЗ выполняется аналогично тому, как это было сделано в разд. 6.3 для ФАП. Если исключить вопрос о детектировании огибающих в двух плечах ПЗ, она похожа на петлю Костаса. Дисперсия ошибки оценки времени в ПЗ обратно пропорциональна ОСШ на выходе петли, которое, в свою очередь, зависит от ОСШ на входе петли и её полосы пропускания. Её качество несколько снижается при квадратичном детектировании.

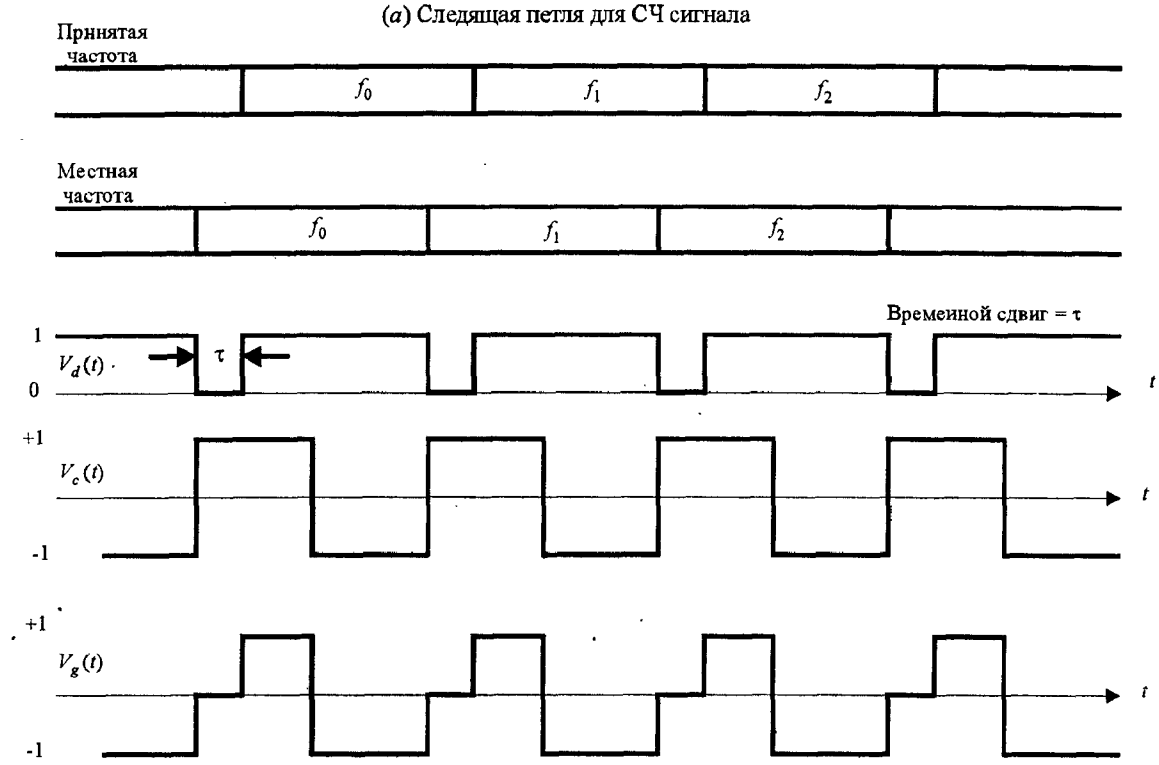
Типичная техника отслеживания иллюстрируется на рис.13.5.8(a).

Этот метод основан на предположении, что достигнуто начальное обнаружение, но имеется небольшая ошибка синхронизации. Полосовой фильтр настроен на единственную промежуточную частоту, и его полоса пропускания имеет порядок $1/T_c$, где T_c – длительность чипа. Сигнал с выхода фильтра детектируется по огибающей и умножается на сигнал задающего генератора (ГУН) для получения трёхуровневого сигнала, который поступает на вход петлевого фильтра и далее на управляющий вход ГУН, как показано на рис.13.5.8(b). Заметим, что если переходы чипов от местного генератора не совпадают во времени с аналогичными переходами во входном сигнале, то на выходе петлевого фильтра

будет либо положительный, либо отрицательный сигнал, в соответствии со знаком временного сдвига. Этот сигнал ошибки с выхода петлевого фильтра будет изменять частоту и фазу ГУН в направлении, обратном имеющейся ошибке синхронизма.



(*a*) Следящая петля для СЧ сигнала



(*b*) Динамика отслеживания СЧ сигнала

Рис. 13.5.8. Метод отслеживания для СЧ сигнала [*Pickholtz* и др.(1982), © 1982 *IEEE*]

13.6. БИБЛИОГРАФИЧЕСКИЕ ЗАМЕЧАНИЯ И ССЫЛКИ

В этом разделе дана краткая вводная трактовка вопросов широкополосных сигналов и достигаемое с ними качество в системах связи. Детали и более специальная трактовка техники обнаружения сигналов, методов отслеживания кодов и гибридных широкополосных систем, как и других общих тем по широкополосным сигналам и системам можно найти в огромном множестве технической литературы, которая ныне существует по этому предмету.

Исторически первая публикация по широкополосным системам связи были в области защиты цифровых систем связи для военного использования. Действительно, до 1970 г. большинство работ в области расчёта и развития широкополосных систем были

засекречены[1]. С тех пор это положение изменилось. Открытая литература сегодня содержит бесчисленные публикации по всем аспектам анализа и синтеза широкополосных сигналов. Более того, мы имеем недавние публикации, касающиеся применений техники широкополосных сигналов для коммерческой связи, такие как межофисная радиосвязь (см. Пахлаван, 1985) и радиосвязь для мобильных пользователей (см. Ю, 1983).

Исторический обзор по разработке широкополосных систем связи за период 1920–1960 гг. дан в статье Шольца. (1982). Учебные трактовки с фокусированием на базовых принципах можно найти в статьях Шольца (1977) и Пикхольца и др. (1982). Эти статьи содержат также большое число ссылок на прежние работы. В дополнение, имеется две статьи Витерби (1979, 1975), которые содержат базовый обзор характеристик качества сигнальной техники ПП и СЧ.

Исчерпывающие обсуждения различных аспектов анализа и синтеза широкополосных сигналов и систем, включая техники синхронизации теперь имеются в распоряжении в публикациях Саймона и др. (1975), Зимера и Питерсона (1985) и Холмса (1982). В дополнение к этим публикациям имеется несколько специальных исследований в *IEEE Transactions on Communications*, посвящённые системам связи с широкополосными сигналами (август 1977 и май 1982) и *IEE Transactions on Selected Areas in Communication* (сентябрь 1985, май 1989, май 1990, июнь 1993). Эти исследования содержат набор статей, посвящённых разным темам, включая технику множественного доступа, технику синхронизации и анализ качества при различных видах интерференции.

Определённое число важных статей, которые были опубликованы в IEEE, также были перепечатаны в книге, выпущенной IEEE Press (Диксон, 1976; Кук и др., 1983).

В заключение, мы рекомендуем читателю, который желает глубже вникнуть в этот предмет, книгу Голомба (1967) как базовую ссылку по последовательностям сдвигового регистра.

ЗАДАЧИ

13.1. Следуя процедуре, изложениой в примере 13.2.2, определите вероятность ошибки широкополосной ПП системы в присутствии гармонической помехи на несущей, если сигнальный импульс определяется так

$$g(t) = \sqrt{\frac{16\mathscr{E}_c}{3T_c}} \cos^2\left[\frac{\pi}{T_c}\left(t - \tfrac{1}{2}T_c\right)\right], \quad 0 \le t \le T.$$

13.2. Изображение на рис. Р.13.2 иллюстрирует спектральные плотности мощности широкополосного ПШ сигнала и узкополосной интерференции при не кодированной (тривиальный код с повтореннем) цифровой системы связи. Ссылаясь на рис.13.2.6, который показывает демодулятор для этого сигнала, нарисуйте (приближённо) спектральные характеристики сигнала и интерференции после умножения $r(t)$ на выходной сигнал ПШ генератора. Определите долю общей интерференции, которая имеется на выходе коррелятора, когда число ПШ чипов на бит равно L.

13.3. Рассмотрите каскадирование кода Рида–Соломона (31,3) (q = 32-ичный алфавит) как внешнего кода с двоичным кодом Адамара (16,5) как внутреннего кода в широкополосной ПП системе. Предположите, что осуществляется декодирование мягких решений по обоим кодам. Определите верхнюю (объединённую) границу для вероятности ошибки на бит, основываясь на минимальном расстоянии каскадного кода.

13.4. Код Адамара $(n,k) = (2^m,\ m+1)$ является низкоскоростным кодом с $d_{\min} = 2^{n-1}$. Определите качество этого класса кодов для широкополосных ПП сигналов при двоичной ФМ и декодировании мягких либо жёстких решений.

---

[1] Эффект сжатия широкополосных сигналов в согласованном фильтре был впервые отмечен в 1956 г. Я.Д. Ширманом [60]. Общие методы синтеза широкополосных сигналов разрабатывались, начиная с 1965 г., Д.Е. Вакманом [61, 62], Л.Е. Варакиным [63, 64] и М.Б. Свердликом [65] (прп).

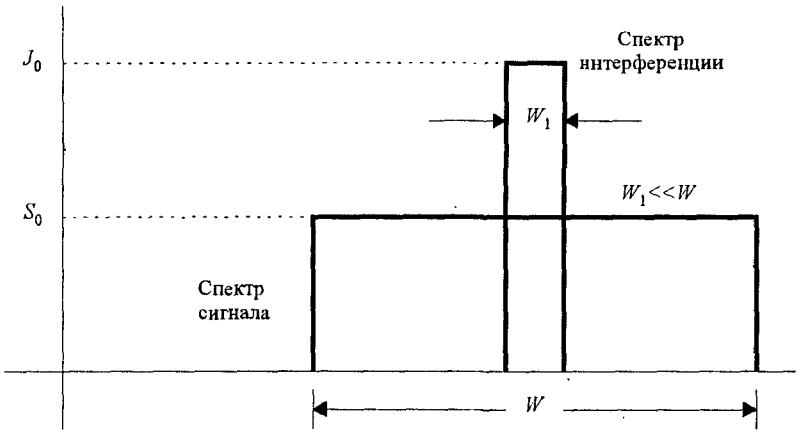


Рис. P13.2

13.5. Свёрточный код со скоростью 1/2 и $d_{св}=10$ используется для кодирования последовательности данных, возникающая со скоростью 1000 бит/с. Используется двоичная ФМ. Широкополосная ПП последовательность имеет скорость образования чипов 10 МГц.

a) Определите выигрыш кодирования.

b) Определите выигрыш обработки.

c) Определите помехозащищённость, предполагая, что $\mathscr{E}_b / J_0 = 10$.

13.6. 30 пользователей с равными мощностями сигналов занимают общий канал связи СДМА. Каждый пользователь передаёт информацию со скоростью 10 кбит/с посредством широкополосной ПП системы и двоичной ФМ. Определите минимальную скорость чипов для получения вероятности ошибки на бит 10^{-5}. При расчёте можно игнорировать аддитивные шумы на приёме.

13.7. Система СДМА спроектирована на основе широкополосных ПП сигналов с выигрышем обработки 1000 и с использованием двоичной ФМ. Определите число пользователей, если каждый пользователь имеет равную мощность и желательный уровень качества вероятность ошибки 10^{-6}. Повторите расчёт для случая, когда выигрыш обработки изменится на 500.

13.8. Широкополосная ПП система передаёт со скоростью 1000 бит/с в присутствии гармонической помехи. Мощность помехи на 20 дБ больше, чем у полезного сигнала, и требуемое $\mathscr{E}_b / J_0$ для достижения нужного качества 10 дБ.

a) Определите ширину спектра ПП сигнала, требуемую для обеспечения условий задачи.

b) Если помеха имеет вид ПВМС, определите «дежурный цикл», который даёт наибольшую величину вероятности ошибки, и найдите величину этой вероятности.

13.9. Система СДМА состоит из 15 пользователей равной мощности, которые передают информацию со скоростью 10000 бит/с. Каждый использует широкополосный ПП сигнал и работает на скорости чипа 1 МГц. Модуляция – двоичная ФМ.

a) Определите $\mathscr{E}_b / J_0$, где J_0 – спектральная плотность мощности общей интерференции.

b) Каков выигрыш обработки?

c) Насколько надо увеличить выигрыш обработки, чтобы позволить работать удвоенному числу пользователей без ухудшения выходного ОСШ?

13.10. Широкополосный ПП сигнал с двоичной ФМ даёт выигрыш обработки 500. Какова помехозащищённость против гармонической помехи, если допустимая вероятность ошибки 10^{-5}?

13.11. Повторите задачу 13.10, если jammer является импульсным jammer с дежурным циклом 1 %.

13.12. Рассмотрите широкополосный ПП сигнал

$$c(t) = \sum_{n=-\infty}^{\infty} c_n p(t - nT_c),$$

где c_n – периодическая m-последовательность с периодом N=127, а $p(t)$ – прямоугольный импульс длительности $T_c = 1$мкс. Определите спектральную плотность мощности сигнала $c(t)$.

13.13. Допустим, что $\{c_{1i}\}$ и $\{c_{2i}\}$ является двумя двоичными $(0,1)$ периодическими последовательностями с периодом N_1 и N_2 соответственно. Определите период последовательности, получаемой суммированием по mod 2 $\{c_{1i}\}$ и $\{c_{2i}\}$.

13.14. Регистр сдвига максимальной длины с $m=10$ используется для генерирования псевдослучайной последовательности в широкополосной ПП системе. Длительность чипа $T_c = 1$мкс, а длительность бита (символа) $T_b = NT_c$, где N – длина (период) m-последовательности.

a) Определите выигрыш обработки системы в децибелах.

b) Определите помехозащищённость, если требуемое $\mathscr{E}_b / J_0 = 10$, а помеха является гармонической со средней мощностью $J_{ср}$.

13.15. Двоичная широкополосная ортогональная система ЧМ при СЧ использует $m=15$-ячеечный линейный сдвиговый регистр с обратными связями, который генерирует последовательность максимальной длины. Каждое состояние регистра сдвига выбирает одну из L неперекрывающихся частотных полосок. Битовая скорость – 100 бит/с, а скорость скачков – один на бит. Демодулятор использует некогерентное детектирование.

a) Определите полосу для скачков в этом канале.

b) Каков выигрыш обработки?

c) Какова вероятность ошибки в присутствии АБГШ?

13.16. Рассмотрите двоичную ортогональную систему с ФМ и СЧ, описанную в задаче 13.15. Предположим, что скорость скачков увеличилась до 2 скачка/бит. Приёмник использует квадратичное сложение для сложения сигналов от двух скачков.

a) Определите полосу скачков для канала.

b) Каков выигрыш обработки?

c) Какова вероятность ошибки в присутствии АБГШ?

13.17. В широкополосной системе с быстрым СЧ информация передаётся посредством ЧМ с некогерентным детектированием. Предположим, что имеется $N=3$ скачка/бит и декодирование жёстких решений сигнала на каждом скачке.

a) Определите вероятность ошибки для этой системы в канале с АБГШ со спектральной плотностью мощности $\frac{1}{2}N_0$ и ОСШ=13 дБ (общее ОСШ на три скачка).

b) Сравните результат a) с вероятностью ошибки широкополосной системы со СЧ, когда скачок совершается один раз на бит.

13.18. Шнрокополосная системе с медленным СЧ и двоичной ЧМ и некогерентным детектированием работает при $\mathscr{E}_b / J_0 = 10$ с полосой скачков 2 ГГц и битовой скоростью 10 кбит/с.

a) Каков выигрыш обработки системы?

b) Если помеха является парциально-полосовой, какая полоса занимается при наихудшей помехе?

c) Какова вероятность ошибки для случая наихудшей парциально-полосовой помехи?

13.19. Определите вероятность ошибки для широкополосной системы со СЧ, в которой используется двоичный свёрточный код в соединении с двоичной ЧМ. Интерференция в канале АБГШ. Выход демодулятора ЧМ подвергается квадратичному детектированию и проходит на декодер, который формирует оптимальное декодирование мягких решений по Витерби, как описано в разделе 8.2. Предположите, что скорость скачков – 1 скачок на кодированный символ.

13.20. Повторите задачу 13.19 при декодировании жёстких решений по Витерби.

13.21. Повторите задачу 13.19, когда осуществляются быстрые скачки частоты со скоростью L скачков на кодированный символ.

13.22. Повторите задачу 13.19, когда осуществляются быстрые скачки частоты с L скачками на кодовый символ , а декодируются жёсткие решения по Витерби. L чипов на кодовом символе детектируется квадратично и сигналы складываются до принятия решения.

13.23. Сигнал TATS, описанный в разделе 13.3.3 демодулируется параллельным блоком из восьми согласованных фильтров (восьмеричная ЧМ) и выход каждого фильтра детектируется квадратично. Восемь выходов, полученных на каждом из семи сигнальных интервалов (общее число выходов 56) используется для формирования 64 возможных величин решения, соответствующие коду Рида–Соломона (7,2). Определите верхнюю (объединённую) границу для вероятности ошибки кодового слова в канале с АБГШ при декодировании мягких решений.

13.24. Повторите задачу 13.23 для наихудшего случая парциально-полосовой интерференции в канале.

13.25. Получите результаты (13.2.62) и (13.2.63) из (13.2.61).

13.26. Покажите, что (13.3.14) следует из (13.3.13).

13.27. Получите (13.3.17) из (13.3.16).

13.28. Порождающие полиномы для конструирования кодовой последовательности Голда длины n=7 равны

$$g_1(p) = p^3 + p + 1$$

$$g_2(p) = p^3 + p^2 + 1$$

Генерируйте все последовательности Голда длины 7 и определите взаимную корреляцию одной последовательности с каждой другой.

13.29. В разделе 13.2.3 мы демонстрировали технику для расчёта вероятности ошибки кодовой системы с перемежением при импульсной интерференции, используя параметр предельной скорости R_0. Используйте кривые вероятности ошибки, данные на рис. Р13.29 при скорости свёрточного кода 1/2 и 1/3 и декодировании мягких решений по Витерби для определения соответствующих вероятностей ошибок для кодированной системы с импульсной интерференцией. Выполните расчёт для $K = 3, 5, 7$.

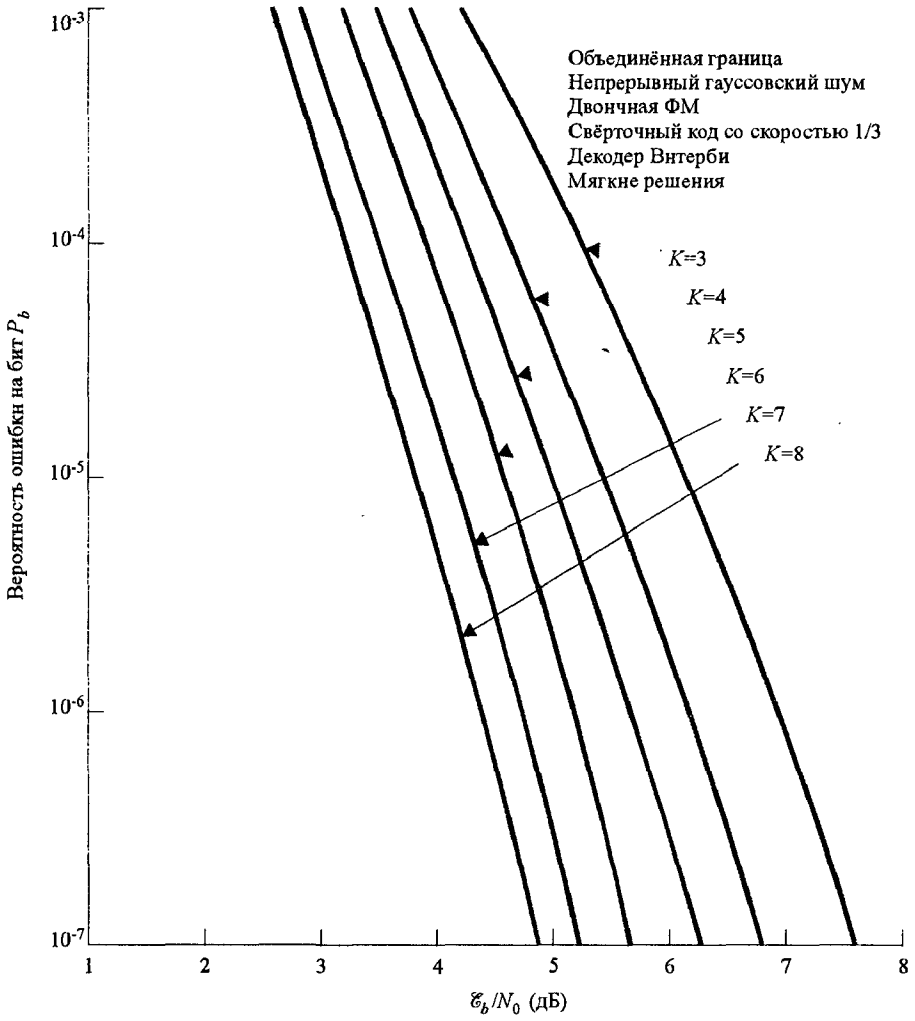


Рис. Р13.29

13.30. В кодированной и с перемежением ПП системе с двоичной ФМ при ПВМС и декодированием мягких решений предельно допустимая скорость равна $R_0 = 1 - \log_2(1 + \alpha e^{-\alpha \mathscr{E}_c / N_0})$,

где α – доля времени, когда система под воздействием jamminga, $\mathscr{E}_c = \mathscr{E}_b R$, R – битовая скорость, а $N_0 \equiv J_0$.

a) Покажите, что ОСШ на бит $\mathscr{E}_b / N_0$ можно выразить так: $\dfrac{\mathscr{E}_b}{N_0} = \dfrac{1}{\alpha R} \ln \dfrac{\alpha}{2^{1-R_0} - 1}$.

b) Определите величину α, которая максимизирует требуемое $\mathscr{E}_b / N_0$ (случай наихудшего ПВМС) и резервирующее значение $\mathscr{E}_b / N_0$.

c) Нарисуйте график $10\lg \mathscr{E}_b / rN_0$ как функцию R_0, где $r = R_0 / R$, для наихудшего случая импульсной помехи и для АБГШ $(\alpha = 1)$. Какие выводы вы сделаете относительно наихудшего воздействия ПВМС?

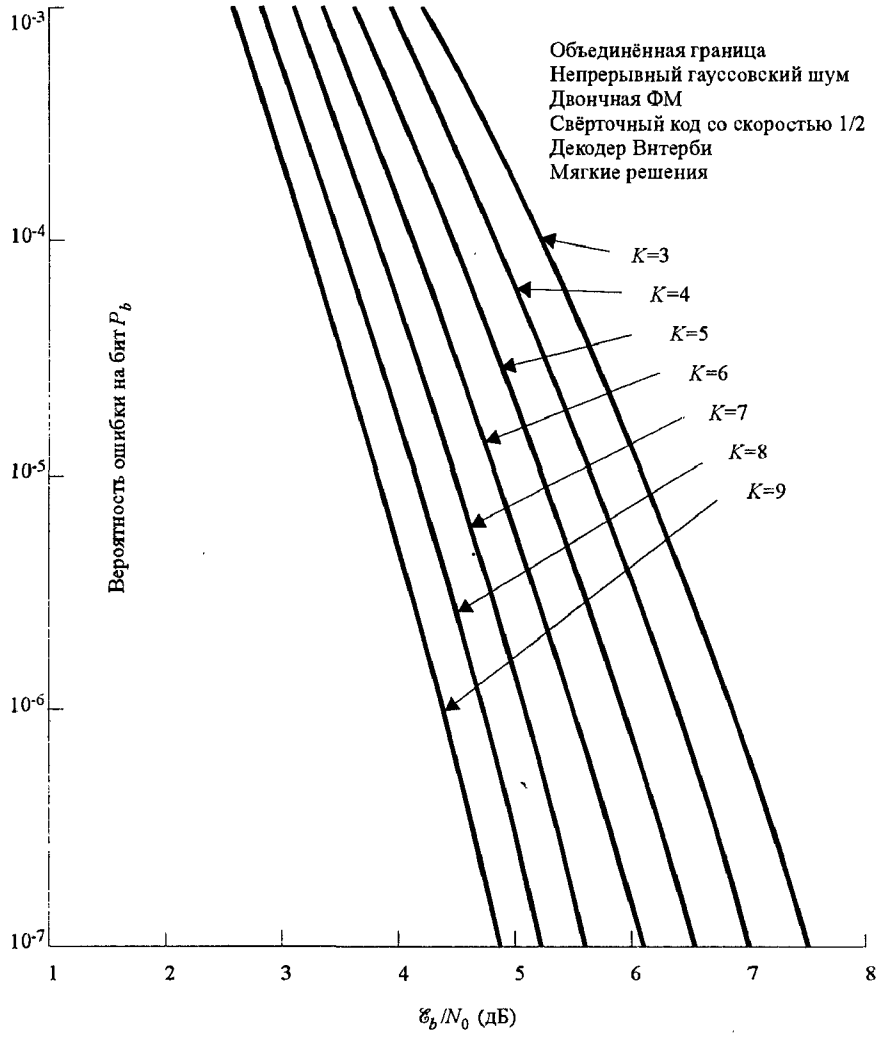


Рис. P13.29 (продолженне)

13.31. В кодированной и с перемежением системе со скачками частоты и q-ичной ЧМ с парциально-полосовой помехой и когерентной демодуляцией с декодированием мягких решений предельная скорость

$$R_0 = \log_2\left[\frac{q}{1+(q-1)\alpha e^{-\alpha \mathscr{E}_c / 2N_0}}\right],$$

где α – доля полосы, поражённая помехой, $\mathscr{E}_c$ – чиповая (одного тона) энергия, а $N_0 = J_0$.

a) Покажите, что ОСШ на бит можно выразить так: $\dfrac{\mathscr{E}_b}{N_0} = \dfrac{2}{\alpha R}\ln\dfrac{(q-1)\alpha}{q2^{-R_0}-1}$.

b) Определите величину α, которая максимизирует требуемое $\mathscr{E}_b / N_0$ (наихудший случай парциально-полосовой помехи) и результирующее значение $\mathscr{E}_b / N_0$.

c) Определите $r = R_0 / R$ в результате $\mathscr{E}_b / N_0$ из b) и нарисуйте график $10\lg \mathscr{E}_b / rN_0$ как функцию от нормированной предельной скорости $R_0 / \log_2 q$ для $q = 2, 4, 8, 16, 32$. Сравните эти графики с результатом задачи 13.30 c). Какие выводы вы делаете относительно воздействия наихудшей парциально-полосовой помехи? Какое влияние имеет увеличение объёма алфавита q? Каковы потери в ОСШ между результатами задачи 13.30 c) и q-ичной ЧМ при $q \to \infty$.

14

ЦИФРОВАЯ СВЯЗЬ ЧЕРЕЗ МНОГОПУТЕВЫЕ КАНАЛЫ С ЗАМИРАНИЯМИ

Предыдущие главы описывали синтез и анализ цифровых систем связи для передачи сообщений или по классическому неискажающему каналу с АБГШ, или по линейному фильтровому каналу с АБГШ. Мы видели, что искажения, свойственные линейному фильтровому каналу, предполагают специальную технику синтеза сигналов и до некоторой степени ухищренные алгоритмы адаптивного выравнивания для достижения хорошего качества. В этой главе мы рассмотрим синтез сигналов, структуру приёмника и качество приема для более сложных каналов, а именно, каналов со случайными переменными во времени импульсными характеристиками. Такие характеристики служат моделью при передаче сигналов по многим радиоканалам, таким как ионосферная радиосвязь в диапазоне частот 3–30 МГц (КВ), радиосвязь с тропосферным рассеянием (за горизонтом) в диапазоне частот 300–3000 МГц (УВЧ) и в полосе частот 3000–30000 МГц (СВЧ) и ионосферным рассеянием вперед в диапазоне частот 30–300 МГц (ОВЧ). Переменная во времени импульсная характеристика этих каналов является следствием постоянно меняющихся физических характеристик среды. Для примера, ионы в ионосферных слоях, которые отражают сигналы, передаваемые в ВЧ диапазоне, находятся все время в движении. Для пользователя канала движение ионов кажется случайным. Следовательно, если один и тот же сигнал передается по ВЧ каналу в двух сильно разнесённых временных интервалах, два принимаемых сигнала будут различными. Меняющиеся во времени отклики канала трактуются статистически (как реализации случайного процесса).

Мы начнём наше рассмотрение передачей цифровых сигналов по многопутевым каналам с замиранием, введя сначала статистическую характеристику канала. Затем мы рассчитаем качество различной базовой техники передачи цифровых сигналов по таким каналам. Результаты качества будут демонстрировать тяжелые потери в ОСШ, которыми приходится платить вследствие замираний принимаемого сигнала. Мы затем покажем, что потери в ОСШ можно существенно уменьшить посредством эффективной техники модуляции-кодирования и демодуляции-декодирования.

14.1. ХАРАКТЕРИСТИКИ МНОГОПУТЕВЫХ КАНАЛОВ С ЗАМИРАНИЯМИ

Если мы передаем предельно короткий импульс, в идеале δ-импульс, по меняющемуся во времени многопутевому каналу, принимаемый сигнал может появиться как ряд импульсов, как показано на рис.14.1.1. Следовательно, одной из характеристик многопутевой среды является время рассеяния сигнала, переданного по каналу. Второй характеристикой является изменение во времени структуры среды. Причина таких изменеий во времени обусловлена природой изменений во времени условий многопутевого распространения. Это значит, если мы повторим эксперимент по передаче импульса по каналу снова и снова, мы сможем наблюдать изменения в принимаемом ряде

импульсов, включающие в себя изменения размеров отдельных импульсов, изменения в относительных задержках между импульсами и, довольно часто, изменения в числе наблюдаемых импульсов принимаемого ряда импульсов, показанных на рис.14.1.1.

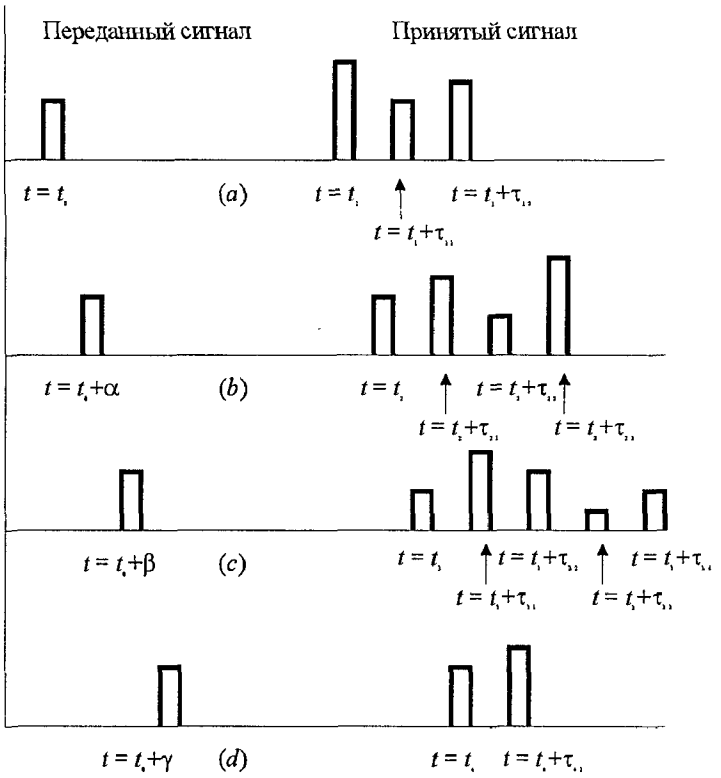


Рис.14.1.1. Примеры откликов меняющегося во времени многолучевого канала на одиночный импульс

Кроме того, возникающие во времени изменения являются непредсказуемыми для пользователя канала. Следовательно, разумно характеризовать меняющийся во времени многопутевой канал статистически. Рассмотрим влияние канала на переданный сигнал, который представим в общем так

$$s(t) = \mathrm{Re}\left[s_l(t)\mathrm{e}^{j2\pi f_c t}\right]. \qquad (14.1.1)$$

Предположим, что имеется многопутевое рассмотрение. С каждым путем связана задержка распространения и множитель ослабления. Как задержка распространения, так и множитель ослабления являются переменными во времени, как результат изменения в структуре среды. Таким образом, принимаемый полосовой сигнал можно выразить в виде

$$x(t) = \sum_n \alpha_n(t)\, s(t - \tau_n(t)), \qquad (14.1.2)$$

где $\alpha_n(t)$ – множитель ослабления принимаемого сигнала по n-му пути и $\tau_n(t)$ – задержка распространения для n-го пути. Подставив (14.1.1) в (14.11.2), получаем результат

$$x(t) = \mathrm{Re}\left\{\left[\sum_n \alpha_n(t)\mathrm{e}^{-j2\pi f_c \tau_n(t)} s_l\left(t - \tau_n(t)\right)\right]\mathrm{e}^{j2\pi f_c t}\right\}. \qquad (14.1.3)$$

Из (14.1.3) очевидно, что эквивалентный низкочастотный принимаемый сигнал равен

$$r_l(t) = \sum_n \alpha_n(t)\,\mathrm{e}^{-j2\pi f_c \tau_n(t)} s_l(t - \tau_n(t)). \qquad (14.1.4)$$

Поскольку $r_l(t)$ является откликом эквивалентного низкочастотного канала на эквивалентный низкочастотный сигнал $s_l(t)$, то следует, что эквивалентный низкочастотный канал описывается переменной во времени импульсной характеристикой

$$c(\tau;t)=\sum_n \alpha_n(t)\mathrm{e}^{-j2\pi f_c\tau_n(t)}\,\delta(\tau-\tau_n(t)). \qquad (14.1.5)$$

Для некоторых каналов, таких как канал тропосферного рассеяния, более подходит рассматривать принимаемый сигнал как состоящий из континуума многопутевых компонент. В этом случае принимаемый сигнал $x(t)$ выражается в интегральном виде

$$x(t)=\int_{-\infty}^{\infty}\alpha(\tau;t)s(t-\tau)d\tau, \qquad (14.1.6)$$

где $\alpha(\tau;t)$ определяет ослабление сигнальных компонент с задержкой (в момент времени t. Подставив теперь в (14.1.6) выражение для $s(t)$ из (14.1.1), получаем

$$x(t)=\mathrm{Re}\left\{\left[\int_{-\infty}^{\infty}\alpha(\tau;t)\,\mathrm{e}^{-j2\pi f_c\tau}s_l(t-\tau)d\tau\right]\mathrm{e}^{-j2\pi f_c t}\right\}. \qquad (14.1.7)$$

Поскольку интеграл в (14.1.7) представляет собой свертку $s_l(t)$ с эквивалентной переменной во времени характеристикой $c(\tau;t)$, то следует, что

$$c(\tau;t)=\alpha(\tau;t)\,\mathrm{e}^{-j2\pi f_c\tau}, \qquad (14.1.8)$$

где $c(\tau;t)$ представляет отклик канала в момент t на δ-импульс, поданный ко входу в момент времени $t-\tau$. Таким образом, (14.1.8) – подходящее определение импульсной характеристики эквивалентного низкочастотного канала, когда канал образуется за счет непрерывной многопутёвости, а (14.1.5) – подходящее определение для канала, который содержит дискретные многопутевые компоненты.

Теперь рассмотрим передачу немодулированного сигнала несущей на частоте f_c. Тогда $s_l(t)=1$ для всех t и, следовательно, принимаемый сигнал для случая дискретной многопутёвости, определяемый (14.1.4), приводит к

$$r_l(t)=\sum_n \alpha_n(t)\,\mathrm{e}^{-j2\pi f_c\tau_n(t)}=\sum_n \alpha_n(t)\mathrm{e}^{-j\theta_n(t)}=x(t)-jy(t), \qquad (14.1.9)$$

где $\theta_n(t)=2\pi f_c\tau_n(t)$, а $x(t)=\sum_n \alpha_n(t)\cos\theta_n(t)$, $y(t)=\sum_n \alpha_n(t)\sin\theta_n(t)$ – суммарные квадратурные компоненты. Таким образом, принимаемый сигнал состоит из суммы переменных во времени векторов, имеющих амплитуды $\alpha_n(t)$ и фазы $\theta_n(t)$. Заметим, что в среде требуются большие динамические изменения для того, чтобы $\alpha_n(t)$ изменялся бы существенно и вызвал бы достаточные изменения принимаемого сигнала. С другой стороны, $\theta_n(t)$ будет меняться на 2π радиан, когда τ_n изменится на $1/f_c$. Но $1/f_c$ – малое число, и, следовательно, θ_n может изменяться на 2π радиан при относительно малых изменениях в среде. Мы также ожидаем, что задержки $\tau_n(t)$, связанные с различными путями сигналов, изменяются с различной скоростью и случайным (непредсказуемым) образом. Это означает, что принимаемый сигнал $r_l(t)$ в (14.1.9) можно моделировать случайным процессом. Если имеется большое число путей, то можно использовать центральную предельную теорему теории вероятностей. Это значит, что $r_l(t)$ можно моделировать как комплексный гауссовский случайный процесс. Нетрудно видеть, что по определению (2.4.19) определяет передаточную функцию канала $C(f_c,t)$, связанную с импульсной характеристикой канала $c(\tau;t)$ парой преобразований Фурье. С учётом сказанного следует, что частотную характеристику канала $C(f_c,t)$ можно моделировать комплексным гауссовским случайным процессом по переменной t. Но поскольку $c(\tau;t)$ связана с $C(f_c,t)$ линейным преобразованием Фурье, то переменные во времени

импульсные характеристики $c(\tau;t)$ являются *комплексным гауссовским случайным процессом* по переменной t.

Многопутевая модель распространения для канала, воплощаемая в принимаемом сигнале $r_l(t)$ согласно (14.1.9), приводит к замиранию сигнала. Феномен замирания – прежде всего результат изменений во времени фаз $\{\theta_n(t)\}$. Это означает, что случайные, меняющиеся во времени фазы $\{\theta_n(t)\}$, связанные с векторами $\{\alpha_n e^{-j\theta_n}\}$, в определенный момент времени играют при суммировании векторов неблагоприятную роль. Когда это происходит, результирующий принимаемый сигнал $r_l(t)$ очень мал или практически равен нулю. В другие моменты времени векторы $\{\alpha_n e^{-j\theta_n}\}$ складываются благоприятно, так что принимаемый сигнал большой. Таким образом, амплитудные изменения принимаемого сигнала, называемые замираниями сигнала, обусловлены переменными во времени многопутевыми характеристиками канала.

Когда импульсная характеристика $c(\tau;t)$ моделируется как комплексный случайный гауссовский процесс с нулевым средним, огибающая $|c(\tau;t)|$ в любой момент t распределена по Релею. В этом случае канал называют каналом с релеевскими замираниями. В случае, когда имеются фиксированные рассеиватели или отражатели сигнала в среде в дополнение к случайно перемещающимся рассеивателям, $c(\tau;t)$ нельзя моделировать процессом с нулевым средним. В этом случае огибающая $|c(\tau;t)|$ имеет райсовское распределение, и канал называют каналом с райсовскими замираниями[1]. Другой функцией распределения, которая используется для моделирования огибающей сигнала, является m-распределение Накагами. Эта модель замираний рассматривается в разделе 14.1.2.

14.1.1. Корреляционная функция канала и спектр мощности

Теперь рассмотрим некоторые используемые корреляционные функции и спектральные плотности мощности, которые определяют характеристики многопутевого

---

[1] Если передаётся узкополосный сигнал с полосой Δf, и в модели (14.1.9) взаимное запаздывание путей $\max|\tau_\nu - \tau_l| << 1/\Delta f$, $\nu \neq l$, то говорят о модели «однолучевого» канала, а компоненты в (14.1.9) называют «подлучами». В «однолучевой» модели разность фаз сигнала на различных частотах близка к нулю. Это приводит к неселективным по частоте замираниям сигнала. Если взаимное запаздывание путей в принимаемом сигнале удовлетворяет условию $\max|\tau_\nu - \tau_l| \geq 1/\Delta f$, $\nu \neq l$, то говорят о многолучевой модели канала. Отдельные лучи в этом случае формируются сильно разнесёнными неоднородностями среды распространения, и посему сигналы отдельных лучей можно считать независимыми случайными процессами. В «многолучевой» модели разность фаз сигнала на различных частотах может существенно отличаться, что приводит к селективным по частоте замираниям сигнала. Весьма общей стохастической моделью «многолучевого» радиоканала, имеющей наглядную физическую интерпретацию и подтверждённую экспериментально [66, 67], является общая гауссовская модель, согласно которой квадратурные компоненты в l-м луче $x_l(t) = \gamma_l(t)\cos\varphi_l(t)$, $y_l(t) = \gamma_l(t)\sin\varphi_l(t)$ являются гауссовскими случайными процессами. Общая гауссовская модель многолучевого канала разрабатывалась в 60-х годах Д.Д. Кловским и его учениками [66, 68, 69, 70]. С 1973 г. эта модель вошла в учебники по теории электросвязи [71/72]. Одномерные распределения амплитуд $\gamma_l = \sqrt{x_l^2 + y_l^2}$ и фаз $\varphi_l = \operatorname{arctg}(y_l/x_l)$ в этом случае называют четырёхпараметрическими, поскольку они зависят от четырёх параметров $m_{x,l}$, $m_{y,l}$, $\sigma^2_{x,l}$, $\sigma^2_{y,l}$. Модели Райса и Релея являются частными случаями 4-параметрического распределения, а m-распределение Накагами для амплитуд сигнала можно считать аппроксимацией 4-параметрического распределения [66] (прп).

канала с замираниями. Нашей исходной точкой является эквивалентная низкочастотная импульсная характеристика $c(\tau;t)$, которая характеризуется как комплексный случайный процесс по переменной t. Предположим, что процесс $c(\tau;t)$ стационарен в широком смысле. Тогда мы определяем автокорреляционную функцию $c(\tau;t)$ так:

$$\phi_c(\tau_1,\tau_2;\Delta t)=\tfrac{1}{2}E\left[c^*(\tau_1;t)c(\tau_2;t+\Delta t)\right]. \qquad (14.1.10)$$

В большинстве сред, где передается радиосигнал, ослабление и сдвиг фазы в канале, связанные с задержкой в пути τ_1, некоррелированы с ослаблением и сдвигом фазы, связанными с задержкой в пути τ_2. Это обычно называется некоррелированным рассеянием. Мы сделаем предположение, что рассеяния при двух различных задержках некоррелированы, и учтем это в (14.1.10), чтобы получить

$$\tfrac{1}{2}E\left[c^*(\tau_1;t)c(\tau_2;t+\Delta t)\right]=\phi_c(\tau_1;\Delta t)\delta(\tau_1-\tau_2). \qquad (14.1.11)$$

Если возьмем $\Delta t=0$, результирующая автокорреляционная функция $\phi_c(\tau;0)\equiv\phi_c(\tau)$ – это просто средняя мощность выхода канала как функция от задержки во времени τ. Из этих соображений $\phi_c(\tau)$ называют *интенсивностью многопутевого профиля* или *спектром мощности задержек* канала. В общем $\phi_c(\tau;\Delta t)$ определяет среднюю мощность выхода канала как функцию от времени задержки τ и разницы моментов наблюдения Δt. На практике функция $\phi_c(\tau;\Delta t)$ измеряется путем передачи по каналу очень короткого импульса и вычислением взаимной корреляции принимаемого сигнала со своей собственной запаздывающей копией. Можно ожидать, что измеренная функция $\phi_c(\tau)$ имеет типичный вид, показанный на рис.14.1.2. Область значений, в которой $\phi_c(\tau)$ существенно больше нуля, называют *многопутевым рассеянием канала* и обозначают T_m.

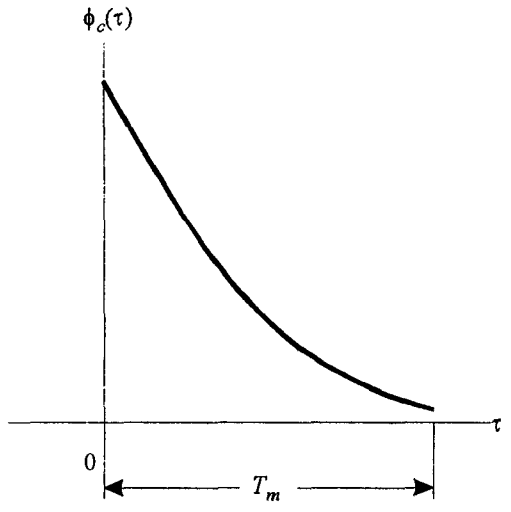


Рис. 14.1.2. Профиль многопутевой интенсивности

Полную характеристику многопутевого переменного во времени канала можно определить и в частотной области. Взяв преобразование Фурье от $c(\tau;t)$, мы получаем переменную во времени передаточную функцию $C(f;t)$, где f – частотная переменная. Итак,

$$C(f;t)=\int_{-\infty}^{\infty}c(\tau;t)\,e^{-j2\pi ft}d\tau. \qquad (14.1.12)$$

В предположении, что канал стационарен в широком смысле, мы определим автокорреляционную функцию

$$\phi_C(f_1,f_2;\Delta t)=\tfrac{1}{2}E\left[C^*(f_1;t)C(f_2;t+\Delta t)\right]. \qquad (14.1.13)$$

Поскольку $C(f;t)$ является преобразованием Фурье от $c(\tau;t)$, то не является неожиданностью, что $\phi_C(f_1,f_2;\Delta t)$ связано с $\phi_c(\tau;\Delta t)$ преобразованием Фурье. Это соотношение легко получить подстановкой (14.1.12) в (14.1.13)

$$\begin{aligned}\phi_C(f_1,f_2;\Delta t) &= \tfrac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E\left[c^*(\tau_1;t)C(\tau_2;t+\Delta t)\right]e^{j2\pi(f_1\tau_1-f_2\tau_2)}d\tau_1 d\tau_2 \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\phi_c(\tau_1;\Delta t)\delta(\tau_1-\tau_2)e^{j2\pi(f_1\tau_1-f_2\tau_2)}d\tau_1 d\tau_2 \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\phi_c(\tau_1;\Delta t)e^{j2\pi(f_1-f_2)\tau_1}d\tau_1 \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\phi_c(\tau_1;\Delta t)e^{-j2\pi\Delta f\tau_1}d\tau_1 \equiv \phi_C(\Delta f;\Delta t),\end{aligned} \tag{14.1.14}$$

где $\Delta f = f_2 - f_1$. Из (14.1.14) мы видим, что $\phi_C(\Delta f;\Delta t)$ является преобразованием Фурье от многопутевой интенсивности профиля. Далее, из предположения, что рассеяние некоррелировано (по отдельным путям) следует, что автокорреляционная функция от $C(f;t)$ по частоте зависит только от разности частот $\Delta f = f_2 - f_1$. Следовательно, подобает называть $\phi_C(\Delta f;\Delta t)$ *совместной корреляционной функцией канала в частотной и временной области*. На практике ее можно измерить путем передачи по каналу двух синусоид с частотным разносом Δf и измерением взаимной корреляции двух отдельно принимаемых сигналов с временной задержкой между ними Δt.

Предположим, что возьмём в (14.1.14) $\Delta t = 0$. Тогда $\phi_C(\Delta f;0) \equiv \phi_C(\Delta f)$, и $\phi_c(\tau;0) \equiv \phi_c(\tau)$, связь между ними упрощается:

$$\phi_C(\Delta f) = \int_{-\infty}^{\infty}\phi_c(\tau)e^{-j2\pi\Delta f\tau}d\tau. \tag{14.1.15}$$

Это соотношение отображено графически на рис.14.1.3.

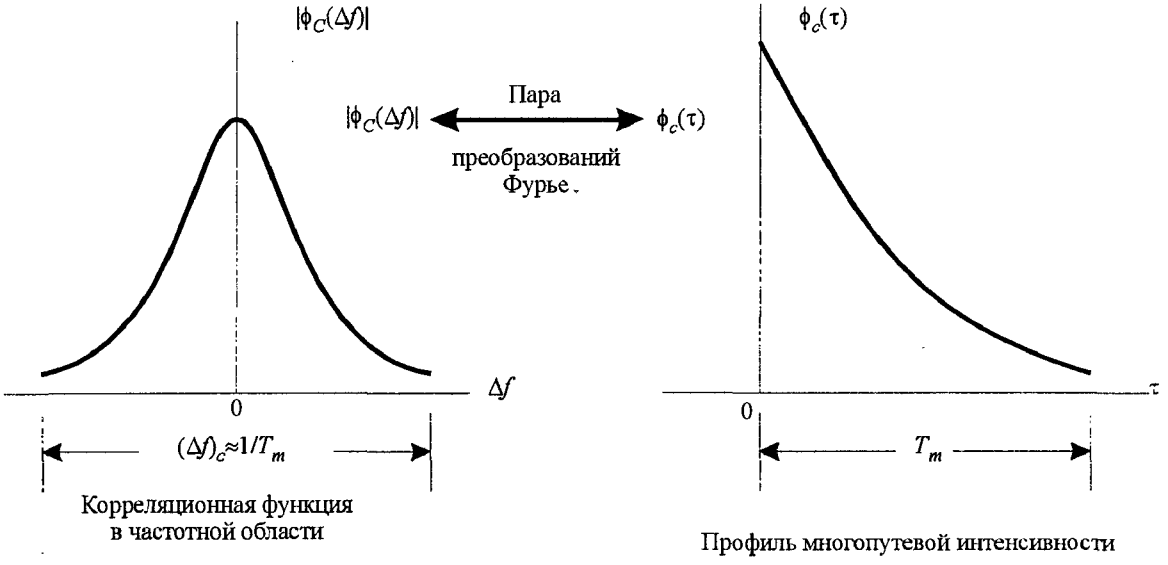


Рис. 14.1.3. Соотношение между $|\phi_C(\Delta f)|$ и $\phi_c(\tau)$

Поскольку $\phi_C(\Delta f)$ является автокорреляционной функцией по частотной переменной, она обеспечивает нам возможность измерить частотную когерентность канала. Как следствие преобразования Фурье между $\phi_C(\Delta f)$ и $\phi_c(\tau)$, обратная величина многопутевого рассеяния является мерой *частотной когерентности канала*.

Это значит, что

$$(\Delta f)_c = \frac{1}{T_m}, \tag{14.1.16}$$

где $(\Delta f)_c$ означает полосу частотной когерентности. Таким образом, две синусоиды с частотным разносом, большим $(\Delta f)_c$, ведут себя различно в канале. Если $(\Delta f)_c$ мало по сравнению с полосой частот переданного сигнала, канал называют частотно селективным. В этом случае сигнал существенно искажается в канале. С другой стороны, если $(\Delta f)_c$ велика по сравнению с полосой частот переданного сигнала, канал называют *частотно неселективным.*

Теперь сосредоточим наше внимание на изменении канала во времени, измеряемом параметром Δt в $\phi_C(\Delta f;\Delta t)$. Изменения во времени характеристик канала свидетельствуют о доплеровском рассеянии и, возможно, также о сдвиге спектральных линий. Чтобы выявить связь эффекта Доплера и изменений во времени канала, определим преобразование Фурье от $\phi_C(\Delta f;\Delta t)$ по переменной Δt, чтобы получить функцию $S_C(\Delta f;\lambda)$. Т.е.

$$S_C(\Delta f;\lambda)=\int_{-\infty}^{\infty}\phi_C(\Delta f;\Delta t)e^{-j2\pi\lambda\Delta t}d\Delta t. \qquad (14.1.17)$$

При $\Delta f=0$ $S_C(0;\lambda)\equiv S_C(\lambda)$, и из (14.1.17) следует

$$S_C(\lambda)=\int_{-\infty}^{\infty}\phi_C(\Delta t)e^{-2\pi\lambda\Delta t}d\Delta t. \qquad (14.1.18)$$

Функция $S_C(\lambda)$ определяет спектр мощности и дает интенсивность сигнала как функцию от частоты Доплера λ. Поэтому $S_C(\lambda)$ называют доплеровским спектром мощности канала.

Из (14.1.18) мы видим, что если канал не меняется во времени, $\phi_C(\Delta t)=K$, и функция $S_C(\lambda)$ становится равной $K\delta(\lambda)$. Следовательно, когда нет изменений канала во времени, не наблюдается спектральное расширение при передаче чистого тона.

Область значений λ, в которой $S_C(\lambda)$ существенно отлично от нуля, называют доплеровским рассеянием в канале B_d. Поскольку $S_C(\lambda)$ связано с $\phi_C(\Delta t)$ преобразованием Фурье, обратная величина B_d является мерой временной когерентности канала, т.е.

$$(\Delta t)_c\approx\frac{1}{B_d}, \qquad (14.1.19)$$

где $(\Delta t)_c$ называют *временем когерентности.* Ясно, что канал с медленными изменениями имеет большую временную когерентность или, что эквивалентно, малое доплеровское рассеяние. Рис.14.1.4 иллюстрирует соотношение между $\phi_C(\Delta t)$ и $S_C(\lambda)$.

Теперь мы установим соотношение Фурье между $\phi_C(\Delta f;\Delta t)$ и $\phi_c(\tau;\Delta t)$, включающих переменные $(\tau,\Delta f)$, и между $\phi_C(\Delta f;\Delta t)$ и $S_C(\Delta f;\lambda)$, включающих переменные $(\Delta t,\lambda)$. Имеются два дополнительных преобразований Фурье, которые мы можем найти и которые служат для связи $\phi_c(\tau;\Delta t)$ и $S_C(\Delta f;\lambda)$, и таким образом замыкается цепь. Требуемое отношение можно получить, определив новую функцию, обозначаемую $S(\tau;\lambda)$, как преобразование Фурье $\phi_c(\tau;\Delta t)$ по переменной Δt, т.е.

$$S(\tau;\lambda)=\int_{-\infty}^{\infty}\phi_c(\tau;\Delta\tau)e^{-j2\pi\lambda\Delta t}d\Delta t. \qquad (14.1.20)$$

Отсюда следует, что $S(\tau;\lambda)$ и $S_C(\Delta f;\lambda)$ являются парой преобразований Фурье. То есть

$$S(\tau;\lambda)=\int_{-\infty}^{\infty}S_C(\Delta f;\lambda)e^{-j2\pi\tau\Delta f}d\Delta f. \qquad (14.1.21)$$

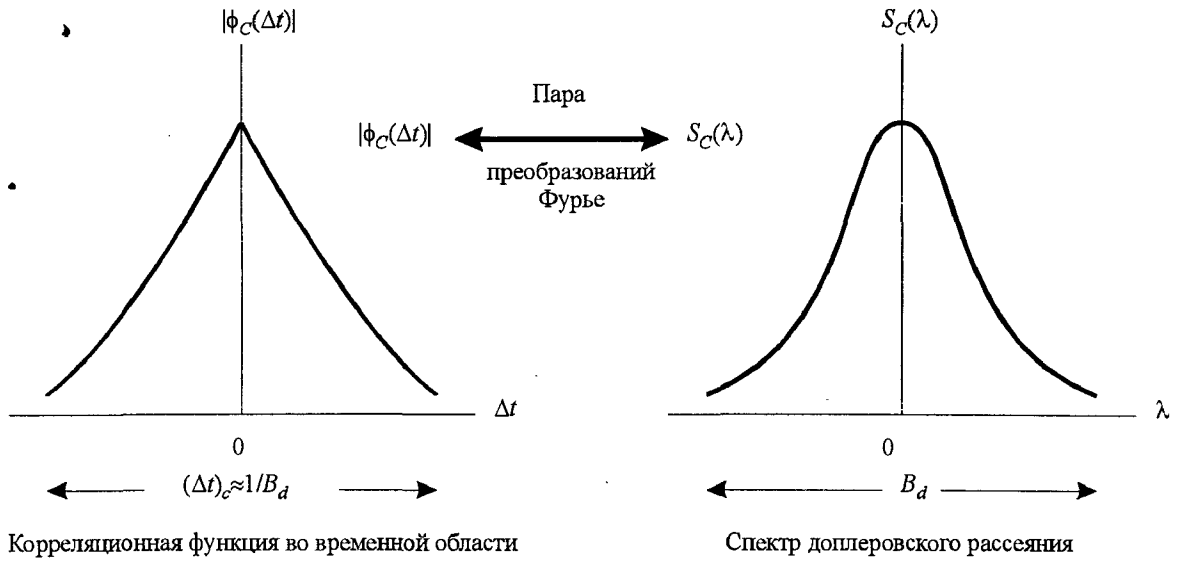


Рис. 14.1.4. Соотношение между $|\phi_C(\Delta t)|$ и $S_C(\lambda)$

Далее, $S(\tau;\lambda)$ и $\phi_C(\Delta f;\Delta t)$ связаны двойным преобразованием Фурье

$$S(\tau;\lambda)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\phi_C(\Delta f;\Delta t)\,e^{-j2\pi\lambda\Delta t}e^{j2\pi\tau\Delta f}\,d\Delta t\,d\Delta f\,. \qquad (14.1.22)$$

Эту новую функцию $S(\tau;\lambda)$ называют *функцией рассеяния канала*. Она определяет меру средней мощности на выходе канала, как функцию времени задержки τ и доплеровской частоты λ.

Соотношения между четырьмя функциями $\phi_C(\Delta f;\Delta t)$, $\phi_c(\tau;\Delta t)$, $\phi_C(\Delta f;\lambda)$ и $S(\tau;\lambda)$ подытожены рисунком 14.1.5.

Функция рассеяния $S(\tau;\lambda)$, измеренная на тропосферной линии рассеяния протяженностью 150 миль, показана на рис.14.1.6. Сигнал, использованный для зондирования канала, имеет разрешение во времени 0,1мкс. Поэтому ось для времени запаздывания проквантована с шагом 0,1мкс. Из рисунка мы видим, что многопутевое рассеяние равно $T_m = 0{,}7$ мкс. С другой стороны, доплеровское рассеяние, которое можно определить как полосу спектра мощности для каждого пути сигнала на уровне 3 дб, оказывается переменной для каждого сигнального пути. Для примера, в одном пути оно меньше 1 Гц в то время как в некоторых других путях оно составляет несколько герц. Для наших целей мы возьмем наибольшее рассеяние по различным путям на уровне 3 дб и назовем ее доплеровским рассеянием.

14.1.2. Статистические модели для каналов с замираниями

Имеются несколько распределений вероятности, которые следует рассмотреть при конструировании модели статистических характеристик канала с замираниями. Когда имеется большое число рассеивателей в канале, которые образуют сигнал на приеме, как в случае ионосферного или тропосферного распространения сигнала, применение центральной предельной теоремы вероятностей приводит к гауссовской модели для характеристики канала. Если процесс с нулевым средним, тогда огибающая характеристики канала в любой момент времени имеют релеевское распределение вероятностей, а фаза распределена равномерно на интервале $(0, 2\pi)$. Релеевское распределение можно записать в виде

$$p_R(r)=\frac{2r}{\Omega}e^{-r^2/\Omega}, \qquad r\geq 0, \qquad (14.1.23)$$

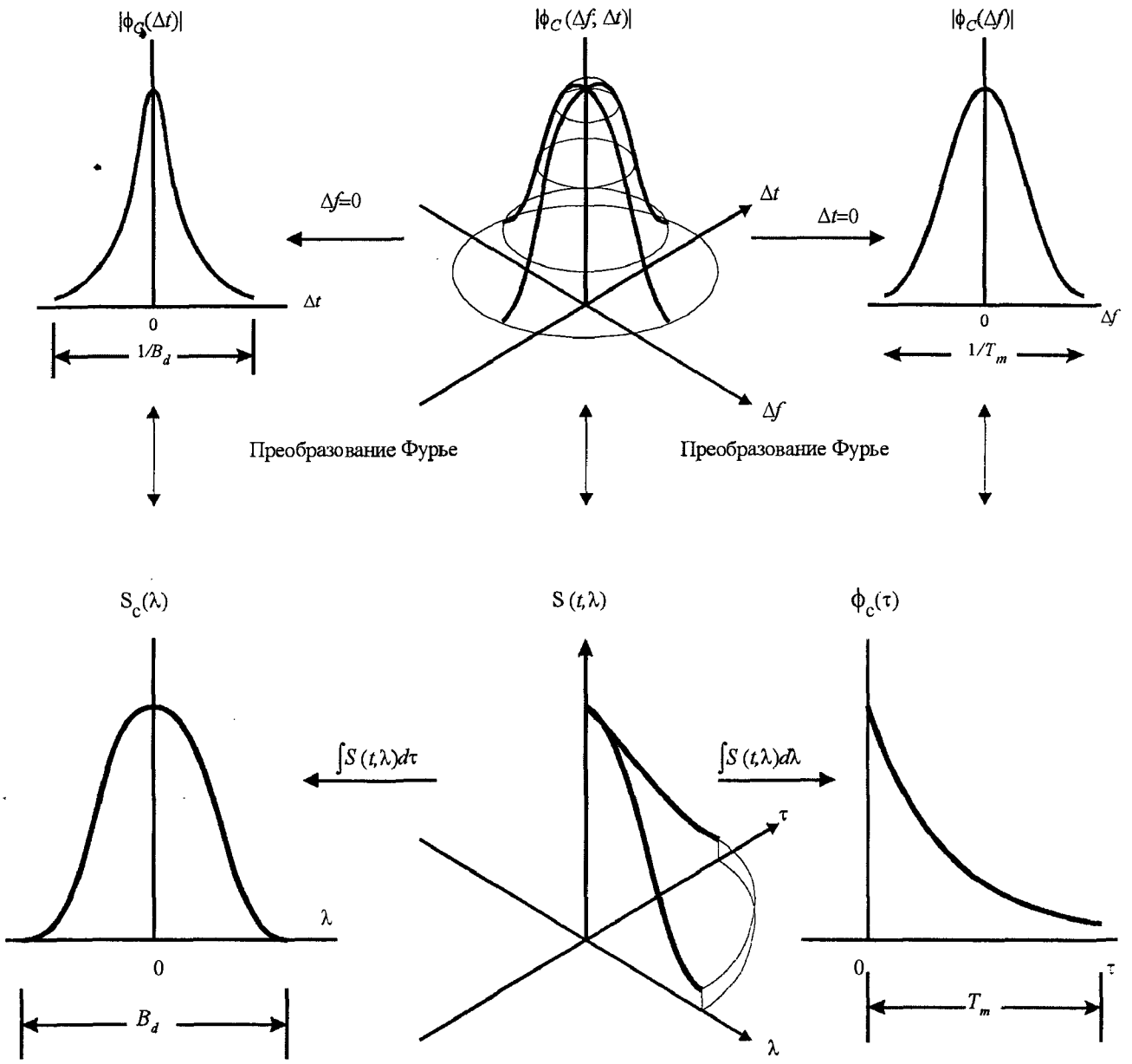


Рис. 14.1.5. Соотношение между корреляционной функцией канала и функцией рассеяния [*Green* (1962)]

где

$$\Omega = E(R^2). \qquad (14.1.24)$$

Мы видим, что релеевское распределение характеризуется единственным параметром $E(R^2)$.

Альтернативной статистической моделью для огибающей характеристики канала является m-распределение Накагами, определяемое ФПВ (2.1.14). В противовес распределению Релея с единственным параметром, который можно использовать для оценки статистики канала с замираниями, m-распределение Накагами является двухпараметрическим, именно, включает параметр m и второй момент $\Omega = E(R^2)$. Как следствие, m-распределение позволяет более гибко и точно оценить наблюдаемую статистику сигнала. Его можно использовать для моделирования условий замираний в канале, которые являются более или менее глубокими, чем определяемые законом Релея, и оно включает распределение Релея как частный случай (m=1). Для примера, Турин (1972) и

Судзуки (1977) показали, что m-распределение Накагами является наиболее подходящим для сигналов, принимаемых в городских многопутевых радиоканалах.

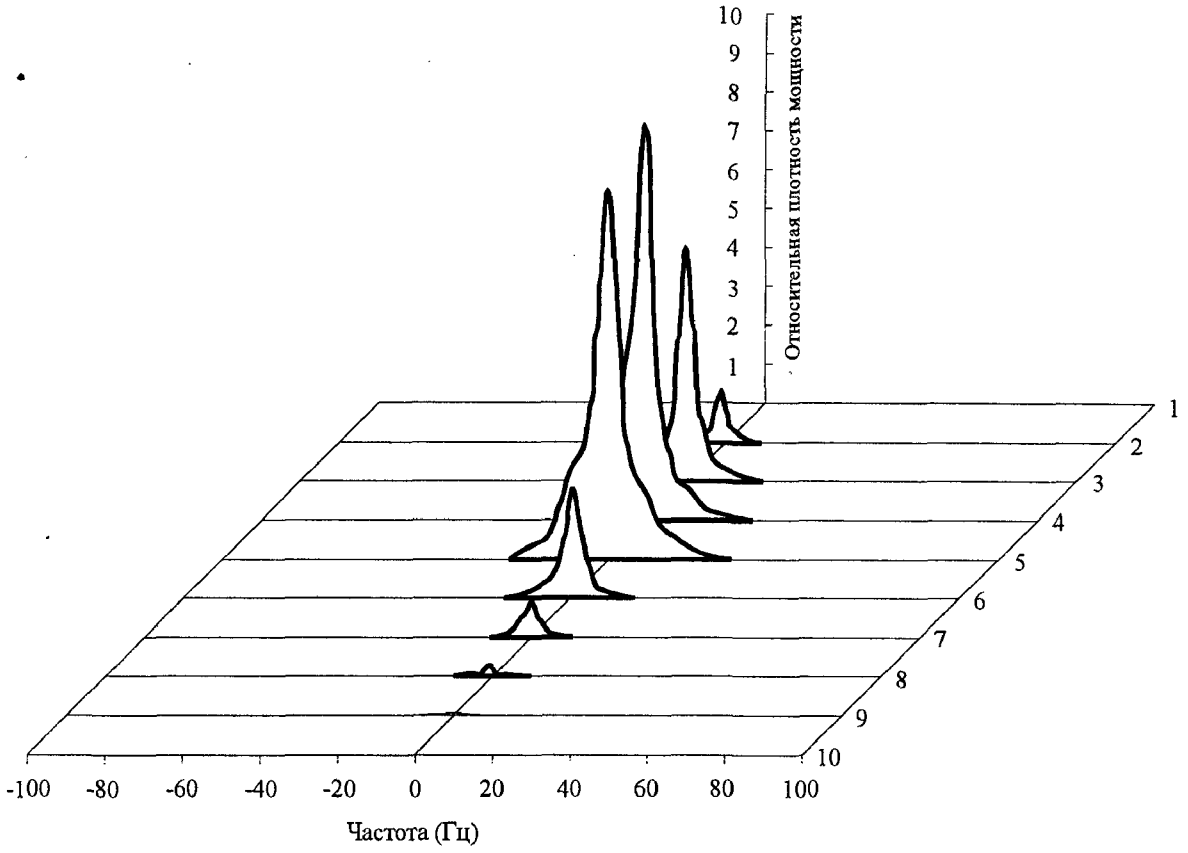


Рис. 14.1.6. Функция рассеяния канала с тропосферным рассеянием. Сечения через 0,1 мкс

Распределение Райса также имеет два параметра. Его можно выразить ФПВ (2.1.41) параметрами s и σ^2. Напомним, что s^2 называется параметром нецентральности в эквивалентном хи-квадрат распределении. Он определяет мощность не замирающей сигнальной компоненты, иногда называемой *зеркальной* (*регулярной*) *компонентой* принимаемого сигнала.

Имеется много радиоканалов, в которых встречаются замирания. В основном это каналы, которые осуществляют связь за пределами прямой видимости (ЗПВ), связь с многопутевыми компонентами, образующимися от вторичных отражений, в частности от местных предметов. В таких каналах число многопутевых компонент невелико и, следовательно, канал можно моделировать некоторой более простой формой. В качестве примера приведём две канальные модели.

В качестве первого примера рассмотрим линию связи между самолетом и землей, в которой имеется прямой путь и одна компонента с задержкой τ_0 относительно прямого пути. Импульсную характеристику такого канала можно моделировать так

$$c(\tau;t)=\alpha\delta(\tau)+\beta(t)\delta(\tau-\tau_0(t)), \qquad (14.1.25)$$

где α – множитель ослабления прямого пути, а $\beta(t)$ представляет меняющуюся во времени многопутевую сигнальную компоненту, образующуюся от местных отражений. Часто $\beta(t)$ можно характеризовать как случайный гауссовский процесс с нулевым средним. Передаточную функцию для этой модели канала можно выразить так:

$$C(f;t)=\alpha+\beta(t)e^{-j2\pi f\tau_0(t)}. \qquad (14.1.26)$$

Для этого канала годится модель Райса, определённая раньше. Прямой путь с ослаблением α представляет регулярную компоненту, а $\beta(t)$ представляет релеевскую замирающую компоненту.

Установили, что похожая модель имеет место в микроволновых ЗПВ радиоканалах, используемых для передачи на большое расстояние речевых и видеопередач телефонными компаниями. Для таких каналов Руммлер (1979) предложил трёхпутевую модель, основанную на канальных измерениях, выполненных на типичных ЗПВ линиях в полосе частот около 6 ГГц. Различие задержек на двух многопутевых компонентах относительно мало и, следовательно, модель, разработанная Руммлером, характеризуется передаточной функцией

$$C(f) = \alpha\left[1 - \beta e^{-j2\pi(f-f_0)\tau_0}\right], \qquad (14.1.27)$$

где α – обычный параметр рассеяния, β называется параметром формы, который относится к многопутевой компоненте, f_0 – минимальная частота замираний, а τ_0 – относительно малое время задержки между прямой и многопутевой компонентами. Эта простая модель была использована для описания сигналов по измерениям в канале.

Руммлер нашел, что параметры α и β можно характеризовать как случайные величины, которые для практических целей можно считать статистически независимыми. По канальным измерениям он нашел, что β имеет распределение вида $(1-\beta)^{2/3}$. Распределение для α хорошо моделируется логнормальным распределением, т.е. $-\log\alpha$ аппроксимируется гауссовским распределением. Было найдено, что для $\beta > 0{,}5$ среднее значение $-20\log\alpha$ равно 25 дБ, а стандартное отклонение равно 5 дБ. Для меньших значений β среднее значение уменьшается до 15 дБ. Параметр задержки, по данным измерений, равен $\tau_0 = 6{,}3$ нс.

Квадрат амплитудной характеристики $C(f)$ равен

$$|C(f)|^2 = \alpha^2\left[1 + \beta^2 - 2\beta\cos 2\pi(f - f_0)\tau_0\right]. \qquad (14.1.28)$$

$|C(f)|$ отображена на рис.14.1.7 как функция частоты $f - f_0$ при $\tau_0 = 6{,}3$ нс.

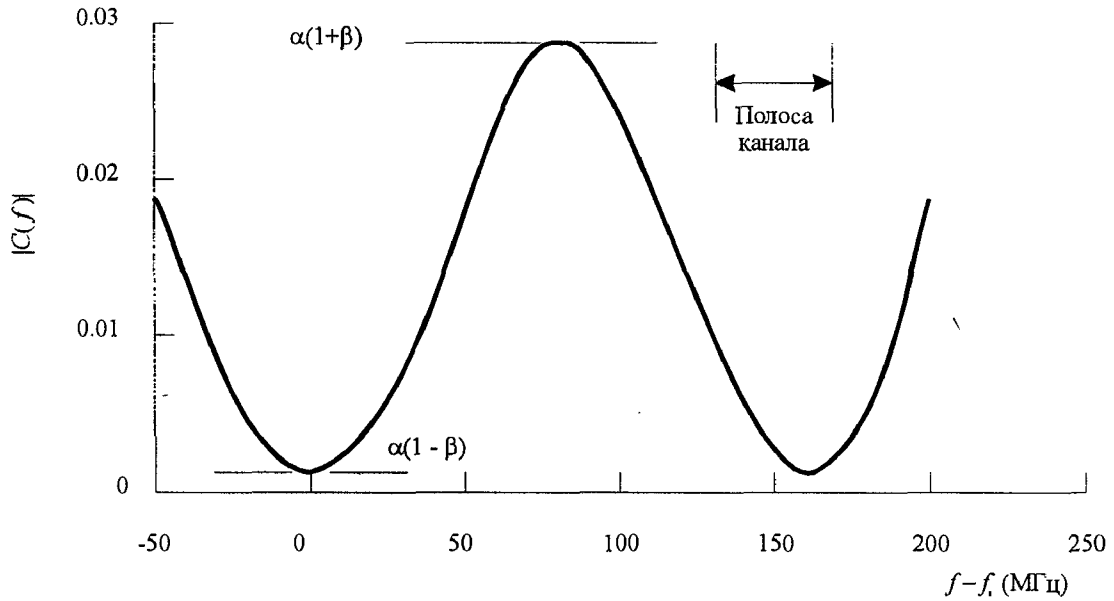


Рис. 14.1.7. Пример АЧХ двухлучевого канала (модель канала ЗПВ)

Заметим, что влияние многопутевой компоненты сводится к созданию глубокого ослабления при $f = f_0$, и оно повторяется по частоте через $1/\tau_0 \approx 159$ МГц. Для сравнения заметим, что типичная полоса канала 30 МГц. Эта модель использована Ландгремом и Руммлером (1979) для определения вероятности ошибки цифровых радиосистем.

14.2. ВЛИЯНИЕ ХАРАКТЕРИСТИК СИГНАЛА НА ВЫБОР МОДЕЛИ КАНАЛА

Обсудив в общем статистические характеристики переменных во времени многопутевых каналов через корреляционные функции, описанные в разделе 14.1, теперь рассмотрим влияние характеристик сигналов на выбор подходящей модели канала. Пусть $s_l(t)$ является эквивалентным низкочастотным сигналом, передаваемым по каналу, а $S_l(f)$ означает его частотное отображение. Тогда эквивалентный низкочастотный принимаемый сигнал, исключая аддитивный шум, можно выразить или через $c(\tau;t)$ $s_l(t)$ во временной области

$$r_l(t) = \int_{-\infty}^{\infty} c(\tau;t) s_l(t-\tau) d\tau, \qquad (14.2.1)$$

или через частотные функции $C(f;t)$ и $S_l(f)$ так

$$r_l(t) = \int_{-\infty}^{\infty} C(f;t) S_l(f) e^{j2\pi ft} df. \qquad (14.2.2)$$

Предположим, что мы передаем цифровую информацию по каналу путем модуляции (по амплитуде или фазе или одновременно по обоим параметрам) базового импульса $s_l(t)$ со скоростью $1/T$, где T – сигнальный интервал. Из (14.2.2) очевидно, что меняющийся во времени канал, характеризуемый передаточной функцией $C(f;t)$ искажает сигнал $S_l(f)$. Если $S_l(f)$ имеет полосу частот W большую, чем полоса когерентности канала $(\Delta f)_c$, $S_l(f)$ подвержено различным ослаблениями и фазовым сдвигам по полосе. В этом случае канал называется частотно-селективным. Дополнительное искажение обусловлено изменениями во времени $C(f;t)$. Этот вид искажений проявляется как изменение интенсивности принимаемого сигнала и называется замиранием. Следует подчеркнуть, что частотная селективность и замирания рассматриваются как два различных вида искажений. Первая зависит от многопутевого рассеяния или, что эквивалентно, от отношения полосы частотной когерентности к полосе передаваемого сигнала W. Второе зависит от изменения канала во времени, которое грубо характеризуется временной когерентностью $(\Delta t)_c$ или, что эквивалентно, доплеровским рассеянием B_d.

Влияние канала на передаваемый сигнал $s_l(t)$ зависит от нашего выбора полосы сигнала и его длительности. Для примера, если мы выбираем длительность сигнала T так, чтобы удовлетворить условие $T \gg T_m$, то канал вводит пренебрегаемый уровень межсимвольной интерференции. Если полоса сигнального импульса $s_l(t)$ равна $W \approx 1/T$, то условие $T >> T_m$ предполагает, что

$$W << \frac{1}{T_m} \approx (\Delta f)_c. \qquad (14.2.3)$$

Это значит, что полоса W намного меньше полосы частотной когерентности канала. Следовательно, канал неселективен по частоте. Другими словами, все частотные компоненты $S_l(f)$ подвергаются одинаковым ослаблениям и фазовым сдвигам при передаче по каналу. Но это подразумевает, что внутри полосы $S_l(f)$ переменная во времени передаточная функция $C(f;t)$ канала является комплексной величиной,

постоянной по частотной переменной. Поскольку $S_l(f)$ имеет концентрацию спектра вблизи $f=0$, то существенное значение имеет $C(f;t)=C(0;t)$. Следовательно, (14.2.2) сводится к

$$r_l(t)=C(0;t)\int_{-\infty}^{\infty}S_l(f)e^{j2\pi ft}df=C(0;t)s_l(t). \qquad (14.2.4)$$

Таким образом, когда полоса частот сигнала W намного меньше полосы частотной когерентности $(\Delta f)_c$ канала, принимаемый сигнал равен просто переданному сигналу, умноженному на комплексный случайный процесс $C(0;t)$, который представляет переменную во времени передаточную функцию канала. В этом случае мы видим, что многопутевые компоненты в принимаемом сигнале неразличимы поскольку $W<<(\Delta f)_c$.

Передаточную функцию $C(0;t)$ для неселективного по частоте канала можно выразить в виде

$$C(0;t)=\alpha(t)e^{-j\phi(t)}, \qquad (14.2.5)$$

где $\alpha(t)$ представляет огибающую, а $\phi(t)$ представляет фазу эквивалентного низкочастотного канала. Если $C(0;t)$ комплексный гауссовский случайный процесс с нулевым средним, то $\alpha(t)$ распределено по Релею для любого фиксированного момента t, а $\phi(t)$ имеет равномерное распределение на интервале $(-\pi,\pi)$. Скорость замираний в неселективном по частоте канале определяется или корреляционной функцией $\phi_C(\Delta t)$, или доплеровским спектром мощности $S_C(\lambda)$. Альтернативно, канальные параметры $(\Delta t)_c$ или B_d можно использовать для характеристики скорости замираний.

Для примера, предположим, что возможно выбрать полосу частот W, удовлетворяющую условию $W<<(\Delta f)_c$ и сигнальный интервал T, удовлетворяющий условию $T<<(\Delta t)_c$. Поскольку T меньше чем интервал временной когерентности канала ослабление канала и фазовый сдвиг по существу постоянны по крайней мере на сигнальном интервале. Когда это условие выполняется мы называем канал *каналом с медленными замираниями*.[1] Далее, когда $W\approx 1/T$ условие, что канал неселективен по частоте и с медленными замираниями предполагает, что произведение T_m и B_d должно удовлетворять условию $T_mB_d<1$.

Произведение T_mB_d называют *фактором рассеяния* канала. Если $T_mB_d<1$, канал считается с низким рассеянием, в противном случае – с высоким рассеянием. Многопутевое рассеяние, доплеровское рассеяние и фактор рассеяния даны в табл.14.2.1 для различных каналов. Мы видим из этой таблицы, что различные радиоканалы, включая и канал, образованный отражением от Луны, как от пассивного отражателя, являются каналами с низким рассеянием. Следовательно, возможно выбрать сигнал $s_l(t)$ так, чтобы эти каналы были неселективны по частоте с медленными замираниями. Условие медленности замираний предполагает, что характеристики канала меняются достаточно медленно, так что их можно измерить.

В разделе 14.3 мы хотим определить вероятность ошибки для двоичной передачи по каналу, неселективному по частоте и с медленными замираниями. Эта модель канала является простейшей для анализа. Важнее то, что по характеристикам качества передачи

[1] Пионерская работа по исследованию каналов с быстрыми измерениями параметров принадлежит Е.З. Финкельштейну [73] (прп).

цифровой информации по каналу с замираниями можно подсказать тот вид сигналов, которые успешно преодолевают замирания, вызванные каналом.

Таблица 14.2.1. Многолучевое рассеяние, доплеровское рассеяние и фактор рассеяния для некоторых многолучевых каналов с замираниями

Тип канала	Многолучевое рассеяние	Доплеровское рассеяние	Фактор рассеяния
Коротковолновое ионосферное распространение (ВЧ – HF)	$10^{-3}...10^{-2}$	$10^{-3}...1$	$10^{-4}...10^{-2}$
Ионосферное распространение по возмущённым полярным трассам (ВЧ – HF)	$10^{-3}...10^{-2}$	10...100	$10^{-2}...1$
Дальнее ионосферное распространение (ОВЧ – VHF)	10^{-4}	10	10^{-3}
Тропосферное возмущение (scatter) (КВЧ – SHF)	10^{-6}	10	10^{-5}
Орбитальная связь (X-диапазон)	10^{-4}	10^{3}	10^{-1}
Луна при макс. либрации ($f_0 = 0{,}4$ км/с)	10^{-2}	10	10^{-1}

Поскольку многопутевые компоненты в принимаемом сигнале не различимы, когда полоса частот W меньше полосы частотной когерентности $(\Delta f)_c$ канала, то принимаемый сигнал достигает приёмника по единственному пути с замираниями. С другой стороны, мы можем выбрать $W >> (\Delta f)_c$ так что канал становится селективным по частоте. Мы покажем далее, что при этом условии многопутевые компоненты в принимаемом сигнале различимы и разрешены во времени при задержках порядка $1/W$. Таким образом, мы покажем, что канал с частотной селективностью можно моделировать как трансверсальный фильтр на линии с задержкой с переменными во времени коэффициентами у отдельных ячеек. Затем мы определим качество двоичной системы сигналов по каналу с частотной селективностью.

14.3. КАНАЛ, НЕСЕЛЕКТИВНЫЙ ПО ЧАСТОТЕ С МЕДЛЕННЫМИ ЗАМИРАНИЯМИ

В этом разделе мы определим вероятность ошибки двоичной ФМ и двоичной ЧМ, когда сигналы передаются по неселективному по частоте каналу с медленными замираниями. Как описано в разделе 4.2 частотный неселективный канал приводит к мультипликативному искажению переданного сигнала $s_l(t)$. Далее условие медленности замираний предполагает, что мультипликативный процесс можно считать неизменным, по крайней мере на сигнальном интервале. Следовательно, если передаваемый низкочастотный сигнал $s_l(t)$, то принимаемый эквивалентный низкочастотный сигнал равен на сигнальном интервале

$$r_l(t) = \alpha e^{-j\phi} s_l(t) + z(t), \qquad 0 \le t \le T, \tag{14.3.1}$$

где $z(t)$ представляет комплексный белый гауссовский шумовой процесс, искажающий сигнал.

Предположим, что замирания в канале достаточно медленны, так что фазу ϕ можно оценить по принимаемому сигналу без ошибок. В этом случае, мы можем обеспечить идеальное когерентное детектирование принимаемого сигнала. Таким образом,

принимаемый сигнал можно обработать, пропуская его через согласованный фильтр в случае двоичной ФМ или через пару согласованных фильтров в случае двоичной ЧМ. Один из методов, который мы можем использовать для определения качества двоичных систем связи сводится к расчёту величин для решения и по ним определить вероятность ошибки.

Однако мы это уже делали для фиксированного (инвариантного во времени) канала. Таким образом, для фиксированного ослабления из (5.2.5) следует выражение для вероятности ошибки двоичной ФМ как функции от ОСШ γ_b принимаемого сигнала

$$P_2(\gamma_b) = Q(\sqrt{2\gamma_b}), \tag{14.3.2}$$

где $\gamma_b = \alpha^2 \mathscr{E}_b / N_0$. Выражение для вероятности ошибки двоичной ЧМ при когерентном детектировании определяется (5.2.10) так

$$P_2(\gamma_b) = Q(\sqrt{\gamma_b}) \tag{14.3.3}$$

Мы можем рассматривать (14.3.2) и (14.3.3) как формулы для условной вероятности ошибки при условии, что α фиксировано. Для получения вероятности ошибки, когда α случайная величина, мы должны усреднить $P_2(\gamma_b)$, определяемых (14.3.2) и (14.3.3) по γ_b с ФПВ $p(\gamma_b)$. Таким образом, мы должны вычислить интеграл

$$P_2 = \int_{-\infty}^{\infty} P_2(\gamma_b) p(\gamma_b) d\gamma_b . \tag{14.3.4}$$

Релеевские замирания. Поскольку α распределено по Релею, α^2 имеет хи-квадрат распределение с двумя степенями свободы. Следовательно, γ_b также распределено по закону хи-квадрат. Легко показать, что

$$p(\gamma_b) = \frac{1}{\bar{\gamma}_b} e^{-\gamma_b / \bar{\gamma}_b}, \qquad \gamma_b \geq 0, \tag{14.3.5}$$

$\bar{\gamma}_b$ – среднее значение ОСШ, определяемое так:

$$\bar{\gamma}_b = \frac{\mathscr{E}_p}{N_0} E(\alpha^2). \tag{14.3.6}$$

Слагаемое $E(\alpha^2)$ – среднее значение α^2.

Теперь мы можем подставить (14.3.5) в (14.3.4) и выполнить интегрирование $P_2(\gamma_b)$ с учётом (14.3.2) и (14.3.3). Результат интегрирования для двоичной ФМ

$$P_2 = \frac{1}{2}\left(1 - \sqrt{\frac{\bar{\gamma}_b}{1+\bar{\gamma}_b}}\right). \tag{14.3.7}$$

При когерентном детектировании ЧМ получаем результат для средней вероятности ошибки

$$P_2 = \frac{1}{2}\left(1 - \sqrt{\frac{\bar{\gamma}_b}{1+2\bar{\gamma}_b}}\right). \tag{14.3.8}$$

При получении результата (14.3.7) и (14.3.8) мы предположили, что оценка фазового сдвига в канале, полученная при медленных замираниях, безошибочная. Такое идеальное условие может не выполняться на практике. В этом случае выражения (14.3.7) и (14.3.8) следует рассматривать как представляющие наилучшее достижимое качество при замираниях в канале. В приложении С мы рассмотрим проблему оценивания фазы в присутствии шума и определим вероятность ошибки для двоичной многопозиционной ФМ.

В каналах, в которых замирания достаточно быстрые для того, чтобы обеспечить стабильную оценку фазы путем усреднения фазы принимаемого сигнала по многим сигнальным интервалам, альтернативный метод передачи является ДФМ. Поскольку ДФМ требует стабильности фазы только по двум соседним сигнальным интервалам эта техника модуляции достаточно проста при наличии замираний сигнала. Чтобы рассчитать качество двоичной ДФМ для канала с замиранием начнем снова с вероятности ошибки для канала без замирания, которая равна

$$P_2(\gamma_b)=\tfrac{1}{2}e^{-\gamma_b}. \qquad (14.3.9)$$

Это выражение подставляется в интеграл (14.3.4) вместе с $p(\gamma_b)$, определяемым (14.3.5). Вычисление интеграла дает среднюю вероятность ошибки для двоичной ДФМ в виде

$$P_2=\frac{1}{2(1+\bar{\gamma}_b)}. \qquad (14.3.10)$$

Если мы вообще не хотим заниматься оценкой сдвига фазы и вместо этого будем использовать некогерентное (по огибающей или квадратичное) детектирование и двоичные ортогональные сигналы ЧМ, вероятность ошибки в канале без замираний

$$P_2(\gamma_b)=\tfrac{1}{2}e^{-\gamma_b/2}. \qquad (14.3.11)$$

Если усреднить $P_2(\gamma_b)$ по ослаблению в канале с релеевским распределением, то получаем для средней вероятности ошибки

$$P_2=\frac{1}{2+\bar{\gamma}_b} \qquad (14.3.12)$$

Кривые вероятности ошибки (14.3.7), (14.3.8), (14.3.10) и (14.3.12) даны на рис.14.3.1.

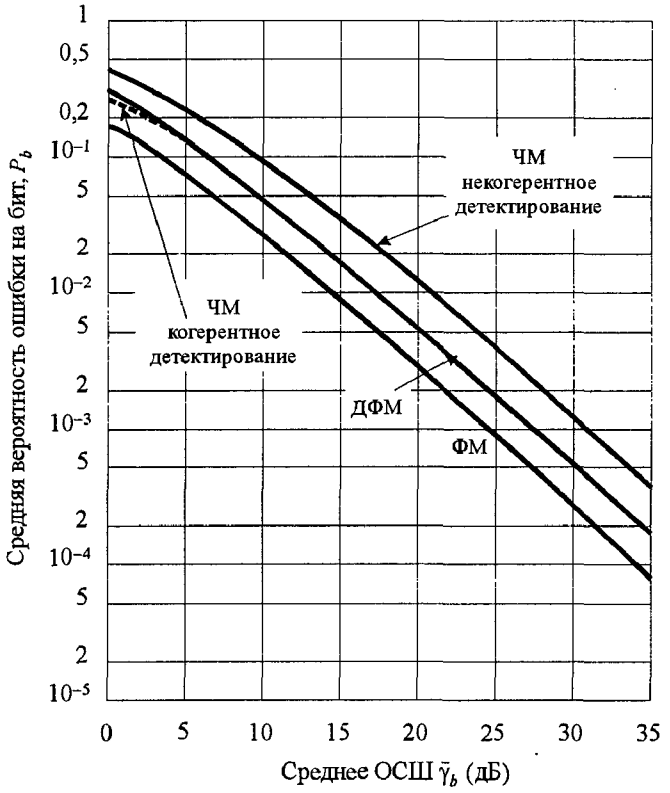


Рис. 14.3.1. Характеристики двоичной передачи по релеевскому замирающему каналу

При сравнении качества четырех двоичных систем сигналов сконцентрируем внимание на вероятности ошибки при больших ОСШ, т.е. $\bar{\gamma}_b >> 1$. При этом условии, формулы для вероятности ошибки (14.3.7), (14.3.8), (14.3.10) и (14.3.12) упрощаются

$$P_2 \approx \begin{cases} 1/4\bar{\gamma}_b & \text{для когерентной ФМ,} \\ 1/2\bar{\gamma}_b & \text{для когерентной ортогональной ЧМ,} \\ 1/2\bar{\gamma}_b & \text{для ОФМ,} \\ 1/\bar{\gamma}_b & \text{для некогерентной ортогональной ЧМ.} \end{cases} \qquad (14.3.13)$$

Из (14.3.13) мы видим, что когерентная ФМ на 3 дБ лучше, чем ДФМ и на 6 дБ лучше некогерентной ЧМ. Более удивительным, однако, является наблюдение, что вероятность ошибки уменьшается только обратно ОСШ. В противоположность этому уменьшение вероятности ошибки в канале без замираний зависит экспоненциально от ОСШ. Это означает, что по каналу с замираниями передатчик должен передавать большие уровни мощности для достижения низкой вероятности ошибки. Во многих случаях большие уровни мощности невозможны технически и (или) экономически. Альтернативное решение проблемы получения приемлемого качества по каналу с замираниями сводится к использованию техники разнесения, обсуждаемой в разделе 14.4.

Замирания по Накагами. Если α характеризуется статистически m-распределением Накагами, то случайная величина $\gamma = \alpha^2 \mathcal{E}_b / N_0$ имеет ФПВ (смотри задачу 14.15)

$$p(\gamma) = \frac{m^m}{\Gamma(m)\bar{\gamma}^m} \gamma^{m-1} e^{-m\gamma/\bar{\gamma}}, \qquad (14.3.14)$$

где $\bar{\gamma} = E(\alpha^2)\mathcal{E}/N_0$.

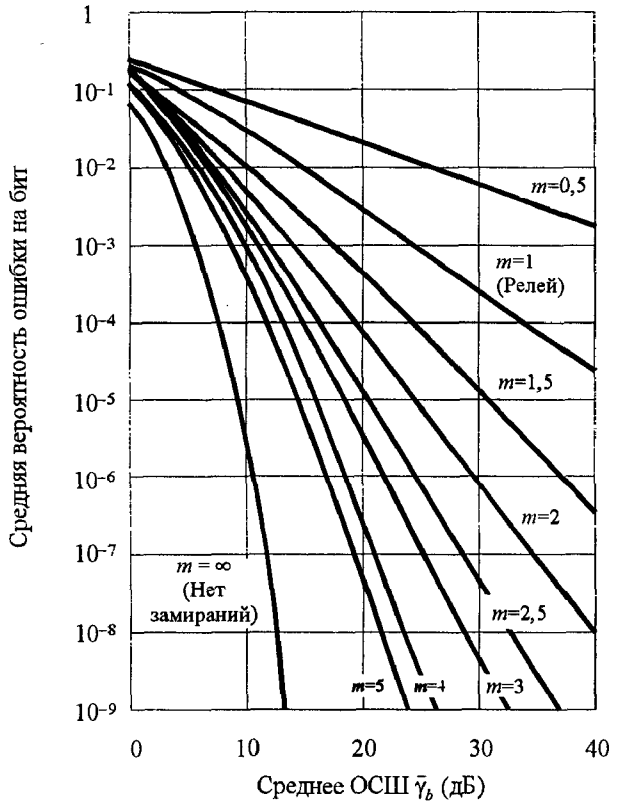


Рис. 14.3.2. Средняя вероятность ошибки для двоичной ФМ при одиночном (неразнесённом) приёме

Среднюю вероятность ошибки для любого метода модуляции легко получить усреднением соответствующей вероятности ошибки в канале без замираний по статистике замирающего сигнала.

В качестве примера расчёта качества, полученного с m-распределением Накагами для канала с замираниями, рис.14.3.2 иллюстрирует вероятность ошибки двоичной ФМ с m как параметром. Напомним, что $m = 1$ соответствует релеевским замираниям. Видим, что качество растет по мере роста m относительно $m = 1$, что говорит о том, что при этом замирания менее суровые. С другой стороны, если $m < 1$, качество хуже, чем при релеевских замираниях.

Другая статистика замираний. Следуя процедуре, описанной выше, можно определить качество при различных методах модуляции при других видах статистики замирающих сигналов, такая, например, как распределение Райса.

Вероятность ошибки для замираний со статистикой Райса можно найти в статье Линдсея (1964), в то время как расчёты со статистикой Накагами читатель может найти в статьях Эспозито (1967), Майгаки и др. (1979), Чараша (1979), Ол Хасини и др. (1985), и Боли и др. (1991)

14.4. ТЕХНИКА РАЗНЕСЕНИЯ ДЛЯ МНОГОПУТЕВЫХ КАНАЛОВ С ЗАМИРАНИЯМИ

Техника разнесения основывается на представлении, что ошибки возникают при приеме, когда ослабление в канале велико, т.е. когда в канале глубокие замирания. Если мы можем подать на приёмник несколько сигналов, переданных по независимым каналам с замиранием с той же информацией, вероятность того, что все сигналы замрут одновременно существенно уменьшается. Это значит, если p – вероятность того, что уровень одного сигнала снизится ниже некоторой критической величины, то p^L – вероятность того, что уровень всех L независимо замирающих сигналов снизится ниже этой величины. Имеется несколько путей, при помощи которых мы можем обеспечить приёмник L независимыми версиями одного и того же информационного сигнала.

Один метод сводится к использованию *частотного разнесения*. В этом случае одинаковый информационный сигнал передается по L несущим, причем разнос между соседними частотами равен или превосходит полосу частотной когерентности канала $(\Delta f)_c$.

Второй метод для получения L независимых замирающих версий одного и того же информационного сигнала сводится к передаче сигнала по L различным временным отрезкам, причем разнос между соседними отрезками равен или превосходит интервал временной когерентности канала $(\Delta t)_c$. Этот метод называется *временным разнесением*.

Заметим, что канал с замираниями подходит к модели канала с группированием ошибок. Далее мы можем рассматривать передачу одинаковой информации на различных частотах или в различных временных интервалах (или обоими способами), как простую форма кодирования с повторением. Передача с разнесением во времени на интервале $(\Delta t)_c$ или по частоте на интервале $(\Delta f)_c$ по существу является формой блокового перемежения символов при использовании кода с повторением с попыткой разорвать пакеты ошибок и, таким образом, получить независимые ошибки. Позже мы покажем, что в общем коды с повторением (тривиальное кодирование) расточительны по полосе частот, если их сравнить с нетривиальным кодированием.

Другой обычно используемый метод для достижения разнесения использует множество антенн. Для примера, мы можем использовать одну передающую антенну и множество приёмных антенн. Последние должны быть достаточно разнесены в пространстве для того, чтобы многопутевые компоненты сигнала имели бы существенно различное *время задержки* до антенн. Обычно требуется разнос между двумя антеннами по крайней мере на 10 длин волны для того, чтобы получить сигналы, которые замирают независимо.

Более изощренный метод для получения разнесения основывается на использовании сигнала с полосой частот, намного большей, чем полоса когерентности $(\Delta f)_c$ канала. Такой сигнал с полосой W позволяет разделять многопутевые компоненты и, таким образом, обеспечит приёмник несколькими независимо замирающими сигналами.

Различение во времени равно $1/W$. Следовательно, при времени рассеяния T_m имеется $T_m W$ разделимых компонент. Поскольку $T_m \approx 1/(\Delta f)_c$, то число разделимых сигнальных компонент можно также выразить так $W/(\Delta f)_c$. Таким образом использование широкополосного сигнала можно рассматривать как другой метод получения частотного разнесения порядка $L \approx W/(\Delta f)_c$. Оптимальный приёмник для обработки широкополосного сигнала будет рассмотрен в разделе 14.5. Его назвали *RAKE* (Рэйк) *коррелятором* или *RAKE согласованным фильтром* и он был разработан Прайсом и Грином (1958).

Имеется и другая техника разнесения, которая используется на практике, такая как угловое разнесение и поляризационное разнесение. Однако они не используются так широко, как те, которые описаны выше.

14.4.1. Двоичные сигналы

Теперь определим вероятность ошибки для двоичной системы связи с разнесением. Начнем с описания математической модели для системы связи с разнесением. Прежде всего предположим, что имеется L каналов разнесения с сигналами, носящими одинаковую информацию. Каждый из каналов считается неселективным по частоте, с медленными замираниями и с распределенной по Релею огибающей сигнала. Процессы замираний в L каналов разнесения считаются статистически независимыми. Сигнал в каждом канале искажается белым гауссовским случайным процессом с нулевым средним. Шумовые процессы в L каналах считаются взаимно статистически независимыми и с одинаковыми автокорреляционными функциями. Эквивалентные низкочастотные принимаемые сигналы для L каналов можно представить в виде

$$r_{lk} = \alpha_k e^{-j\phi_k} s_{km}(t) + z_k(t), \qquad k = 1, 2, \ldots, L, \qquad m = 1, 2, \tag{14.4.1}$$

где $\{\alpha_k e^{-j\phi_k}\}$ представляют множители ослабления и фазовые сдвиги в L каналах, $s_{km}(t)$ обозначает m-й сигнал переданный по k-му каналу, а $z(t)$ обозначает аддитивный белый гауссовский шум в k-ом канале. Все сигналы ансамбля $s_{km}(t)$ имеют одинаковую энергию.

Оптимальный демодулятор для сигнала в k-м канале состоит из двух согласованных фильтров. Один имеет импульсную характеристику

$$b_{k1}(t) = s_{k1}^*(T - t), \tag{14.4.2}$$

а второй имеет импульсную характеристику

$$b_{k2}(t) = s_{k2}^*(T - t). \tag{14.4.3}$$

Конечно, если используется двоичная ФМ, то $s_{k1}(t) = -s_{k2}(t)$. Следовательно, для двоичной ФМ требуется один согласованный фильтр. После согласованных фильтров следует устройство сложения, в котором формируются две величины для принятия решения. В устройстве сложения, обеспечивающие наилучшее качество, выход каждого согласованного фильтра умножается на соответствующий комплексный (сопряжённый) канальный множитель $\alpha_k e^{j\phi_k}$. Цель этого умножения – компенсировать фазовый сдвиг в канале и взвесить сигналы множителем, пропорциональным уровню сигнала. Таким образом сильный сигнал получает больший вес, чем слабый. После того, как операция комплексного взвешивания закончена, формируются две суммы. Одна содержит взвешенный выход фильтров и соответствует передаче 0. Вторая содержит взвешенный выход фильтров и соответствует передаче 1. Такое оптимальное сложение называют *сумматором максимальных отношений* Бреннана (1959). Конечно, реализация такого сумматора основана на предположении, что канальные ослабления $\{\alpha_k\}$ и фазовые сдвиги

$\{\phi_k\}$ известны точно. (Влияние шума оценок на вероятность ошибки многопозиционной ФМ рассмотрено в приложении С). Блок-схема, иллюстрирующая модель двоичной системы связи, описанной выше, показана на рис.14.4.1.

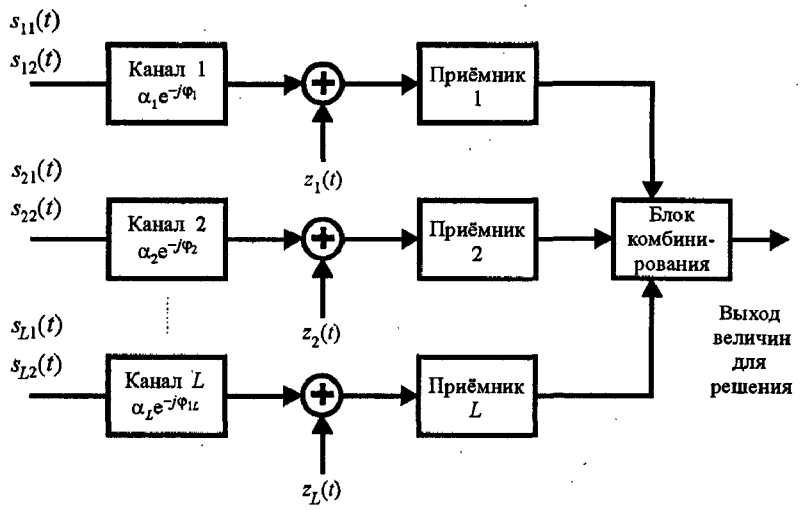


Рис. 14.4.1. Модель двоичной цифровой системы связи с разнесением

Сначала рассмотрим качество двоичной ФМ с L-кратным разнесением. Выход сумматора максимальных отношений можно выразить через единственную величину решения в виде

$$U = \mathrm{Re}\left(2\mathscr{E}\sum_{k=1}^{L}\alpha_k^2 + \sum_{k=1}^{L}\alpha_k N_k\right) = 2\mathscr{E}\sum_{k=1}^{L}\alpha_k^2 + \sum_{k=1}^{L}\alpha_k N_{kr}, \qquad (14.4.4)$$

где N_{kr} означает реальную часть от комплексной шумовой гауссовской случайной величины

$$N_k = e^{j\phi_k}\int_0^T z_k(t)s_k^*(t)dt. \qquad (14.4.5)$$

Будем следовать подходу, использованному в разделе 14.3 при расчете вероятности ошибки. Это значит, что сначала находим вероятность ошибки при фиксированных значениях множителей ослабления $\{\alpha_k\}$. Затем эта условная вероятность ошибки усредняется по $\{\alpha_k\}$.

Релеевские замирания. При фиксированных значениях параметров $\{\alpha_k\}$ величина для решения U является случайной гауссовской со средним

$$E(U) = 2\mathscr{E}\sum_{k=1}^{L}\alpha_k^2 \qquad (14.4.6)$$

и дисперсией

$$\sigma_U^2 = 2\mathscr{E}N_0\sum_{k=1}^{L}\alpha_k^2. \qquad (14.4.7)$$

При таких значениях среднего и дисперсии вероятность того, что $U<0$, равна

$$P_2(\gamma_b) = Q\left(\sqrt{2\gamma_b}\right) \qquad (14.4.8)$$

где ОСШ на γ_b определяется так

$$\gamma_b = \frac{\mathscr{E}}{N_0}\sum_{k=1}^{L}\alpha_k^2 = \sum_{k=1}^{L}\gamma_k, \qquad (14.4.9)$$

где $\gamma_k = \mathscr{E}\alpha_k^2/N_0$ – мгновенные ОСШ в k-м канале. Теперь мы должны определить ФПВ $p(\gamma_b)$. Эту функцию существенно проще определить через характеристическую функцию γ_b. Сначала заметим, что при $L=1$ $\gamma_b \equiv \gamma_1$ имеет хи-квадрат плотность вероятности, определенную (14.3.5). Характеристическую функцию γ_1 легко найти

$$\psi_{\gamma_1}(j\upsilon) = E\left(e^{j\upsilon\gamma_1}\right) = \frac{1}{1-j\upsilon\bar{\gamma}_c} . \tag{14.4.10}$$

где $\bar{\gamma}_c$ – среднее значение ОСШ на канал, которое считается одинаковым для всех каналов. Это значит, что

$$\bar{\gamma}_c = \frac{\mathscr{E}}{N_0} E\left(\alpha_k^2\right) \tag{14.4.11}$$

независимо от k. Это предположение используется для всех результатов этого раздела. Поскольку замирания по всем L каналам статистически независимы, $\{\gamma_k\}$ статистически независимы и, следовательно, характеристическая функция суммы равна результату (14.4.10) в L-й степени, т.е.

$$\psi_{\gamma_b}(j\upsilon) = \frac{1}{(1-j\upsilon\bar{\gamma}_c)^L} . \tag{14.4.12}$$

Но эта характеристическая функция случайной величин с хи-квадрат распределением с $2L$ степенями свободы. Из (2.1.107) следует, что ФПВ $p(\gamma_b)$ равна

$$p(\gamma_b) = \frac{1}{(L-1)!\bar{\gamma}_c^L}\gamma_b^{L-1}e^{-\gamma_b/\bar{\gamma}_c} . \tag{14.4.13}$$

Заключительная ступень расчета сводится к усреднению условной вероятности ошибки (14.4.8) по статистике замираний, т.е. к вычислению интеграла.

$$P_2 = \int_0^\infty P_2(\gamma_b)p(\gamma_b)d\gamma_b . \tag{14.4.14}$$

Имеется замкнутая форма решения (14.4.14), которую можно выразить так

$$P_2 = \left[\frac{1}{2}(1-\mu)\right]^L \sum_{k=0}^{L-1}\binom{L-1+k}{k}\left[\frac{1}{2}(1+\mu)\right]^k , \tag{14.4.15}$$

где по определению

$$\mu = \sqrt{\frac{\bar{\gamma}_c}{1+\bar{\gamma}_c}} . \tag{14.4.16}$$

Когда средняя ОСШ на канал $\bar{\gamma}_c$ удовлетворяет условию $\bar{\gamma}_c >> 1$, слагаемое $\frac{1}{2}(1+\mu) \approx 1$, а слагаемое $\frac{1}{2}(1-\mu) \approx 1/4\bar{\gamma}_c$. Далее

$$\sum_{k=0}^{L-1}\binom{L-1+k}{k} = \binom{2L-1}{L} . \tag{14.4.17}$$

Следовательно, когда $\bar{\gamma}_c$ достаточно велико (больше 10 дБ) вероятность ошибки (14.4.15) можно выразить так

$$P_2 \approx \left(\frac{1}{4\bar{\gamma}_c}\right)^L \binom{2L-1}{L} . \tag{14.4.18}$$

Из (14.4.18) мы видим, что вероятность ошибки меняется как $1/\bar{\gamma}_c$ в степени L. Таким образом при разнесении вероятность ошибки уменьшается обратно пропорционально L-й степени ОСШ.

Получив вероятность ошибки для двоичной ФМ при разнесении, теперь обратим наше внимание на двоичную ортогональную ЧМ при когерентном детектировании. В этом случае, две величины для решения на выходе сумматора максимальных отношений можно выразить так

$$U_1 = \mathrm{Re}\left(2\mathscr{E}\sum_{k=1}^{L}\alpha_k^2 + \sum_{k=1}^{L}\alpha_k N_{k1} \right)$$
$$U_2 = \mathrm{Re}\left(\sum_{k=1}^{L}\alpha_k N_{k2} \right), \qquad (14.4.19)$$

причём мы предположили, что был передан сигнал $s_{k1}(t)$, а $\{N_{k1}\}$ и $\{N_{k2}\}$ являются двумя ансамблями шумовых компонент на выходе согласованных фильтров.

Вероятность ошибки равно вероятности того, что $U_2 > U_1$. Этот расчет подобен расчету для ФМ, исключая того, что мы теперь имеем удвоение мощности шума. Следовательно, когда $\{\alpha_k\}$ фиксированы, условная вероятность ошибки равна для ЧМ

$$P_2(\gamma_b) = Q\left(\sqrt{\gamma_b}\right). \qquad (14.4.20)$$

Используем (14.4.13) для усреднения $P_2(\gamma_b)$ по статистике замираний. Не удивительно, что мы найдем, что результат (14.4.15) остается в силе с заменой $\bar{\gamma}_c$ на $\frac{1}{2}\bar{\gamma}_c$. Т.е., (14.4.15) определяет вероятность ошибки для двоичной ортогональной ЧМ при когерентном детектировании с параметром μ, равным

$$\mu = \sqrt{\frac{\bar{\gamma}_c}{2+\bar{\gamma}_c}}. \qquad (14.4.21)$$

Далее, для больших значений $\bar{\gamma}_c$ вероятность P_2 можно выразить так

$$P_2 \approx \left(\frac{1}{2\bar{\gamma}_c}\right)^L \binom{2L-1}{L}. \qquad (14.4.22)$$

Сравнивая (14.4.22) и (14.4.18), видим, что разница в 3 дБ по ОСШ между ФМ и ортогональной ЧМ при когерентном детектировании, существующая в канале без замираний и без рассеяния, остается такой же в канале с замираниями.

В проведенном выше обследовании двоичных систем ЧМ и ФМ при когерентном детектировании, мы предположили, что на приеме используются свободные от шума оценки комплексных канальных параметров $\{\alpha_k e^{-j\phi_k}\}$. Поскольку канал меняется во времени параметры $\{\alpha_k e^{-j\phi_k}\}$ нельзя оценить точно. Действительно, в некоторых каналах изменения во времени могут быть достаточно быстрыми и препятствовать применению когерентного детектирования. В этом случае мы рассмотрим использование либо ДФМ или ЧМ с некогерентным детектированием.

Сначала рассмотрим ДФМ. Чтобы использовать ДФМ изменения в канале должны быть достаточно медленными для того, чтобы фазовые сдвиги $\{\phi_k\}$ не менялись заметно на двух соседних сигнальных интервалах. В нашем анализе мы предположим, что канальные параметры $\{\alpha_k e^{-j\phi_k}\}$ остаются постоянными на два соседних сигнальных интервала. Таким образом сумматор для двоичных ДФМ выдает на выходе величину для решения

$$U = \mathrm{Re}\left[\sum_{k=1}^{L} \left(2\mathscr{E}\alpha_k e^{-j\phi_k} + N_{k2}\right)\left(2\mathscr{E}\alpha_k e^{j\phi_k} + N_{k1}^*\right) \right], \qquad (14.4.23)$$

где $\{N_{k1}\}$ и $\{N_{k2}\}$ означают принимаемые шумовые компоненты на выходе согласованных фильтров на двух соседних сигнальных интервалах. Вероятность ошибки равна вероятности того, что $U<0$. Поскольку U – это частный случай общей квадратичной формы комплексных гауссовских случайных величин, обсуждаемой в приложении В, вероятность ошибки можно найти непосредственно из результатов, данных в этих приложениях. Альтернативно, мы можем использовать вероятность ошибки, определяемую (12.1.3), для сигналов двоичной ДФМ, передаваемых по L неизменных во времени каналам, и усреднить её по релеевской статистике замирающего сигнала. Таким образом мы имеем условную вероятность ошибки

$$P_2(\gamma_b)=\left(\frac{1}{2}\right)^{2L-1} e^{-\gamma_b}\sum_{k=0}^{L-1} b_k\gamma_b^2, \tag{14.4.24}$$

где γ_b определяется (14.4.9) и

$$b_k=\frac{1}{k!}\sum_{n=0}^{L-1-k}\binom{2L-1}{n}. \tag{14.4.25}$$

Усреднение $P_2(\gamma_b)$ с учетом ФПВ $p(\gamma_b)$, определяемый (14.4.13), легко приводит к результату

$$P_2=\frac{1}{2^{2L-1}(L-1)!(1+\bar{\gamma}_c)^L}\sum_{k=0}^{L-1} b_k(L-1+k)!\left(\frac{\bar{\gamma}_c}{1+\bar{\gamma}_c}\right)^k. \tag{14.4.26}$$

Отметим, что результат (14.4.26) можно преобразовать к виду (14.4.15), который применим также для когерентной ФМ и ЧМ. Для двоичной ДФМ параметр μ в (14.4.15) определяется так (смотри приложение С)

$$\mu=\frac{\bar{\gamma}_c}{1+\bar{\gamma}_c}. \tag{14.4.27}$$

Для $\bar{\gamma}_c>>1$ вероятность ошибки в (14.4.26) можно аппроксимировать выражением

$$P_2\approx\left(\frac{1}{2\bar{\gamma}_c}\right)^L\binom{2L-1}{L}. \tag{14.4.28}$$

Последний тип сигналов, который мы рассмотрим в этом разделе, – это сигналы двоичной ЧМ с некогерентным детектированием. Такие сигналы подходят как при медленных, так и при быстрых замираниях. Однако анализ качества, проведенный ниже, основывается на предположении, что замирания достаточно медленные, так что канальные параметры $\{\alpha_k e^{-j\phi_k}\}$ остаются постоянными на сигнальном интервале. Сумматоры многоканальных сигналов являются сумматорами квадратов выходов согласованных фильтров. Их выходы определяют две величины для решения

$$\begin{aligned} U_1&=\sum_{k=1}^{L}\left|2\&\alpha_k e^{-j\phi_k}+N_{k1}\right|^2 \\ U_2&=\sum_{k=1}^{L}\left|N_{k2}\right|^2, \end{aligned} \tag{14.4.29}$$

причём считается, что U_1 содержит полезный сигнал. Вероятность ошибки это вероятность того, что $U_2>U_1$. Как и в случае ДФМ, мы имеем выбор из двух подходов в определении качества ЧМ при квадратичном сложении. В разделе 12.1 мы указали, что выражение для вероятности ошибки двоичной ЧМ при квадратичном сложении сигналов такое же как для ДФМ с заменой γ_b на $\frac{1}{2}\gamma_b$. Это значит, что двоичная ЧМ требует

дополнительно увеличение ОСШ на 3 дБ для достижения неизменного качества в канале с постоянными параметрами. Следовательно, условная вероятность ошибки ДФМ, определяемая (14.4.24), остается при квадратичном сложении справедливой и для ЧМ если γ_b заменить на $\frac{1}{2}\gamma_b$. Результат, полученный при усреднении (14.2.24) по статистике замираний и даваемый (14.4.26) также применим для ЧМ с заменой $\bar{\gamma}_c$ на $\frac{1}{2}\bar{\gamma}_c$. Но мы также установили прежде, что (14.4.26) и (14.4.15) эквивалентны. Следовательно вероятность ошибки, даваемое (14-4-15) также справедлива при квадратичном сложении ЧМ с параметром μ, определяемом так

$$\mu = \frac{\bar{\gamma}_c}{2+\bar{\gamma}_c}. \qquad (14.4.30)$$

Альтернативный подход, использованный Пирсом (1958) для получения вероятности того, что $U_2 > U_1$ так же прост, как метод, описанный выше. Он начинается с ФПВ $p(U_1)$ и $p(U_2)$. Поскольку комплексные случайные величины $\{\alpha_k e^{-j\phi_k}\}$, $\{N_{k1}\}$ и $\{N_{k2}\}$ распределены по Гауссу с нулевыми средними, величины для решения U_1 и U_2 распределены согласно хи-квадрат распределению с $2L$ степенями свободы. Это значит

$$p(U_1) = \frac{1}{(2\sigma_1^2)^L (L-1)!} U_1^{L-1} \exp\left(-\frac{U_1}{2\sigma_1^2}\right), \qquad (14.4.31)$$

где

$$\sigma_1^2 = \frac{1}{2} E\left(\left|2\mathscr{E}\alpha_k e^{-j\phi_k} + N_{k1}\right|^2\right) = 2\mathscr{E}N_0\left(1+\bar{\gamma}_c\right).$$

Аналогично

$$p(U_2) = \frac{2}{(2\sigma_2^2)^L (L-1)!} U_2^{L-1} \exp\left(-\frac{U_2}{2\sigma_2^2}\right), \qquad (14.4.32)$$

где

$$\sigma_2^2 = 2\mathscr{E}N_0.$$

Вероятность ошибки и есть вероятность того, что $U_2 > U_1$. Читателю оставляется в качестве упражнения показать, что эта вероятность определяется (14.4.15), где μ определяется (14.4.30).

Если $\bar{\gamma}_c >> 1$, вероятность ошибки для ЧМ с квадратичным детектированием можно упростить, как мы это делали раньше для других двоичных многоканальных систем. В этом случае вероятность ошибки хорошо аппроксимируется выражением

$$P_2 \approx \left(\frac{1}{\bar{\gamma}_c}\right)^L \binom{2L-1}{L}. \qquad (14.4.33)$$

Вероятность ошибки для ФМ, ДФМ, ортогональной ЧМ с некогерентным детектированием иллюстрируется на рис.14.4.2 при L=1, 2 и 4.

Качество определяется как функция от среднего ОСШ на бит $\bar{\gamma}_b$, которое связано со средним ОСШ на канал $\bar{\gamma}_c$ формулой

$$\bar{\gamma}_b = L\bar{\gamma}_c. \qquad (14.4.34)$$

Результаты рис.14.4.2 ясно иллюстрируют выгоду разнесения как средство для преодоления тяжелых потерь в ОСШ, вызванные замираниями.

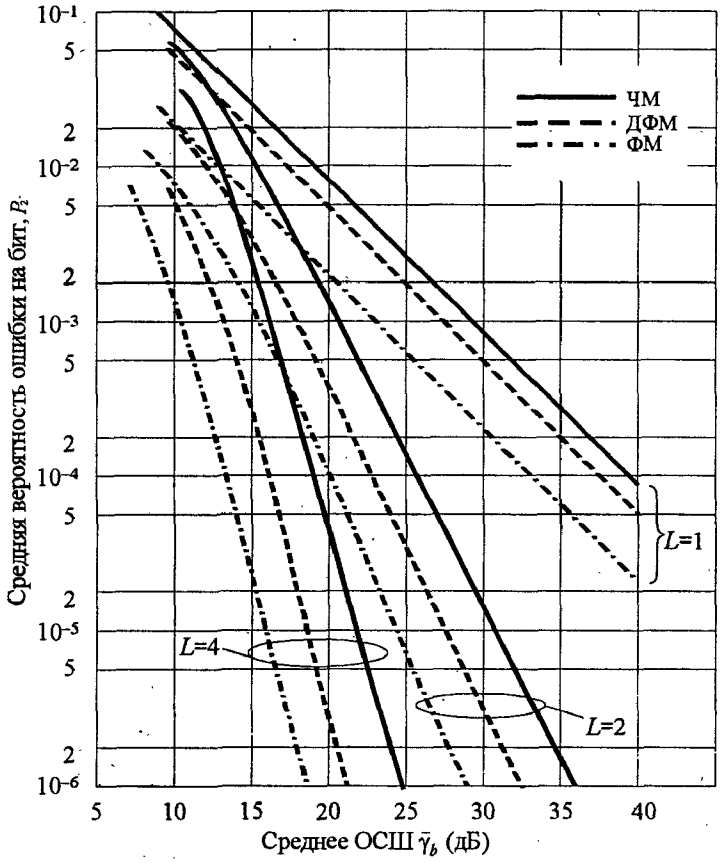


Рис. 14.4.2. Качество двоичных сигналов с разнесением

14.4.2. Многофазные сигналы

В приложении С детально рассматривается передача многофазных сигналов через канал с релеевскими замираниями. Наша основная цель в этом разделе привести общие результаты для вероятности ошибки на символ в M-ичной ФМ и ДФМ и для вероятности ошибки на бит в четырёхфазной ФМ и ДФМ. Общий результат для вероятности ошибки на символ для M-ичной ФМ и ДФМ

$$P_M = \frac{(-1)^{L-1}(1-\mu^2)^L}{\pi(L-1)!}\left(\frac{\partial^{L-1}}{\partial b^{L-1}}\left\{\frac{1}{b-\mu^2}\left[\frac{\pi}{M}(M-1)- \right.\right.\right.$$
$$\left.\left.\left. -\frac{\mu\sin(\pi/M)}{\sqrt{b-\mu^2\cos^2(\pi/M)}}\operatorname{arcctg}\frac{-\mu\cos(\pi/M)}{\sqrt{b-\mu^2\cos^2(\pi/M)}}\right]\right\}\right)_{b=1} \qquad (14.4.35)$$

где

$$\mu = \sqrt{\frac{\bar{\gamma}_c}{1+\bar{\gamma}_c}} \qquad (14.4.36)$$

для когерентной ФМ (для ФМ при когерентном приеме) и

$$\mu = \frac{\bar{\gamma}_c}{1+\bar{\gamma}_c} \qquad (14.4.37)$$

для ДФМ. Опять $\bar{\gamma}_c$ это среднее значение ОСШ на канал. ОСШ на бит $\bar{\gamma}_b = L\bar{\gamma}_c/k$, где $k = \log_2 M$.

Вероятность ошибки на бит для четырехфазной ФМ и ДФМ получена в предположении, что пара информационных бита отображается в четыре фазы согласно коду Грея. Выражение для вероятности ошибки на бит, полученное в приложении С, равно

$$P_b = \frac{1}{2}\left[1 - \frac{\mu}{\sqrt{2-\mu^2}} \sum_{k=0}^{L-1} \binom{2k}{k} \left(\frac{1-\mu^2}{4-2\mu^2}\right)^k\right], \qquad (14.4.38)$$

где μ-снова определяется (14.4.36) и (14.4.37) для ФМ и ДФМ соответственно.

Рис.14.4.3 иллюстрирует вероятность ошибки на символ для ДФМ и когерентной ФМ при $M = 2, 4, 8$ и $L = 1$. Заметим, что разница в качестве между ДФМ и когерентной ФМ равна примерно 3 дБ для всех значений M. Действительно, когда $\bar{\gamma}_b >> 1$ и $L = 1$, (14.4.35) хорошо аппроксимируется так

$$P_M \approx \frac{M-1}{(M \log_2 M)\left[\sin^2(\pi/M)\right]\bar{\gamma}_b} \qquad (14.4.39)$$

для ДФМ и

$$P_M \approx \frac{M-1}{(M \log_2 M)\left[\sin^2(\pi/M)\right]2\bar{\gamma}_b} \qquad (14.4.40)$$

для ФМ. Таким образом, при больших ОСШ когерентная ФМ на 3 дБ лучше, чем ДФМ, в канале с релеевскими замираниями. Эта разница также сохраняется при увеличении L.

Вероятность ошибки на бит отображается на рис.14.4.4 для двухфазовой, четырехфазной и восьмифазной ДФМ при L=1, 2, 4. →

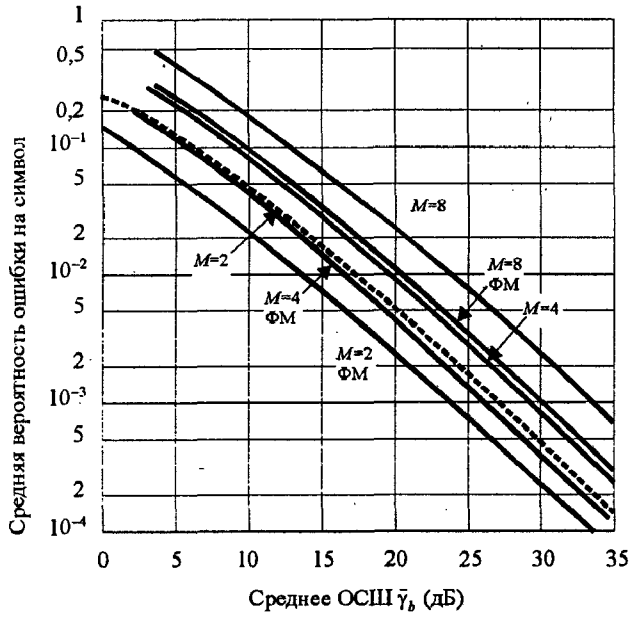


Рис. 14.4.3. Средняя вероятность ошибки на символ для ФМ и ДФМ в релеевском канале

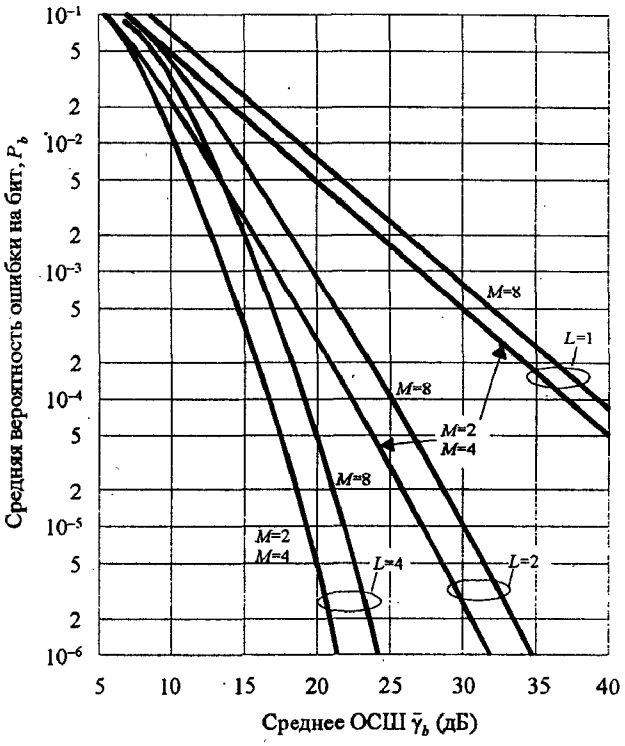


Рис. 14.4.4. Вероятность ошибки на бит для ДФМ в релеевском канале с разнесением

Выражение для вероятности ошибки на бит восьмифазной ДФМ с кодом Грея здесь не дается, но она имеется в статье Прокиса (1968). В этом случае мы видим, что качество для двух- и четырехфазной ДФМ приблизительно одинаково, в то время как восьмифазная ДФМ на 3 дБ хуже . Хотя мы не привели вероятность ошибки на бит для когерентной ФМ, можно показать, что 2- и 4-фазные когерентные ФМ имеет примерно одинаковое качество.

14.4.3. *M*-позиционные ортогональные сигналы

В этом подразделе мы определим качество M-ичной ортогональной системы сигналов, передаваемых по каналу с релеевскими замираниями, и оценим выгоду сигналов с

большим объёмом алфавита M относительно двоичных. Ортогональные сигналы можно рассматривать как сигналы M-ичной ЧМ с минимальным разносом частот $1/T$, где T– сигнальный интервал. Сигнал с той же информацией передается по L разнесенным каналам. Каждый разнесенный канал считается частотно-неселективным и с медленными замираниями, а процесс замираний по L каналам считается статистически взаимно независимым. В каждом канале сигнал искажается аддитивным белым гауссовским шумом. Считаем, что эти аддитивные шумовые процессы статистически независимы .

Хотя относительно легко сформулировать структуру и выполнить анализ качества сумматора с максимальным отношением для разнесенных каналов с M-ичными сигналами, более вероятно, что разработчик выберет некогерентное детектирование. Следовательно, мы обратим наше внимание на квадратичное сложение разнесенных сигналов. Выход сумматора, содержащего сигнал, равен

$$U_1 = \sum_{k=1}^{L} \left| 2\mathscr{E}\alpha_k e^{-j\phi_k} + N_{k1} \right|^2, \tag{14.4.41}$$

в то время как выходы остальных сумматоров

$$U_m = \sum_{k=1}^{L} |N_{km}|^2, \qquad m = 2, 3, 4, \ldots, M. \tag{14.4.42}$$

Вероятность ошибки равно 1 минус вероятность того, что $U_1 > U_m$ для $m = 2, 3, \ldots, M$. Поскольку сигналы ортогональны, а аддитивный шумовой процесс в отдельных каналах разнесения статистически независим, случайные величины $U_1, U_2, \ldots, U_M$ также статистически независимы. ФПВ для U_1 определяется (14.4.31). С другой стороны $U_2, \ldots, U_M$ одинаково распределены с функцией плотности вероятности, определяемой (14.4.32). При фиксированном U_1 совместная вероятность $P(U_2 < U_1, U_3 < U_1, \ldots, U_m < U_1)$ равна вероятности $P(U_2 < U_1)$ в степени $M-1$. Теперь

$$P(U_2 < U_1) = \int_0^{U_1} p(U_2) dU_2 = 1 - \exp\left(-\frac{U_1}{2\sigma_2^2}\right) \sum_{k=0}^{L-1} \frac{1}{k!} \left(\frac{U_1}{2\sigma_2^2}\right)^k, \tag{14.4.43}$$

где $\sigma_2^2 = 2\mathscr{E}N_0$. Чтобы получить вероятность правильного приема, $(M-1)$-я степень от этой вероятности затем усредняется по U_1.

Если мы вычтем этот результат из 1, мы получим вероятность ошибки в виде, данном Ханом (1962):

$$\begin{aligned} P_M &= 1 - \int_0^{\infty} \frac{1}{(2\sigma_1^2)^L (L-1)!} U_1^{L-1} \exp\left(-\frac{U_1}{2\sigma_2^2}\right) \times \\ &\times \left[1 - \exp\left(-\frac{U_1}{2\sigma_2^2}\right) \sum_{k=0}^{L-1} \frac{1}{k!} \left(-\frac{U_1}{2\sigma_2^2}\right)^k \right]^{M-1} dU_1 = \\ &= 1 - \int_0^{\infty} \frac{1}{(1+\bar{\gamma}_c)^L (L-1)!} U_1^{L-1} \exp\left(-\frac{U_1}{1+\bar{\gamma}_c}\right) \times \\ &\times \left(1 - e^{-U_1} \sum_{k=0}^{L-1} \frac{U_1^k}{k!} \right)^{M-1} dU_1, \end{aligned} \tag{14.4.44}$$

где $\bar{\gamma}_c$ – среднее ОСШ на канал разнесения. Среднее ОСШ на бит $\bar{\gamma}_b = L\bar{\gamma}_c / \log_2 M = L\bar{\gamma}_c / k$.

Интеграл в (14.4.44) можно выразить в замкнутой форме через двойную сумму. Это можно показать, если написать

$$\left(\sum_{k=0}^{L-1}\frac{U_1^k}{k!}\right)^m = \sum_{k=0}^{m(L-1)}\beta_{km}U_1^k, \qquad (14.4.45)$$

где β_{km} – набор коэффициентов. Тогда (14.4.44) приводится к виду

$$P_M = \frac{1}{(L-1)!}\sum_{m=1}^{M-1}\frac{(-1)^{m+1}\binom{M-1}{m}}{(1+m+m\bar{\gamma}_c)^L}\cdot\sum_{k=0}^{m(L-1)}\beta_{km}(L-1+k)!\left(\frac{1+\bar{\gamma}_c}{1+m+m\bar{\gamma}_c}\right)^k. \qquad (14.4.46)$$

Если нет разнесения $(L=1)$, вероятность ошибки (14.4.46) приводит к простой форме

$$P_M = \sum_{m=1}^{M-1}\frac{(-1)^{m+1}\binom{M-1}{m}}{1+m+m\bar{\gamma}_c}. \qquad (14.4.47)$$

Вероятность ошибки символа P_M можно преобразовать в эквивалентную вероятность ошибки на бит, умножая P_M на $2^{k-1}/(2^k-1)$.

Хотя выражение для P_M (14.4.46) находится в замкнутой форме, оно затруднительно для вычислений для больших значений M и L.

Альтернативно можно вычислить P_M численным интегрированием, используя выражение (14.4.44). Результаты, иллюстрируемые ниже на графиках, были получены из (14.4.44).

Сначала рассмотрим вероятность ошибки для M-ичной ортогональной системы сигналов с квадратичным сложением, как функция от порядка разнесения. Рис.14.4.5 и 14.4.6 иллюстрируют характеристики P_M для M=2 и 4, как функция L, когда суммарное ОСШ, определенное как $\bar{\gamma}_L = L\bar{\gamma}_c$, остается фиксированным. Эти результаты указывают на то, что имеется оптимальный порядок разнесения для каждого $\bar{\gamma}_L$. Это значит, что для любого $\bar{\gamma}_L$ имеется величина L, при которой P_M минимальна. Тщательное исследование этих графиков обнаруживает, что минимум P_M получается, когда $\bar{\gamma}_c = \bar{\gamma}_t/L \approx 3$. Оказывается, что этот результат не зависит от объёма алфавита M.

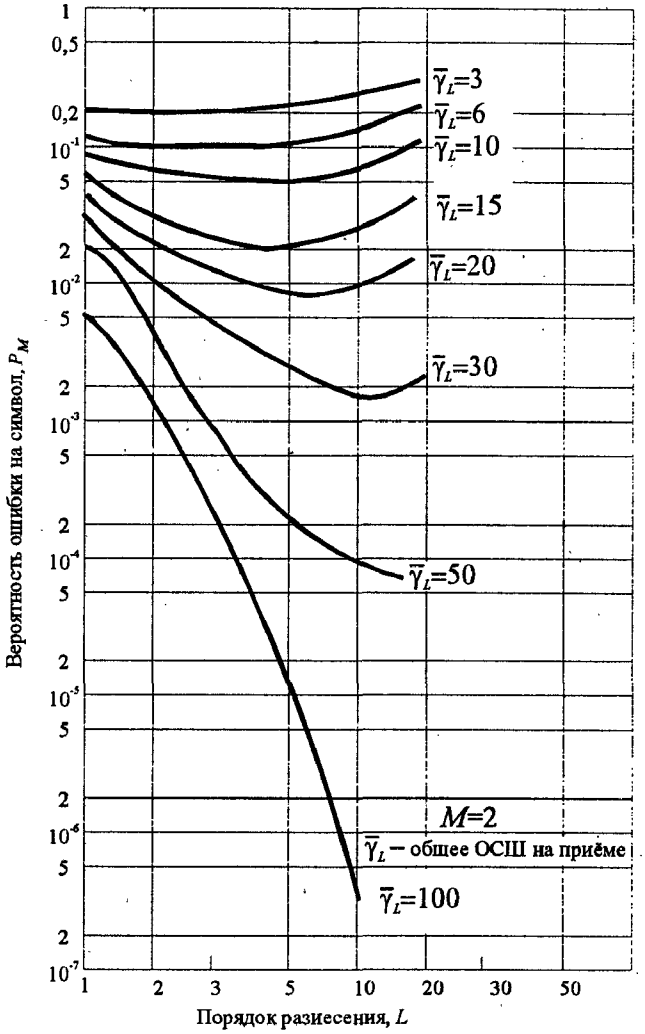


Рис. 14.4.5. Характеристика двоичных ортогональных сигналов при квадратичном детектировании с разнесением

Во-вторых, рассмотрим вероятность ошибки P_M как функцию от среднего ОСШ на бит, определяемого $\bar{\gamma}_b = L\bar{\gamma}_c/k$. (Если мы интерпретируем M-ичную ортогональную ЧМ как форму кодирования[1], а порядок разнесения L как число повторений символа в коде с повторением, тогда $\bar{\gamma}_b = \bar{\gamma}_c/R_c$, где $R_c = k/L$ – скорость кода). Зависимость P_M от $\bar{\gamma}_b$ для $M = 2$, 4, 8, 16, 32 и $L = 1$, 2, 4 показаны на рис. 14.4.7.

Эти результаты иллюстрируют выигрыш в качестве по мере роста M и L. Сначала заметим, что достаточный выигрыш в качестве получается при увеличении L. Второе, мы заметим, что выигрыш в качестве при росте M относительно небольшой при малых L. Однако при увеличении L выигрыш, получаемый с ростом M, также растёт. Поскольку увеличение любого из этих параметров влияет на полосу частот, так как

$$B_c = \frac{LM}{\log_2 M},$$

то результаты, показанные на рис. 14.4.7, указывает на то, что увеличение L более эффективно, чем соответствующее увеличение M.

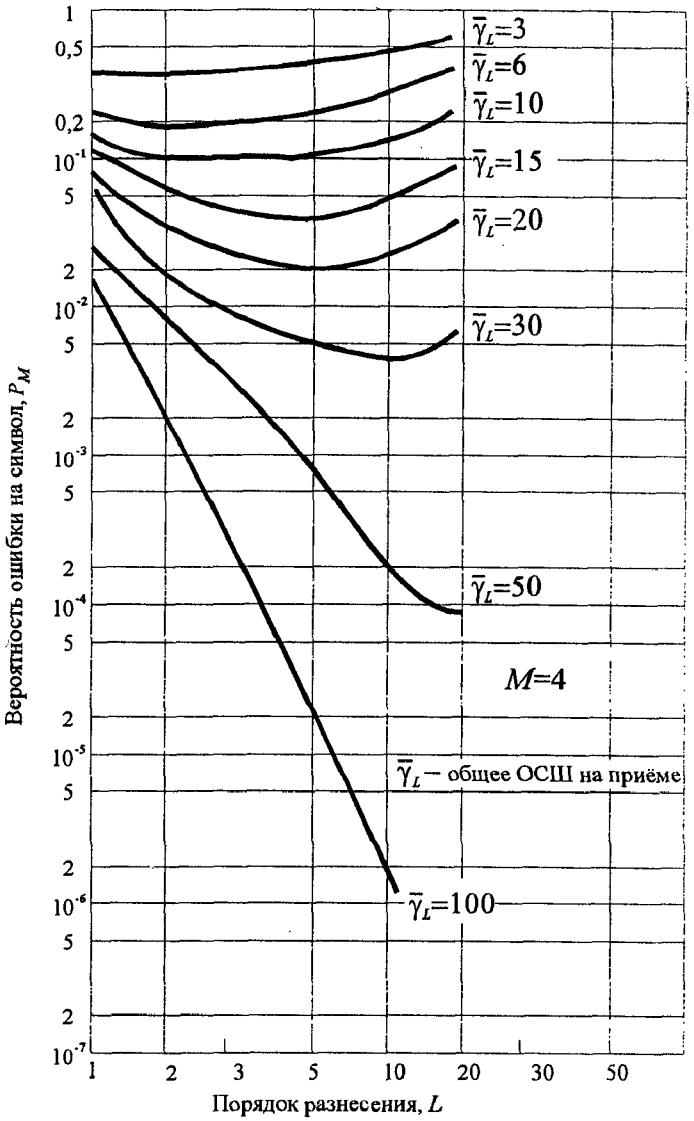


Рис. 14.4.6. Характеристика ортогональных сигналов с $M = 4$ при квадратичном детектировании с разнесением

Как мы увидим в разделе 14.6, кодирование является эффективным по полосе частот средством для получения разнесения сигнала, переданного по каналу с замираниями.

Граница Чернова. Перед окончанием этого раздела, мы определим верхнюю границу Чернова для вероятности ошибки двоичной ортогональной системы сигналов с разнесением L-го порядка, которое будет полезным в нашем обсуждении кодирования для каналов с замираниями, что является предметом раздела 14.6. Наша исходная точка это выражение для двух величин для решения U_1 и U_2, определяемых (14.4.29), где U_1 содержит слагаемые сигнала и шума при квадратном суммировании, а U_2 содержит только слагаемые шума при квадратичном суммировании. Вероятность ошибки двоичной системы сигналов, обозначенная здесь $P_2(L)$, равна

$$P_2(L) = P(U_2 - U_1 > 0) = P(X > 0) = \int_0^\infty p(x)dx, \qquad (14.4.48)$$

где случайная величина X определена так

[1] В разделе 14.6 мы покажем, что M-ичную ортогональную систему ЧМ с разнесением можно рассматривать как блоковый ортогональный код.

$$X = U_2 - U_1 = \sum_{k=1}^{L} \left(|N_{k2}|^2 - |2\mathscr{E}\alpha_k + N_{k1}|^2 \right). \qquad (14.4.49)$$

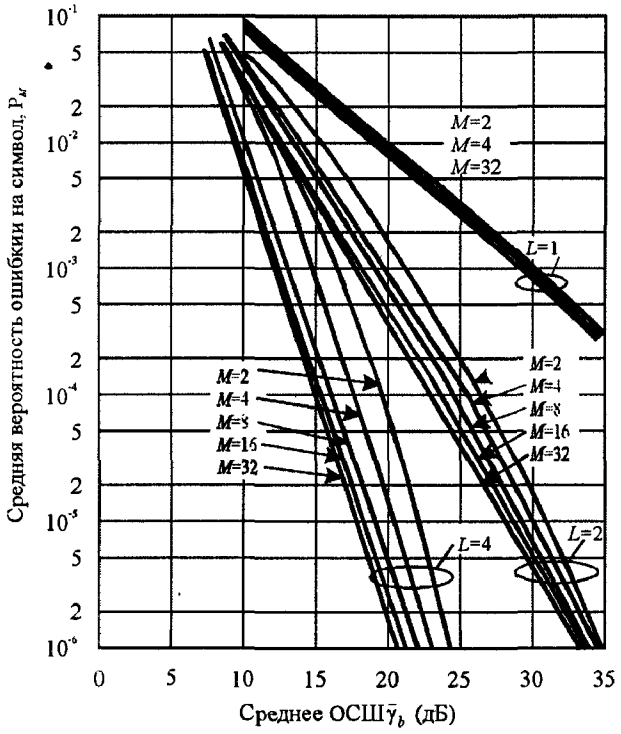


Рис. 14.4.7. Помехоустойчивость передачи с использованием M-позиционных ортогональных сигналов при L-кратном разнесении

Фазовые слагаемые $\{\varphi_k\}$ в U_1 опущены, так как они не влияют на качество при квадратичном детектировании.

Пусть $S(X)$ означает единичную функцию. Тогда вероятность ошибки в (14.4.48) можно выразить в виде

$$P_2(L) = E[S(X)]. \qquad (14.4.50)$$

Согласно разделу 2.1.5, граница Чернова ограничивает сверху единичную функцию экспоненциальной функцией, т.е.

$$S(X) \le e^{\zeta X}, \qquad \zeta \ge 0, \qquad (14.4.51)$$

где параметр ζ оптимизируется для получения плотной границы. Таким образом, мы имеем

$$P_2(L) = E[S(X)] \le E\left(e^{\zeta X}\right). \qquad (14.4.52)$$

Если вместо случайной величины X подставим (14.4.49) и учтём, что случайные слагаемые в сумме статистически независимы, получаем результат

$$P_2(L) \le \prod_{k=1}^{L} E\left(e^{\zeta |N_{k2}|^2}\right) E\left(e^{-\zeta |2\mathscr{E}\alpha_k + N_{k1}|^2}\right). \qquad (14.4.53)$$

Но

$$E\left(e^{\zeta |N_{k2}|^2}\right) = \frac{1}{1 - 2\zeta\sigma_2^2}, \qquad \zeta < \frac{1}{2\sigma_2^2} \qquad (14.4.54)$$

и

$$E\left(e^{-\zeta |2\mathscr{E}\alpha_k + N_{k1}|^2}\right) = \frac{1}{1 + 2\zeta\sigma_1^2}, \qquad \zeta > \frac{-1}{2\sigma_1^2}, \qquad (14.4.55)$$

где $\sigma_2^2 = 2\mathscr{E}N_0$, $\sigma_1^2 = 2\mathscr{E}N_0(1 + \bar{\gamma}_c)$, а $\bar{\gamma}_c$ – среднее ОСШ на канал разнесения. Заметим, что σ_1^2 и σ_2^2 не зависят от k, т.е. параметры аддитивного шума в L разнесённых каналах, как и его распределение, одинаковы. Следовательно (14.4.53) приводит к

$$P_2(L) \le \left[\frac{1}{(1 - 2\zeta\sigma_2^2)(1 + 2\zeta\sigma_1^2)} \right]^2, \qquad 0 \le \zeta \le \frac{1}{2\sigma_2^2}. \qquad (14.4.56)$$

Дифференцируя правую часть (14.4.56) по ζ, находим, что верхняя граница минимизируется, если

$$\zeta = \frac{\sigma_1^2 - \sigma_2^2}{4\sigma_1^2\sigma_2^2}. \qquad (14.4.57)$$

Подстановка (14.4.57) для ζ в (14.4.56) даёт верхнюю границу Чернова в виде

$$P_2(L) \le \left[\frac{4(1+\bar{\gamma}_c)}{(2+\bar{\gamma}_c)^2}\right]^L \qquad (14.4.58)$$

Интересно отметить, что (14.4.58) можно также выразить так

$$P_2(L) \le [4p(1-p)]^L \qquad (14.4.59)$$

где $p = 1/(2+\bar{\gamma}_c)$ – вероятность ошибки для двоичных ортогональных сигналов в канале с замираниями без разнесения.

Сравнение границы Чернова (14.4.58) с точным значением вероятности ошибки для двоичной ортогональной системы сигналов, при квадратичном сложении по L каналам разнесения, определяемым формулой

$$P_2(L) = \left(\frac{1}{1+\bar{\gamma}_c}\right)^L \sum_{k=0}^{L-1} \binom{L-1+k}{k} \left(\frac{1+\bar{\gamma}_c}{2+\bar{\gamma}_c}\right)^k = p^L \sum_{k=0}^{L-1} \binom{L-1+k}{k} (1-p)^k, \qquad (14.4.60)$$

обнаруживает достаточную плотность найденной границы. Рис.14.4.8 иллюстрирует сравнение. Мы видим, что верхняя граница Чернова отличается примерно на 6 дБ от точных значений вероятности ошибки при $L=1$, но по мере роста L она становится плотнее. Для примера, разница между границей и точным значением вероятности ошибки равна примерно 2,5 дБ при $L=4$.

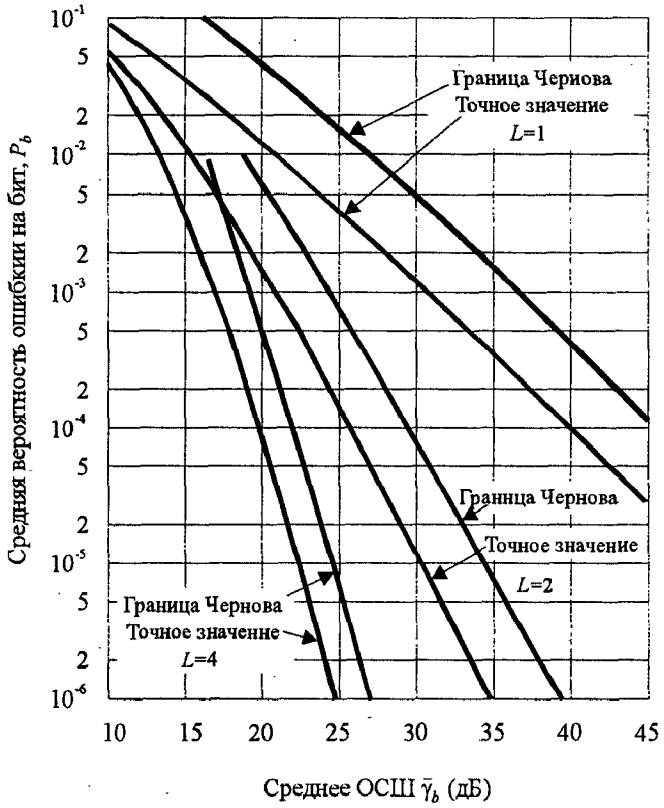


Рис. 14.4.8. Сравиение границы Чернова с точным значением вероятности ошибки

В заключение мы напомним, что вероятность ошибки для M-ичной ортогональной системы сигналов с разнесением можно оценить сверху объединённой границей

$$P_M \le (M-1)P_2(L), \qquad (14.4.61)$$

где $P_2(L)$ определяется или точным значением (14.4.60), или границей Чернова (14.4.58).

14.5. ЦИФРОВАЯ ПЕРЕДАЧА ПО ЧАСТОТНО-СЕЛЕКТИВНОМУ КАНАЛУ С МЕДЛЕННЫМИ ЗАМИРАНИЯМИ

Когда фактор рассеяния канала удовлетворяет условию $T_m B_d << 1$, возможно выбрать сигналы, имеющие полосу $W << (\Delta f)_c$ и длительность $T << (\Delta t)_c$. Это значит, что канал частотно-неселективен и с медленными замираниями. В таком канале можно использовать технику разнесения, чтобы преодолеть тяжёлые последствия замираний.

Если полоса $W >> (\Delta f)_c$, удобно для пользователя разделить канал на определённое число частотно-разделённых (FDM) подканалов, имеющих взаимное расстояние центральных частот по крайней мере $(\Delta f)_c$. Затем один и тот же сигнал можно передать по всем FDM подканалам и, таким образом, получается частотное разнесение. В этом разделе мы опишем альтернативный метод.

14.5.1. Модель канала в виде линии задержки с отводами

Как мы теперь покажем, более прямой метод для достижения по существу того же результата сводится к использованию широкополосного сигнала, заполняющего полосу частот W. Канал по-прежнему считается с медленным замираниями при выполнении условия $T << (\Delta t)_c$. Теперь предположим, что W – это полоса частот, занимаемая реальным полосовым сигналом. Тогда полоса частот, занимаемая эквивалентным низкочастотным сигналом $s_l(t)$, равна $|f| \le \frac{1}{2}W$. Поскольку $s_l(t)$ ограничен по полосе $|f| \le \frac{1}{2}W$, использование теоремы отсчётов приводит к представлению сигнала

$$s_l(t) = \sum_{n=-\infty}^{\infty} s_l\left(\frac{n}{W}\right) \frac{\sin[\pi W(t - n/W)]}{\pi W(t - n/W)} \tag{14.5.1}$$

Преобразование Фурье для $s_l(t)$ равно

$$S_l(f) = \begin{cases} \dfrac{1}{W} \displaystyle\sum_{n=-\infty}^{\infty} s_l(n/W) e^{-j2\pi f n/W} & \left(|f| \le \dfrac{1}{2}W\right) \\ 0 & \left(|f| > \dfrac{1}{2}W\right) \end{cases} \tag{14.5.2}$$

Принимаемый сигнал без шума в частотно-селективном канале был раньше представлен в виде

$$r_l(t) = \int_{-\infty}^{\infty} C(f;t) S_l(f) e^{j2\pi f t} df, \tag{14.5.3}$$

где $C(f;t)$ – переменная во времени передаточная функция канала. Подстановка (14.5.2) для $s_l(t)$ в (14.5.3) даёт

$$r_l(t) = \frac{1}{W} \sum_{n=-\infty}^{\infty} s_l(n/W) \int_{-\infty}^{\infty} C(f;t) e^{j2\pi f(t - n/W)} df = \frac{1}{W} \sum_{n=-\infty}^{\infty} s_l(n/W) c(t - n/W;t), \tag{14.5.4}$$

где $c(\tau;t)$ – переменная во времени импульсная характеристика канала. (14.5.4) имеет форму свёртки. Её можно выразить в альтернативной форме

$$r(t) = \frac{1}{W} \sum_{n=-\infty}^{\infty} s_l(t - n/W) c(n/W;t). \tag{14.5.5}$$

Удобно определить ансамбль переменных во времени коэффициентов канала так

$$c_n(t) = \frac{1}{W} c\left(\frac{n}{W}; t\right). \qquad (14.5.6)$$

Тогда (14.5.5), выраженная через эти коэффициенты канала, принимают вид

$$r(t) = \sum_{n=-\infty}^{\infty} c_n(t) s_l(t - n/W). \qquad (14.5.7)$$

Форма принимаемого сигнала (14.5.7) подразумевает, что переменный во времени частотно-селективный сигнал можно моделировать или представить как линию задержки с отводами, задержка между которыми равна $1/W$, и со взвешивающими коэффициентами $\{c_n(t)\}$. Действительно, мы заключаем из (14.5.7), что низкочастотная импульсная характеристика канала равна

$$c(\tau; t) = \sum_{n=-\infty}^{\infty} c_n(t) \delta(\tau - n/W), \qquad (14.5.8)$$

а соответствующая переменная во времени передаточная функция

$$C(f; t) = \sum_{n=-\infty}^{\infty} c_n(t) e^{-j2\pi f n/W}. \qquad (14.5.9)$$

Итак, при помощи эквивалентного низкочастотного сигнала с полосой $\frac{1}{2}W$, где $W >> (\Delta f)_c$, мы достигаем разрешение во времени $1/W$ по многопутевому профилю запаздывания. Поскольку суммарное многопутевое рассеяние равно T_m, для всех практических целей модель канала на линии задержки можно сконструировать на $L = [T_m W] + 1$ ячеек. Тогда принимаемый сигнал без шума можно выразить в виде

$$r_l(t) = \sum_{n=1}^{L} c_n(t) s_l\left(t - \frac{n}{W}\right). \qquad (14.5.10)$$

Сконструированная модель линии задержки с отводами показана на рис.14.5.1.

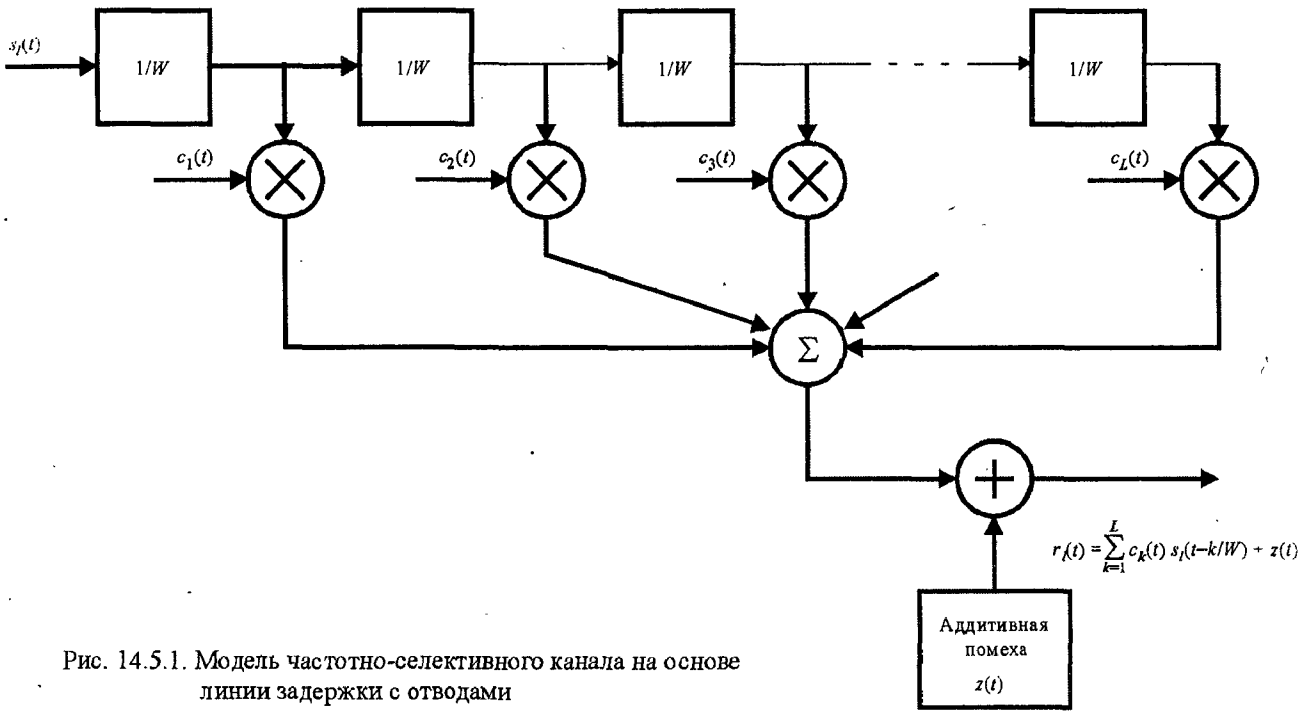


Рис. 14.5.1. Модель частотно-селективного канала на основе линии задержки с отводами

В соответствии со статистической характеристикой, представленной в разделе 14.1, переменные во времени веса в отводах $\{c_n(t)\}$ являются стационарными комплексными случайными процессами. В частном случае релеевских замираний амплитуды $|c_n(t)| \equiv \alpha_n(t)$ распределены по Релею, а фазы $\phi_n(t)$ имеют равномерное распределение. Поскольку $\{c_n(t)\}$ представляет веса отводов, соответствующих L различным задержкам $\tau = n/W$, $n = 1, 2, ..., L$, предположение о некоррелированном рассеянии, сделанное в разделе 7.1, предполагает, что процессы $\{c_n(t)\}$ взаимно некоррелированы. Когда $\{c_n(t)\}$ гауссовские случайные процессы, они также статистически независимы.

14.5.2. RAKE демодулятор

Теперь рассмотрим проблему цифровой передачи по частотно-селективному каналу, который моделируется линией задержки с отводами и со статистически независимыми, меняющимися во времени, весами $\{c_n(t)\}$. Очевидно, однако, в самом начале, что модель линии задержки с отводами со статистически независимыми весами отводов даёт нам L образцов одного и того же переданного сигнала. Следовательно, приёмник, который обрабатывает принимаемый сигнал оптимальным образом, может достичь качества эквивалентной системы связи с разнесением L-го порядка.

Рассмотрим двоичную передачу по каналу. Мы имеем два сигнала равной энергии $s_{l1}(t)$ и $s_{l2}(t)$, которые или противоположные или ортогональные. Их длительность T выбирается так, чтобы удовлетворять условию $T >> T_m$. Таким образом, мы можем пренебречь межсимвольной интерференцией, обусловленной многопутёвостью. Поскольку полоса сигнала превышает полосу частотной когерентности канала, принимаемый сигнал выражается так

$$r_l(t) = \sum_{k=1}^{L} c_k(t) s_{li}(t - k/W) + z(t) = \upsilon_i(t) + z(t), \qquad 0 \le t \le T, \qquad i = 1, 2, \quad (14.5.11)$$

где $z(t)$ – комплексный белый гауссовский шумовой процесс с нулевым средним. Предположим на время, что веса канальных ячеек известны. Тогда оптимальный приёмник состоит из двух фильтров, согласованных с $\upsilon_1(t)$ и $\upsilon_2(t)$, за ними следуют стробирующие устройства и блок решения, который выбирает сигнал, соответствующий наибольшему выходу. Эквивалентный оптимальный приёмник использует взаимную коррекцию вместо согласованной фильтрации. В любом случае величины для решения при когерентном детектировании двоичных сигналов выражаются так

$$U_m = \mathrm{Re}\left[\int_0^T r_l(t) v_m^*(t) dt\right] = \mathrm{Re}\left[\sum_{k=1}^{L} \int_0^T r_l(t) c_k^*(t) s_{lm}^*(t - k/W) dt\right], \qquad m = 1, 2 \quad (14.5.12)$$

Рис.14.5.2 иллюстрирует операции, выполняемые при расчёте величин для решения.

В этой реализации оптимального приёмника два опорных сигнала задерживаются и коррелируются с принимаемым сигналом $r_l(t)$.

Альтернативная реализация оптимального приёмника сводится к использованию единственной линии задержки через которую пропускается принимаемый сигнал $r_l(t)$. В каждом отводе этот сигнал коррелируется с $c_k(t) s_{lm}^*(t)$, где $k = 1, 2, ..., L$ и $m = 1, 2$. Структура приёмника показана на рис.14.5.3. По своему действию приёмник с линией задержки пытается собрать энергию сигнала со всех принимаемых сигнальных путей, которые находятся в области линии задержки и несут одинаковую информацию.

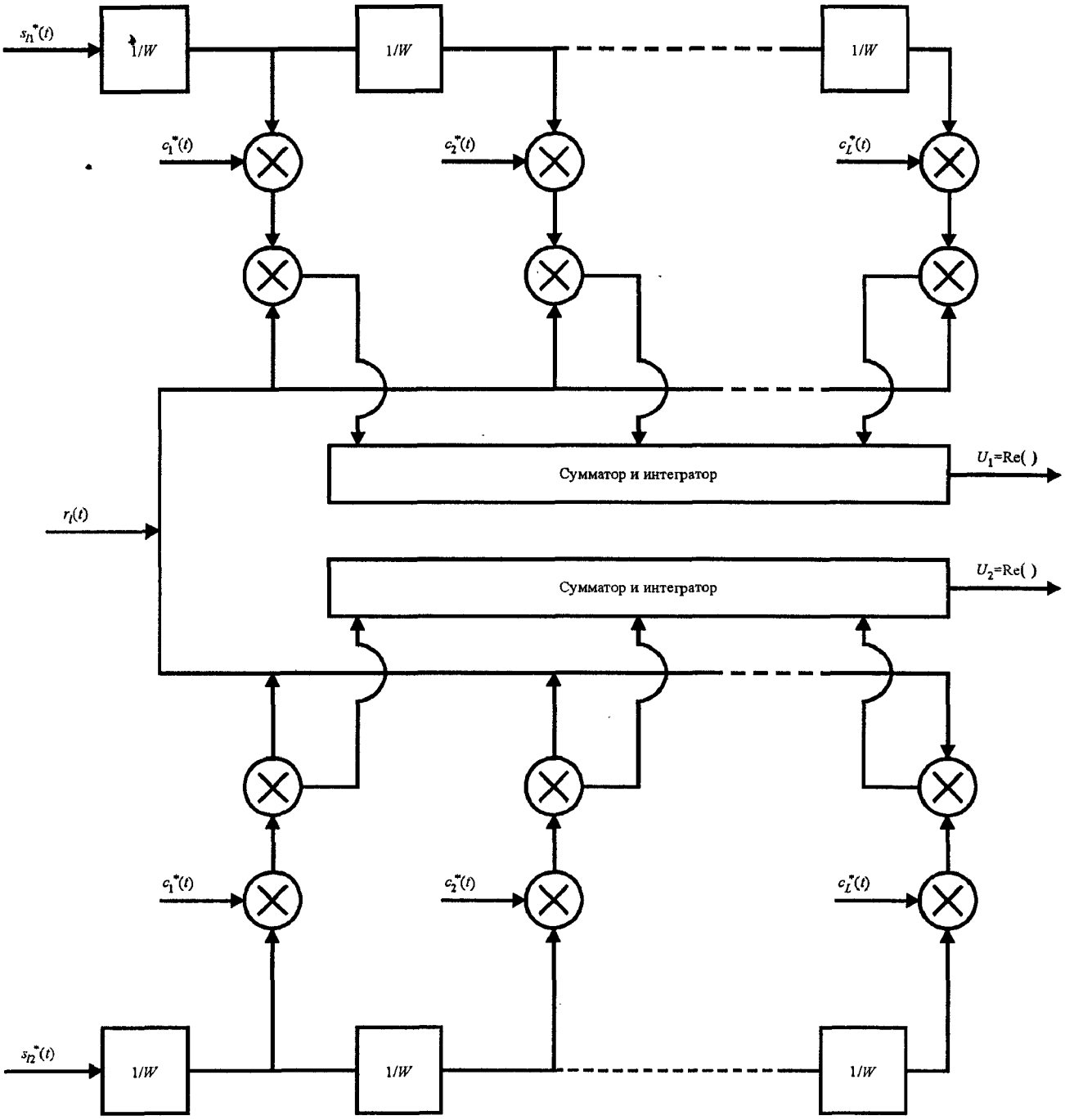


Рис. 14.5.2. Оптимальный демодулятор для широкополосных двоичных сигналов (конфигурация с задержкой эталонного сигнала)

Его действие в определённой степени аналогично обычным садовым граблям (rake) и поэтому имя "RAKE-приёмник" было предложено для этой структуры приёмника Прайсом и Грином (1958).

14.5.3. Качество RAKE приёмника

Теперь рассчитаем качество RAKE приёмника при условии, что замирания сигналов достаточно медленные и позволяют оценить $c_k(t)$ точно (без шума). Далее, на сигнальном

интервале $c_k(t)$ считается константой и обозначается c_k. Таким образом, величины для решения в (14.5.12) можно выразить в виде:

$$U_m = \mathrm{Re}\left[\sum_{k=1}^{L} c_k^* \int_0^T r(t) s_{lm}^*(t - k/W)dt\right], \qquad m = 1, 2. \qquad (14.5.13)$$

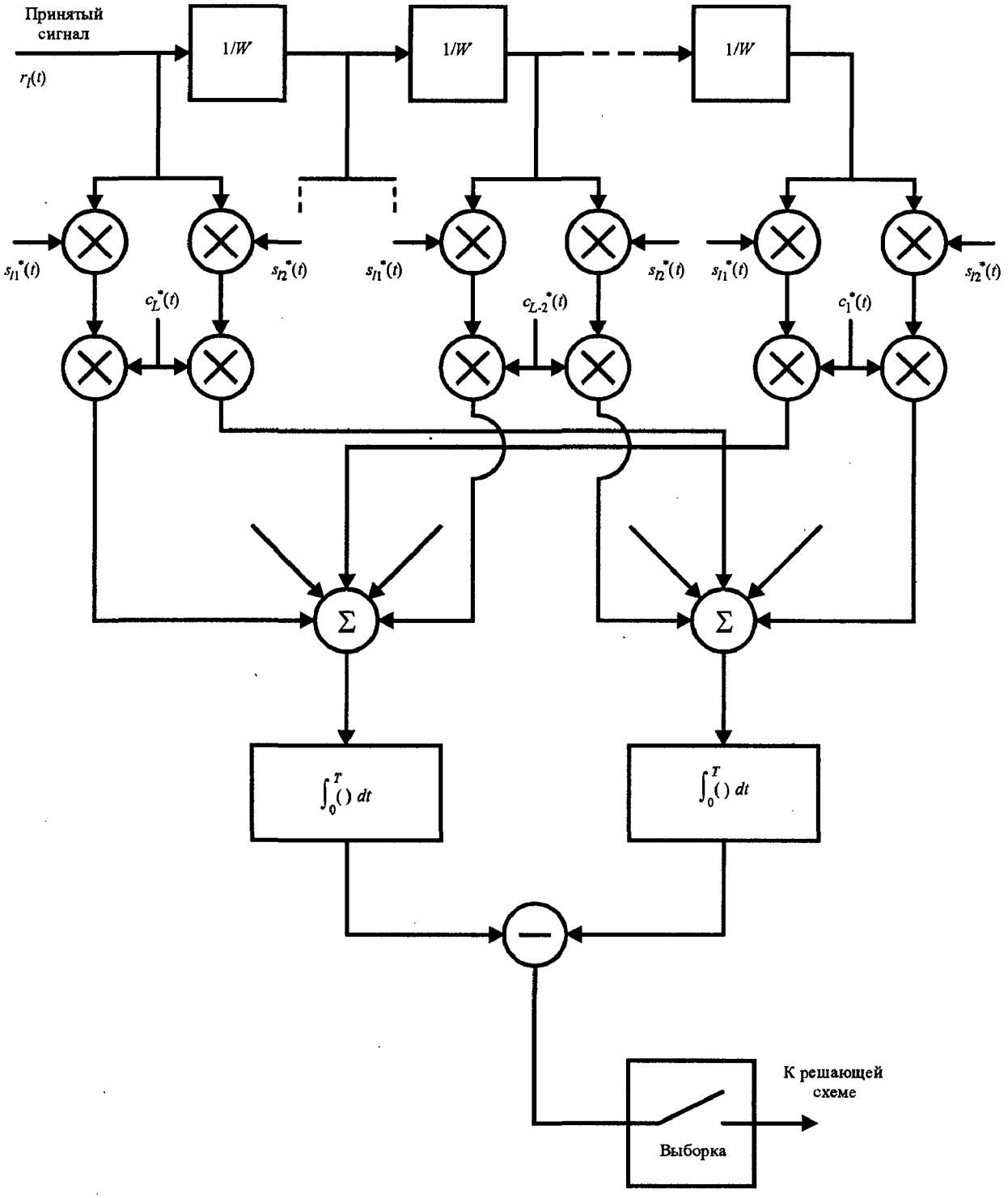


Рис. 14.5.3. Оптимальный демодулятор для широкополосных двоичных сигналов (конфигурация с задержкой принятого сигнала)

Предположим, что передаётся сигнал $s_{l1}(t)$; тогда принимаемый сигнал

$$r_l(t) = \sum_{n=1}^{L} c_n s_{l1}(t - n/W) + z(t), \qquad 0 \le t \le T. \tag{14.5.14}$$

Подстановка (14.5.14) в (14.5.13) даёт

$$\begin{aligned} U_m = \mathrm{Re}\left[\sum_{k=1}^{L} c_k^* \sum_{n=1}^{L} c_n \int_0^T s_{l1}(t - n/W) s_{lm}^*(t - k/W)\,dt\right] + \\ + \mathrm{Re}\left[\sum_{k=1}^{L} c_k^* \int_0^T z(t) s_{lm}^*(t - k/W)\,dt\right], \qquad m = 1, 2. \end{aligned} \tag{14.5.15}$$

Обычно широкополосные сигналы $s_{l1}(t)$ и $s_{l2}(t)$ генерируются посредством псевдослучайных последовательностей, которые образуют сигналы со свойствами

$$\int_0^T s_{li}(t - n/W) s_{li}^*(t - k/W)\,dt \approx 0, \qquad k \ne n, \qquad i = 1, 2 \tag{14.5.16}$$

Если мы предположим, что наши двоичные сигналы синтезируются так, что удовлетворяется это условие тогда (14.5.15) упрощается до[1]

$$\begin{aligned} U_m = \mathrm{Re}\left[\sum_{k=1}^{L} |c_k|^2 \int_0^T s_{l1}(t - k/W) s_{lm}^*(t - k/W)\,dt\right] \\ + \mathrm{Re}\left[\sum_{k=1}^{L} c_k^* \int_0^T z(t) s_{lm}^*(t - k/W)\,dt\right], \qquad m = 1, 2. \end{aligned} \tag{14.5.17}$$

Если двоичные сигналы противоположные, достаточно иметь единственную величину для решения. В этом случае (14.5.17) приводит к

$$U_1 = \mathrm{Re}\left(2\mathscr{E}\sum_{k=1}^{L} \alpha_k^2 + \sum_{k=1}^{L} \alpha_k N_k\right), \tag{14.5.18}$$

где $\alpha_k = |c_k|$, и

$$N_k = e^{j\phi_k} \int_0^T z(t) s_l^*(t - k/W)\,dt. \tag{14.5.19}$$

Но (14.5.18) идентично величине для решения, определяемой (14.4.4), которая соответствует выходу сумматора с максимальным отношением в системе с разнесением L-го порядка. Следовательно, RAKE приёмник с точными оценками весов канальных отводов эквивалентен сумматору максимальных отношений в системе с разнесением L-го порядка. Так, если все веса отводов имеют одинаковое значение средних квадратов, т.е. $E(\alpha_k^2)$ одинаково для всех k, вероятность ошибки приёмника RAKE определяется (14.4.15) и (14.4.16). С другой стороны, если величина средних квадратов не равна при всех к, расчёт вероятности ошибки надо повторить поскольку (14.4.15) теперь не приемлемо.

Мы хотим рассчитать вероятность ошибки для двоичных противоположных и ортогональных сигналов при условии, что величины средних квадратов $\{\alpha_k\}$ отличаются.

---

[1] Хотя свойство ортогональности (14.5.16) можно удовлетворить соответствующим выбором псевдослучайных последовательностей, реально взаимная корреляция $S_{l1}(t\text{-}n/W)$ и $S_{li}^*(\text{-}k/W)$ приводит к сигналу, образующему собственный шум, который безусловно ограничивает качество. Для простоты мы не будем рассматривать соответствующий шум в дальнейших расчётах. Следовательно, результаты качества, представленные ниже, следует рассматривать как нижнюю границу (идеальная Rake). Аппроксимацию качества Rake можно получить путём трактовки собственного шума, как дополнительной компоненты гауссовского шума с мощностью, равной его дисперсии.

Начнём с условной вероятности ошибки

$$P_2(\gamma_b)=Q\left(\sqrt{\gamma_b(1-\rho_r)}\right), \tag{14.5.20}$$

где $\rho_r=-1$ для противоположных сигналов, $\rho_r=0$ для ортогональных сигналов и

$$\gamma_b=\frac{\mathscr{E}}{N_0}\sum_{k=1}^{L}\alpha_k^2=\sum_{k=1}^{L}\gamma_k. \tag{14.5.21}$$

Каждое слагаемое $\{\gamma_k\}$ распределено по закону хи-квадрат с двумя степенями свободы. Это значит, что

$$p(\gamma_k)=\frac{1}{\bar{\gamma}_k}e^{-\gamma_k/\bar{\gamma}_k}, \tag{14.5.22}$$

где $\bar{\gamma}_k$ среднее ОСШ для *k*-го пути, определяемые так

$$\bar{\gamma}_k=\frac{\mathscr{E}}{N_0}E(\alpha_k^2). \tag{14.5.23}$$

Далее, из (14.4.10) мы знаем, что характеристическая функция γ_k равна

$$\psi_{\gamma k}(j\upsilon)=\frac{1}{1-j\upsilon\bar{\gamma}_k}. \tag{14.5.24}$$

Поскольку γ_b – это сумма L статистически независимых компонент $\{\gamma_k\}$, характеристическая функция γ_b равна

$$\psi_{\gamma_b}(j\upsilon)=\prod_{k=1}^{L}\frac{1}{1-j\upsilon\bar{\gamma}_k}. \tag{14.5.25}$$

Обратное преобразование Фурье характеристической функции (14.5.25) определяет ФПВ γ_b в виде

$$p(\gamma_b)=\sum_{k=1}^{L}\frac{\pi_k}{\bar{\gamma}_k}e^{-\gamma_b/\bar{\gamma}_k}, \qquad \gamma_b\ge 0, \tag{14.5.26}$$

где π_k определено так

$$\pi_k=\prod_{\substack{i=1\\ i\ne k}}^{L}\frac{\bar{\gamma}_k}{\bar{\gamma}_k-\bar{\gamma}_i}. \tag{14.5.27}$$

Когда условная вероятность ошибки (14.5.20) усредняется по γ_b с учётом ФПВ (14.5.26), следует результат

$$P_2=\frac{1}{2}\sum_{k=1}^{L}\pi_k\left[1-\sqrt{\frac{\bar{\gamma}_k(1-\rho_r)}{2+\bar{\gamma}_k(1-\rho_r)}}\right]. \tag{14.5.28}$$

Эту вероятность ошибки можно аппроксимировать (при $\bar{\gamma}_k>>1$) так

$$P_2\approx\binom{2L-1}{L}\prod_{k=1}^{L}\frac{1}{2\bar{\gamma}_k(1-\rho_r)}. \tag{14.5.29}$$

Сравнивая (14.5.29) для $\rho_r=-1$ с (14.4.18), видим, что как в случае неравных значений ОСШ на один путь, так и в случае равных значений ОСШ на один путь имеет место одинаковый вид асимптотического поведения вероятности ошибки.

При определении вероятности ошибки приёмника Rake мы предположили, что оценки весов канальных отводов являются точными. На практике относительно хорошие оценки можно получить, если замирания в канале достаточно медленные, например $(\Delta t)_c/T \geq 100$, где T – сигнальный интервал. Рис.14.5.4 иллюстрирует метод оценивания весов ячеек, когда двоичные сигналы ортогональны.

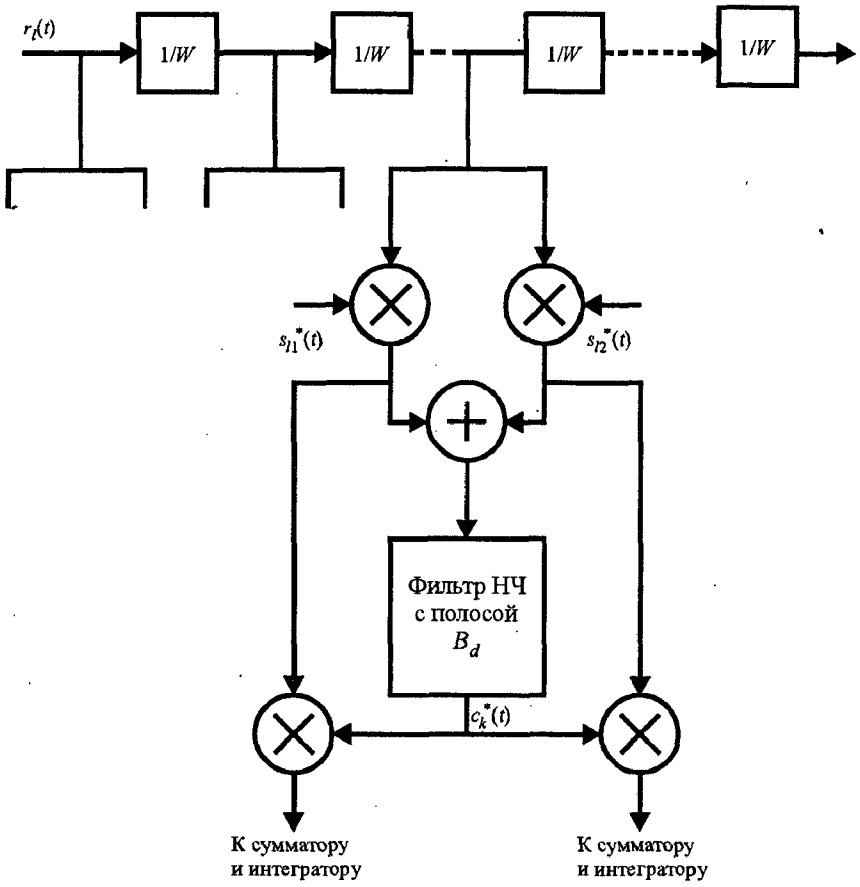


Рис. 14.5.4. Оценивание весов отводов для двоичных ортогональных сигналов

Оценки являются выходом ФНЧ в каждом отводе. В любой момент времени приходящий сигнал или $s_{l1}(t)$ или $s_{l2}(t)$. Следовательно, выход ФНЧ, используемого для оценки $c_k(t)$, содержит сигнал плюс шум одного из корреляторов и только шум от других корреляторов. Этот метод оценивания канала не годится для противоположных сигналов, поскольку сумма двух выходов корреляторов приводит к погашению сигнала. Вместо этого для противоположных сигналов можно использовать один коррелятор. Его выход питает вход ФНЧ после восстановления информационного сигнала (?). Чтобы это выполнить мы должны ввести задержку на сигнальный интервал при выполнении процедуры оценивания канала, как показано на рис.14.5.5. Это значит, что сначала приёмник должен решить является ли информационный символ в принимаемом сигнале +1 или –1 и, затем он использует это решение для восстановления информации на выходе коррелятора до его подачи на ФНЧ.

Если мы не хотим заниматься оценкой весов ячеек для селективного по частоте канала, мы можем использовать сигналы ДФМ или некогерентное детектирование ортогональных сигналов. Структура приёмника Rake для ДФМ иллюстрируется на рис.14.5.6.

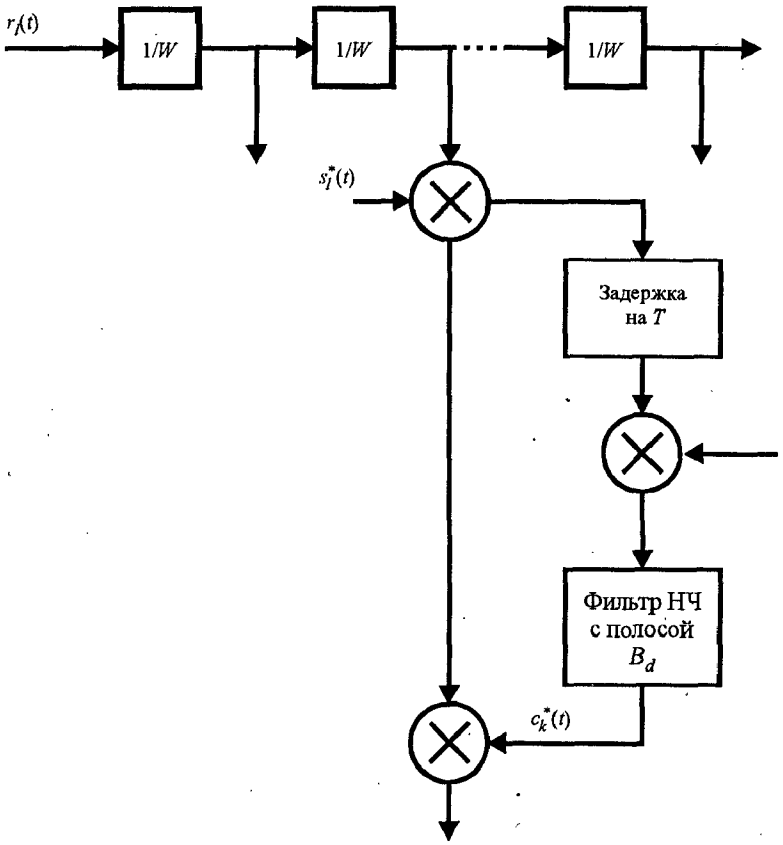


Рис. 14.5.5. Оценивание весов отводов для двоичных противоположных сигналов

Очевидно, что когда передаваемый сигнал $s_l(t)$ удовлетворяет свойству ортогональности (14.5.16) величины для решения идентичны тем, которые определяются (14.4.23) для системы разнесения L-го порядка. Как следствие, вероятность ошибки приёмника Rake для двоичной ДФМ аналогично той, которой определятся (14.4.15) с $\mu = \bar{\gamma}_c/(1+\bar{\gamma}_c)$, когда все пути сигналов имеют одинаковые ОСШ $\bar{\bar{\gamma}}_c$. С другой стороны, если ОСШ $\{\bar{\gamma}_k\}$ различны, вероятность ошибки можно получить путём усреднения (14.4.24), что определяет условную вероятность ошибки для канала с постоянными параметрами, по γ_b с ФПВ (14.5.26). Результат такого интегрирования

$$P_2 = \left(\frac{1}{2}\right)^{2L-1} \sum_{m=0}^{L-1} m! b_m \sum_{k=1}^{L} \frac{\pi_k}{\bar{\gamma}_k} \left(\frac{\bar{\gamma}_k}{1+\bar{\gamma}_k}\right)^{m+1}, \qquad (14.5.30)$$

где π_k определяется (14.5.27), а b_m определяется (14.4.25).

В заключение мы рассмотрим передачу двоичных ортогональных сигналов по селективноу по частоте каналу с квадратичным детектированием в приёмнике. Этот тип сигналов подходит, когда или замирания настолько быстрые, что исключают хорошую оценку весов канальных отводов или когда цена реализации оценивания канала высока. Приёмник Rake с квадратичным сложением сигналов от каждого отвода показан на рис.14.5.7.

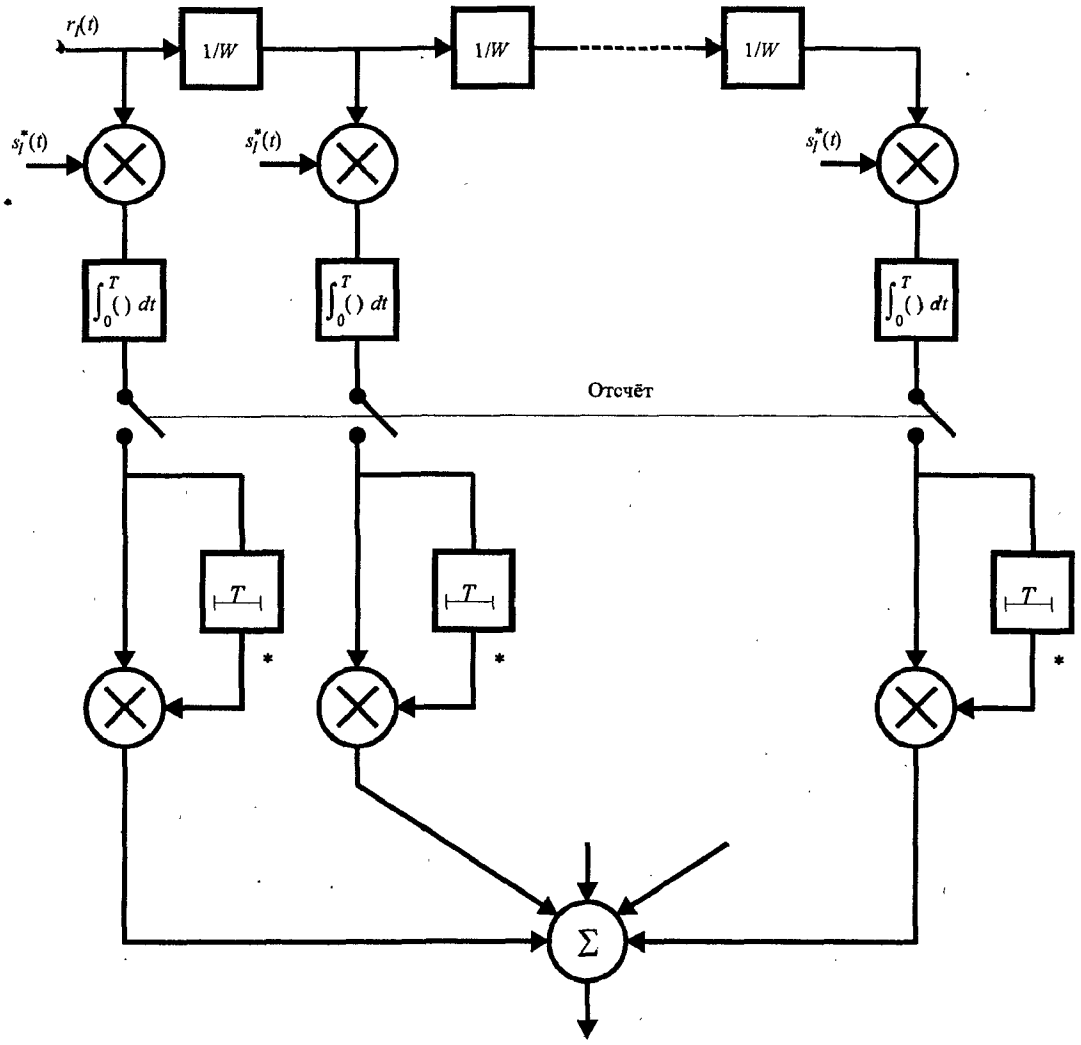


Рис. 14.5.6. Демодулятор Rake для ДФМ сигнала

При вычислении его качества мы снова предположим, что выполняется свойство ортогональности (14.5.16). Тогда величины для решения на выходе приёмника Rake равны

$$U_1 = \sum_{k=1}^{L} |2\mathscr{E}c_k + N_{k1}|^2$$
$$U_2 = \sum_{k=1}^{L} |N_{k2}|^2, \qquad (14.5.31)$$

где мы предположили, что передаваемым сигналом был $s_{l1}(t)$. Снова заметим, что величины для решения идентичны тем, которые определяются (14.4.29) для ортогональных сигналов с разнесением L-го порядка. Следовательно, качество приёмника Rake при квадратичном детектировании ортогональных сигналов определяется (14.4.15) с $\mu = \bar{\gamma}_c/(2+\bar{\gamma}_c)$, когда все сигнальные пути имеют одинаковые ОСШ. Если ОСШ различны мы можем усреднить условную вероятность ошибки (14.4.24), при замене γ_b на $\frac{1}{2}\gamma_b$ по γ_b с ФПВ $p(\gamma_b)$, определяемой (14.5.26). Результат такого усреднения определяется (14.5.30) при замене $\bar{\bar{\gamma}}_k$ на $\frac{1}{2}\bar{\gamma}_k$.

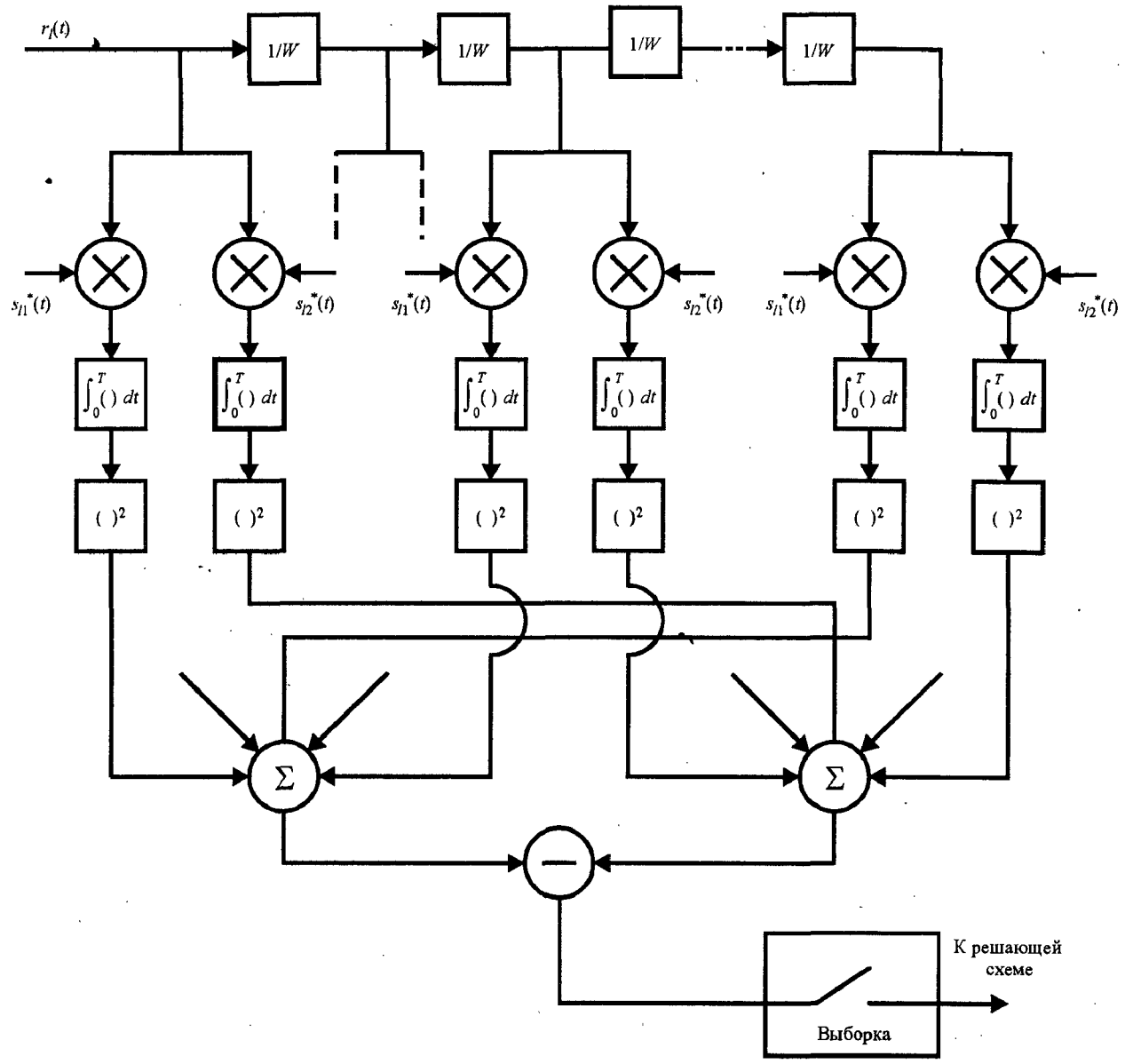


Рис. 14.5.7. Демодулятор Rake с квадратичным сложением ортогональных сигналов

В вышеприведённом анализе считалось, что в демодуляторе приёмника Rake, показанном на рис.14.5.7, с квадратичным сложением ортогональных сигналов, сигнальные компоненты имеются при каждой задержке. Если это не так, то качество ухудшается, поскольку некоторые из корреляторов ячеек будут вносить только шумы. В этих условиях составляющие с наиболее низким уровнем, содержащие только шум, должны быть исключены, как показано Чаем и др. (1988).

Этим мы заканчиваем наше обсуждение передачи сигналов по селективному по частоте каналу. Схемы приёмника Rake, представленные в этом разделе, можно легко обобщать на многопозиционные сигналы. Действительно, если выбрана M -ичная ФМ или ДФМ, то структура приёмника Rake, представленная в этом разделе, остаётся без изменений. Только детекторы ФМ и ДФМ, следующие за Rake коррелятором, различны.

14.6. КОДИРОВАННЫЕ СИГНАЛЫ ДЛЯ КАНАЛОВ С ЗАМИРАНИЯМИ

До сих пор мы показали, что техника разнесения очень эффективна при преодолении вредных эффектов замираний, вызванных меняющимися во времени характеристиками

рассеяния канала. Техника разнесения во времени и (или) частоте можно рассматривать как форму блокового кодирования с повторением информационной последовательности. С этой точки зрения, техника суммирования, описанная выше, представляет декодирование мягких решений для кода с повторением. Поскольку код с повторением – это тривиальная форма кодирования, мы теперь рассмотрим дополнительные преимущества, получаемые от более эффективных типов кодов. В частности, мы покажем, что кодирование обеспечивает эффективное средство разнесения по каналу с замираниями. Величина (порядок) разнесения, обеспечиваемая кодом, прямо связано с его минимальным расстоянием.

Как показано в разделе 14.4, разнесение во времени получается передачей сигнальных компонент с той же информацией по многим временным интервалам, взаимно разделённых на величину равной или большей времени когерентности $(\Delta t)_c$ канала. Аналогично, частотное разнесение получается передачей сигнальных компонент, несущих одинаковую информацию, по многим частотным интервалам, взаимно разнесённых на величину, по крайней мере, равной полосе частотной когерентности канала $(\Delta f)_c$. Таким образом, сигнальные компоненты, несущие одинаковую информацию, подвергаются статистически независимым замираниям.

Чтобы расширить эти понятия на кодированную информационную последовательность, мы просто потребуем, чтобы сигнал, соответствующий кодовому биту или кодовому символу, замирал независимо от замираний сигналов, соответствующих другому кодовому биту или кодовому символу. Это требование может привести к неэффективному использованию имеющегося в распоряжение частотно-временного пространства и существование больших неиспользуемых участков в этом двухмерном сигнальном пространстве. Чтобы этого не произошло, применяют перемежение: определённое число кодовых слов можно разнести во времени, частоте или одновременно по времени и частоте, таким образом, что сигналы, соответствующие битам или символам каждого кодового слова замирают независимо. Таким образом, мы предполагаем, что частотно-временное сигнальное пространство разделяется на неперекрывающиеся частотно-временные ячейки. Сигнал, соответствующий кодовому биту или кодовому символу, передаётся внутри такой ячейки.

Дополнительно к предположению о статистической независимости замираний сигнальных компонент данного кодового слова мы также предполагаем, что компоненты аддитивного шума, поражающие принимаемые сигналы, являются белыми гауссовскими процессами, которые статистически независимы и одинаково распределены в отдельных ячейках частотно-временного пространства. Также предполагаем, что между соседними ячейками имеется достаточный разнос, так что интерференцией между ячейками можно пренебречь.

Важным является исследование техники модуляции, которая используется для передачи кодированных информационных последовательностей. Если замирания в канале достаточно медленные для того, чтобы позволить надёжно оценить фазу, тогда можно использовать ФМ или ДФМ. Если это невозможно, тогда подходящим является ЧМ с некогерентным детектированием. В нашей трактовке мы предполагаем, что невозможно установить точные значения фазы для сигналов в различных ячейках, занимаемых передаваемым сигналом. Следовательно, мы выбираем ЧМ с некогерентным декодированием.

Модель цифровой системы связи, для которой будет рассчитана вероятность ошибки, показана на рис.14.6.1. Кодер может быть двоичным, недвоичным или каскадным объединением из недвоичного кодера и двоичного кодера. Далее, код, создаваемый

кодером, может быть блоковым, свёрточным или, в случае каскадирования, смесь блокового и свёрточного кодов.

Чтобы объяснить модуляцию, демодуляцию и декодирование для сигналов типа ЧМ (ортогональных) рассмотрим линейный блоковый код, в котором k информационных символов кодируются в блок из n символов. Для упрощения и без потери общности мы предположим, что все n символов кодового слова передаются одновременно по каналу по многим частотным ячейкам. Кодовое слово $\mathbf{C}_i$, имеющее символы $\{c_{ij}\}$, отображается в ЧМ сигнал следующим образом. Если $c_{ij}=0$, передаётся частота f_{0j}, если $c_{ij}=1$, то передаётся частота f_{1j}. Это означает, что для передачи n символов кодового слова требуется $2n$ частот или ячеек, но на интервале кодового слова передаются только n частот. Поскольку каждое кодовое слово содержит k информационных символов, показатель расширения полосы частот для ЧМ равен $B_c=2n/k$.

Демодулятор для принимаемого сигнала разделяет сигнал на $2n$ спектральных компонент, соответствующих используемым на передаче частотам. Таким образом, демодулятор можно реализовать как банк из $2n$ фильтров, причём каждый фильтр согласован с одной из переданных частот. Выходы $2n$ фильтров детектируются некогерентно. Поскольку релеевские замирания и аддитивные белые гауссовские шумы в $2n$ частотных ячейках взаимно независимые и одинаково распределённые случайные процессы, оптимальное максимально правдоподобное декодирование мягких решений требует, чтобы отклики этих фильтров были бы продетектированы квадратично и соответствующим образом просуммированы для каждого кодового слова, чтобы формировать $M=2^k$ величин для решения. Выбирается кодовое слово, соответствующее максимальной величине решения. Если используется декодирование жёстких решений, оптимальный максимально правдоподобный декодер выбирает кодовое слово, имеющее минимальное расстояние Хемминга относительно принятого кодового слова.

Хотя в выше представленном обсуждении предполагалось использование блокового кода свёрточный кодер можно легко применить в блок-схеме, показанной на рис.14.6.1. Для примера, если используется двоичный свёрточный код, каждый символ в его выходной последовательности можно передать двоичной ЧМ. Максимально-правдоподобное правило декодирования мягких решений для свёрточного кода можно эффективно реализовать посредствам алгоритма Витерби (АВ), в котором метрики для выживших последовательностей в любой точке решётки состоят из суммы квадратичных выходов для соответствующих путей по решётке. С другой стороны, если используется декодирование жёстких решений, АВ применяется с использованием в качестве метрик расстояния Хемминга.

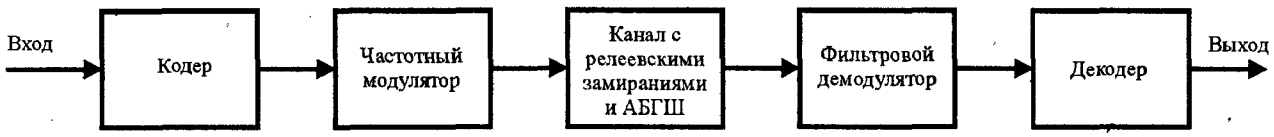


Рис. 14.6.1. Модель системы связи с ЧМ

14.6.1. Вероятность ошибки при декодировании мягких решений и использовании двоичного блокового кода

Рассмотрим декодирование линейного двоичного (n,k) кода при передаче по каналу с релеевскими замираниями, как описано выше. Оптимальный декодер мягких решений,

основанный на правиле максимально правдоподобия, формирует $M = 2^k$ величин для решения

$$U_i = \sum_{j=1}^{n} \left[(1 - c_n) |y_{0j}|^2 + c_{ij} |y_{1j}|^2 \right] = \sum_{j=1}^{n} \left[|y_{0j}|^2 + c_{ij} \left(|y_{1j}|^2 - |y_{0j}|^2 \right) \right], \qquad i = 1, 2, \ldots, 2^k, \qquad (14.6.1)$$

где $|y_{rj}|^2$, $j = 1, 2, \ldots, n$ и $r = 0, 1$ представляют квадраты огибающих выходов $2n$ фильтров, которые настроены на $2n$ возможных переданных частот. Решение делается в пользу кодового слова, которому соответствует максимальное значение величины решения из набора $\{U_i\}$.

Наша цель в этом разделе сводится к определению вероятности ошибки декодера мягких решений. Для этого предположим, что передаётся кодовое слово $\mathbf{C}_1$ из одних нулей. Среднее принимаемое отношение сигнала/шум на частоту (на ячейку) обозначим $\bar{\gamma}_c$. Суммарные принимаемые ОСШ для n частот равно $n\bar{\gamma}_c$ и, следовательно, среднее ОСШ на бит

$$\bar{\gamma}_b = \frac{n}{k} \bar{\gamma}_c = \frac{\bar{\gamma}_c}{R_c}, \qquad (14.6.2)$$

где R_c – скорость кода.

Величина для решения U_1, соответствующая кодовому слову $\mathbf{C}_1$, определяется (14.6.1) при $c_{ij} = 0$ для всех j. Вероятность того, что решение принято в пользу m-го кодового слова равно

$$\begin{aligned} P_2(m) &= P(U_m > U_1) = P(U_1 - U_m < 0) \\ &= P\left[\sum_{j=1}^{n} (c_{1j} - c_{mj}) \left(|y_{1j}|^2 - |y_{0j}|^2 \right) < 0 \right] \\ &= P\left[\sum_{j=1}^{w_m} \left(|y_{0j}|^2 - |y_{1j}|^2 \right) < 0 \right], \end{aligned} \qquad (14.6.3)$$

где w_m – вес m-го кодового слова. Но вероятность (14.6.3) как раз вероятность ошибки при квадратичном сложении двоичных ортогональных сигналов ЧМ с порядком разнесения w_m. Это значит

$$P_2(m) = p^{w_m} \sum_{k=0}^{w_m - 1} \binom{w_m - 1 + k}{k} (1 - p)^k ; \qquad (14.6.4)$$

$$P_2(m) \le p^{w_m} \sum_{k=0}^{w_m - 1} \binom{w_m - 1 + k}{k} = \binom{2w_m - 1}{w_m} p^{w_m}, \qquad (14.6.5)$$

где

$$p = \frac{1}{2 + \bar{\gamma}_c} = \frac{1}{2 + R_c \bar{\gamma}_b}. \qquad (14.6.6)$$

Альтернативно мы можем использовать верхнюю границу Чернова, полученную в разделе 14.4, которые можно представить здесь так

$$P_2(m) \le [4p(1 - p)]^{w_m}. \qquad (14.6.7)$$

Сумма двоичных событий по всем $M-1$ кодовым словам с ненулевым весом даёт верхнюю границу вероятности ошибки. Итак,

$$P_M \le \sum_{m=2}^{M} P_2(m). \qquad (14.6.8)$$

Поскольку минимальное расстояние линейного кода равно минимальному весу то следует

$$\left(1+R_c\bar{\gamma}_b\right)^{-w_m} \le \left(2+R_c\bar{\gamma}_b\right)^{-d_{\min}}.$$

Использование этих отношений вместе с (14.6.5) и (14.6.8) даёт простую, ... свободную, верхнюю границу, которую можно выразить в виде

$$P_M < \frac{\sum_{m=2}^{M} \begin{pmatrix} 2w_m - 1 \\ w_m \end{pmatrix}}{\left(2+R_c\bar{\gamma}_b\right)^{d_{\min}}}. \qquad (14.6.9)$$

Эта простая граница указывает на то, что кодирование обеспечивает эффективный порядок разнесения, равный $d_{\min}$. Ещё более простой является верхняя объединённая граница

$$P_M < (M-1)\left[4p(1-p)\right]^{d_{\min}}, \qquad (14.6.10)$$

которая получается из границы Чернова, данной (14.6.7).

В качестве примера, иллюстрирующего выгоду кодирования в релеевском канале с замираниями, мы привели на рис.14.6.2 кривые вероятности ошибки, полученные посредством расширенного кода Голея (24,12) и двоичной ЧМ и четверичной ЧМ с двойным разнесением.

Поскольку расширенный код Голея требует в целом 48 ячеек и $k=12$, показатель расширения полосы $B_e = 4$. Это также показатель расширения полосы для двоичной и четверичной ЧМ с $L=2$. Следовательно, три типа сигналов сравниваются при одинаковом показателе расширения полосы частот. Заметим, что при $P_b = 10^{-4}$ код Голея превосходит четверичную ЧМ больше чем на 6 дБ, а при $P_b = 10^{-5}$ разница, примерно, 10 дБ.

Объяснение высокого качества кода Голея – его большое минимальное расстояние $(d_{\min} = 8)$, которое переводится в эквивалентное разнесение восьмого порядка $(L=8)$. В противоположность этому двоичные и четверичные сигналы ЧМ имеют только разнесение второго порядка. Итак, код Голея более эффективно использует предоставляемую полосу частот канала. Цену, которую мы должны заплатить за высокое качества кода – увеличение сложности декодирования.

14.6.2. Вероятность ошибки при декодировании жёстких решений и использовании линейных двоичных блоковых кодов

Границы качества, получаемые при декодировании жёстких решений для линейного двоичного (n,k) кода, уже даны в разделе 8.1.5. Эти границы применимы к произвольному каналу без памяти с двоичным входом и двоичным выходом (двоичный симметричный канал) и, следовательно, они приемлемы без изменения для канала с релеевскими замираниями и АБТШ при статистически независимых замираниях сигналов отдельных символов кодового слова. Вероятность ошибки на бит, необходимая для расчёта этих границ, когда используется двоичная ЧМ с некогерентным декодированием как техника модуляции и демодуляции, дана (14.6.6).

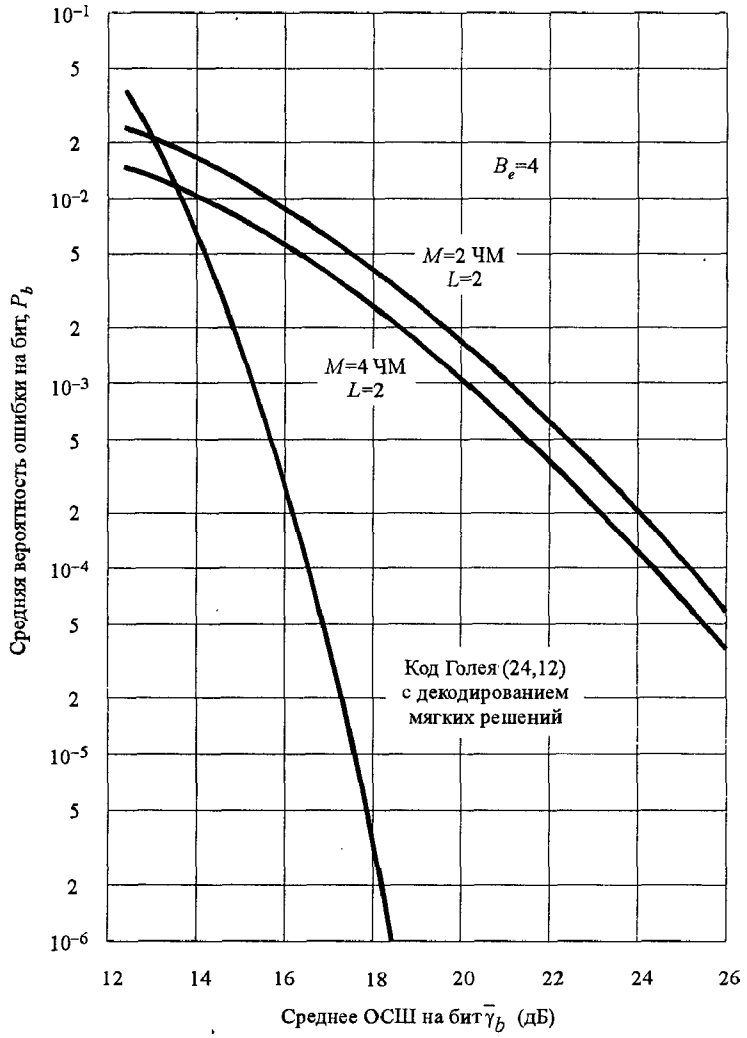


Рис. 14.6.2. Пример сравнительных характеристик помехоустойчивости при разнесении и кодировании с B_e=4

Частный интересный результат получается, если мы используем верхнюю границу Чернова для вероятности ошибки при декодировании жёстких решений, определяемой (8.1.89). Таким образом есть

$$P_2(m) \le [4p(1-p)]^{w_m/2}, \qquad (14.6.11)$$

а P_M ограничена сверху (14.6.8). Для сравнения верхняя граница Чернова для $P_2(m)$ при использовании декодирования мягких решений определяется (14.6.7). Мы видим, что влияние декодирования жёстких решений сводится к сокращению расстояния между кодовыми словами в 2 раза. Когда минимальные расстояния кода относительно мало сокращение расстояния в 2 раза намного больше заметно в канале с замираниями, чем в канале без замираний.

С целью иллюстрации мы на рис.14.6.3 показали качество кода Голея (23, 12) при использовании декодирования жёстких и мягких решений.

Разница в качестве при $P_b = 10^{-5}$ примерно равно 6 дБ. Эта существенная разница в качестве по сравнению с 2 дБ между декодированием мягких и жёстких решений в канале без замираний и АБГШ. Также заметим, что разница в качестве увеличивается по мере уменьшения P_b. Короче, эти результаты указывают на выгоду декодирования мягких решений относительно декодирования жёстких решений в канале с замираниями.

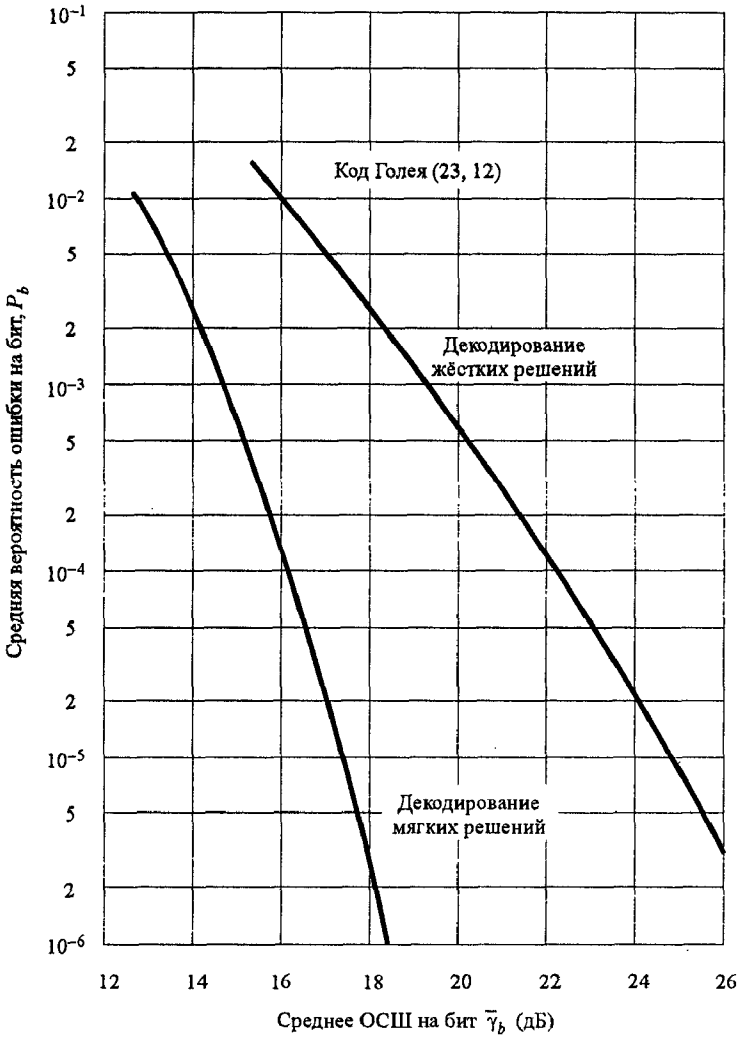


Рис. 14.6.3. Сравнение декодирования мягких и жёстких решений

14.6.3. Верхние границы качества свёрточных кодов в канале с релеевскими замираниями

В этом подразделе мы определим качество двоичных свёрточных кодов, когда используется канал с релеевскими замираниями и АБГШ. На кодер поступает за определённое время k двоичных символа и выходит за это время n двоичных символа. Таким образом, скорость кода $R_c = k/n$. Двоичные символы с выхода кодера передаются по каналу с релеевскими замираниями посредством сигналов двоичной ЧМ, которые на приёме детектируются квадратичным детектором. Детектор обеспечивает максимально правдоподобное оценивание последовательности сигналов, детектируя мягкие или жёсткие решения, что эффективно реализуется посредством АВ.

Сначала рассмотрим декодирование мягких решений. В этом случае метрики, вычисленные алгоритмом Витерби, равны сумме квадратов детекторных выходов демодулятора. Предположим, что передана последовательность из одних нулей. Следуя процедуре, изложенной в разделе (8.2.3), легко найти, что вероятность ошибки при парном сравнении метрики, соответствующей последовательности из одних нулей, с метрикой, соответствующей другой последовательности, которая сливается первый раз с состоянием, соответствующим одним нулям, равна

$$P_2(d) = p^d \sum_{k=0}^{d-1} \binom{d-1+k}{k} (1-p)^k, \qquad (14.6.12)$$

где d – число позиций символов, в которых отличаются две последовательности, а p определяется (14.6.6). Таким образом, $P_2(d)$ как раз вероятность ошибки для двоичной ЧМ с квадратичным детектированием при разнесении порядка d. Альтернативно мы можем использовать границу Чернова (14.6.7) для $P_2(d)$. В любом случае, вероятность ошибки на бит ограничена сверху, как показано в разделе 8.2.3 выражением

$$P_b < \frac{1}{k} \sum_{d=d_{св}}^{\infty} \beta_d P_2(d), \qquad (14.6.13)$$

где весовые коэффициенты $\{\beta_d\}$ в сумме получаются из выражения первой производной передаточной функции $T(D, N)$, определяемой (8.2.25).

Если на приеме осуществляется декодирование жёстких решений применимы границы для вероятности ошибки для двоичных свёрточных кодов, полученные в разделе 8.2.4. Это значит, что P_b снова ограничена сверху выражением (14.6.13), где $P_2(d)$ определяется (8.2.28) для нечетных d и (8.2.29) для четных d, или ограничены сверху (граница Чернова) (8.2.31), а p определяется (14.6.6).

Интересно отметить, что как и при блоковом кодировании, когда используются границы Чернова для $P_2(d)$, при свёрточном декодировании эффект декодирования жёстких решений сводится к сокращению расстояний между кодовыми словами (порядка разнесения) в два раза по сравнению с декодированием мягких решений.

Следующие численные результаты иллюстрируют вероятность ошибки двоичных свёрточных кодов со скоростью $1/n$ и максимальным свободным расстоянием при $n = 2$, 3 и 4 при декодировании мягких решений алгоритмом Витерби. Прежде всего, рис.14.6.4 показывает качество свёрточных кодов со скоростью 1/2 и кодовых ограничений 3, 4 и 5. Показатель расширения полосы частот для двоичной ЧМ равен $B_e = 2n$. Поскольку увеличение кодового ограничения ведет к увеличению сложности декодера при соответствующем увеличении минимального свободного расстояния, разработчик системы может взвесить два этих фактора при выборе кода.

Другой путь для увеличения расстояния без увеличения кодового ограничения кода сводится к повторению каждого выходного символа m раз. Это эквивалентно сокращению скорости кода в m раз или расширению полосы частот на ту же величину. Результатом является свёрточный код с минимальным свободным расстоянием $md_{св}$, где $d_{св}$ – свободное расстояние исходного кода без повторений. Такой код всегда настолько хорош с точки зрения минимального расстояния, как код со скоростью $1/mn$, и с тем же максимальным свободным расстоянием. Вероятность ошибки кода с повторением ограничена сверху так

$$P_b < \frac{1}{k} \sum_{d_{св}}^{\infty} \beta_d P_2(md), \qquad (14.6.14)$$

где $P_2(md)$ определяется (14.6.5). Рис.14.6.5 иллюстрирует качество кодов с повторениями $(m = 1, 2, 3, 4)$ со скоростью $1/2$ и кодовым ограничением 5.

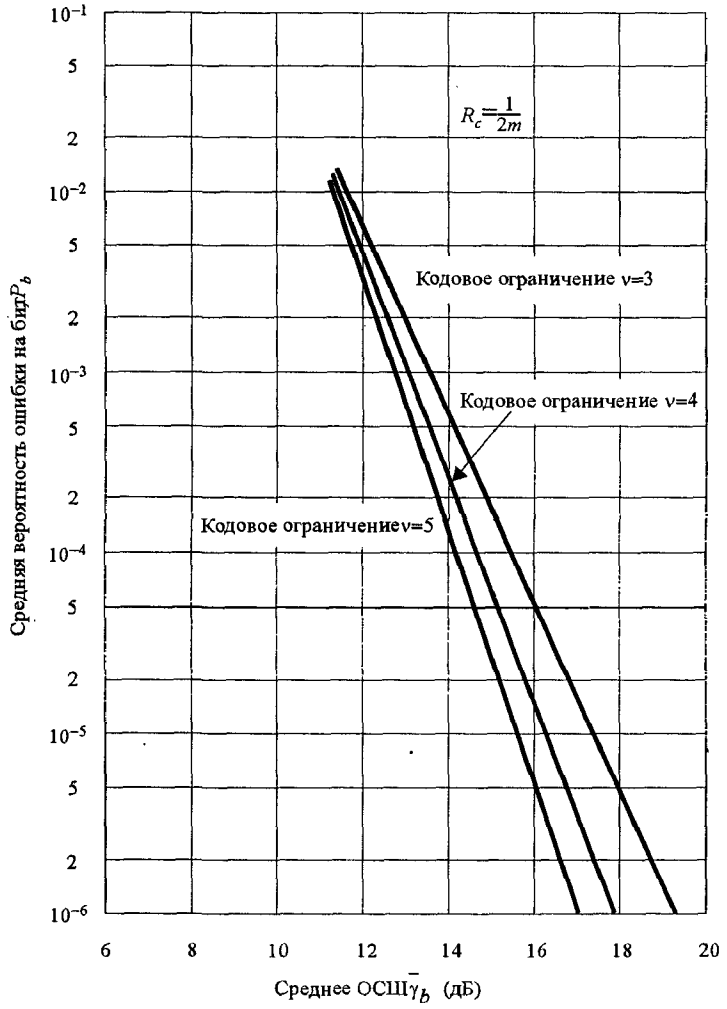


Рис. 14.6.4. Характеристики двоичного свёрточного кода со скоростью 1/2 с декодированием мягких решений

14.6.4. Использование кодов с постоянным весом и каскадных кодов для канала с замираниями

До сих пор наша трактовка кодирования для канала с релеевскими замираниями базировалась на использовании двоичной ЧМ, как техника модуляции для передачи каждого двоичного символа кодового слова. При такой технике модуляции все 2^k кодовых слов в коде (n,k) имеют одинаковые передаваемые энергии. Далее, при условии, что замирания n передаваемых частот статистически взаимно независимы и одинаково распределены, средняя энергия принимаемого сигнала для всех $M=2^k$ возможных кодовых слов также одинаковая. Следовательно, при декодировании мягких решений, решения принимаются в пользу кодового слова, имеющего наибольшую из сравниваемых величину для решения.

Условие, что принимаемые кодовые слова имеют одинаковые средние ОСШ, имеет важное значение при реализации приёмника. Если принимаемые кодовые слова не имеют одинаковое среднее ОСШ, приёмник должен обеспечить определенную выравнивающую компенсацию для каждого принимаемого кодового слова так, чтобы сделать его с равной энергией. В общем, определение соответствующих выравнивающих слагаемых трудно для реализации, поскольку это требует оценивание средней мощности принимаемого сигнала. Следовательно, условие равной энергии принимаемых кодовых слов значительно упрощает обработку сигналов в приёмнике.

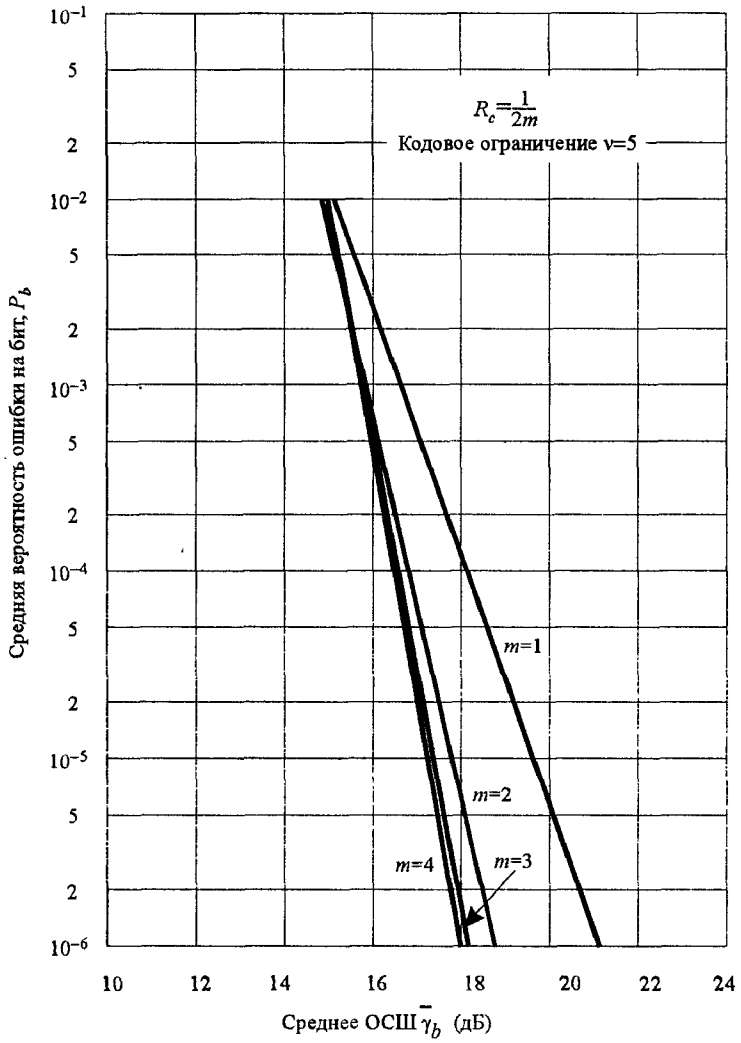


Рис. 14.6.5. Характеристики двоичного свёрточного кода со скоростью 1/2m с кодовым ограничением ν=5 и декодированием мягких решений

Имеется альтернативный метод для генерирования сигналов кодовых слов с равной энергией, когда код имеет постоянный вес, т.е. когда каждое кодовое слово имеет одинаковое число единиц. Заметим, что такой код нелинеен. Тем не менее предположим, что мы выделяем одну частоту или ячейку для каждого кодового символа 2^k кодовых слов. Таким образом, (n,k) двоичный блоковый код имеет n выделяемых частот. Сигналы конструируются путем передачи частот на тех позициях кодового слова, где имеется символ 1; на других позициях сигнального интервала частота не передаётся. Такая техника модуляции для передачи кодовых слов называется амплитудной манипуляцией (OOK – on-off keying). Поскольку код имеет постоянный вес w, каждое кодовое слово состоит из w передаваемых частот, которые определяются позициями единиц в кодовом слове.

Как и при ЧМ считается, что все частоты в OOK системе, которые передаются по каналу, замирают независимо по полосе частот и во времени от одного кодового слова к другому. Огибающая принимаемого сигнала для каждой частоты описывается статистически распределением Релея. Считается, что в каждой частотной ячейке присутствует статистически независимый аддитивный белый гауссовский шум.

Приёмник использует максимально правдоподобное декодирование (мягких решений) для отображения принимаемых сигналов в один из M возможных переданных кодовых

слов. Для этой цели используется n согласованных фильтров, каждый с одной из n частот. В предположении статистической независимости замирающих сигналов и аддитивного белого гауссовского шума в n частотных ячейках огибающие выходов согласованных фильтров возводятся в квадрат и суммируются для образования M величин для решения

$$U_i = \sum_{j=1}^{n} c_{ij} |y_j|^2, \qquad i = 1, 2, \ldots, 2^k, \tag{14.6.15}$$

где $|y_j|^2$ соответствует квадрату огибающей фильтра для j-й частоты, где

$$j = 1, 2, \ldots, n.$$

Можно ожидать, что условие постоянства веса кода серьезно ограничит выбор кода. Однако это не так. Чтобы это проиллюстрировать, мы сжато опишем некоторые методы для конструирования кодов с постоянным весом. Это обсуждение не является исчерпывающим.

Метод 1: Нелинейные преобразования линейного кода. В общем, если при образовании каждого кодового нового слова произвольного двоичного кода мы используем одну двоичную последовательность при каждом появлении 0 и другую последовательность для каждой 1, то можно получить двоичный блоковый код с постоянным весом, если две используемые последовательности имеют равный вес и равную длину. Если длина используемой последовательности v и её исходный код (n, k), тогда результирующий код с постоянным весом будет кодом (vn, k). Вес кодового слова будет в n раз больше веса использованных последовательностей, а минимальное расстояние будет равно минимальному расстоянию исходного кода, умноженному на расстояние между двумя используемыми последовательностями. Таким образом, использование дополнительных последовательностей, когда v четно, даёт код с минимальным расстоянием $vd_{\min}$ и весом $\frac{1}{2}vn$.

Простейшая форма этого метода это случай, когда $v = 2$ и, когда каждый 0 заменяется на пару 01 и каждая 1 заменяется последовательности 10 (или наоборот). В качестве примера возьмём в качестве исходного кода расширенный код Голея (24, 12). Параметры исходного и результирующего кодов с постоянным весом даны в табл.14.6.1.

Таблица 14.6.1. Пример кода с постоянным весом, формируемого методом 1

Параметры кода	Исходный код Голея	Код с постоянным весом
n	24	48
k	12	12
M	4096	4096
$d_{\min}$	8	16
w	переменный	34

Метод 2: Вычёркивание. В этом методе мы начинаем с произвольного двоичного блокового кода и выбираем из него подмножество, состоящее из всех слов определённого веса. Несколько различных кодов с постоянным весом можно получить из одного исходного кода, меняя выбор веса w. Поскольку кодовые слова результирующего кода с вычёркиванием можно рассматривать как подмножество всех перестановок любого кодового слова множества, Гаардер (1971) использовал для этого кода термин *двоичная перестановочная модуляция с вычёркиванием* (binary expurgated permutation modulation – BEXPERM). Действительно, двоичные блоковые коды с постоянным весом, сконструированные другими методами, также можно рассматривать как BEXPERM коды. Этот метод генерирования кодов с постоянным весом по существу противоположен

первому методу, в котором длина слова n держится постоянной, а объем (размер) кода M меняется. Ясно, что минимальное расстояние для подмножества с постоянным весом не меньше, чем в исходном коде.

В качестве примера мы рассмотрим код Голея (24, 12) и образуем два различных кода с постоянным весом, показанные в табл.14.6.2.

Таблица 14.6.2. Примеры кодов с постоянным весом, формируемых вычёркиванием

Параметры кодов	Исходный код Голея	Код с пост. весом № 1	Код с пост. весом № 2
n	24	24	24
k	12	9	12
M	4096	759	2576
$d_{\min}$	8	≥8	≥8
w	переменный	8	12

Метод 3: Матрица Адамара. Может казаться, что этот метод формирует двоичный блоковый код с постоянным весом непосредственно, но фактически он является частным случаем метода вычеркивания. В этом методе формируется матрица Адамара, как описано в разделе 8.1.2 и создаётся код с постоянным весом путем выбора столбцов (кодовых слов) из этой матрицы. Напомним, что матрица Адамара это $n \times n$ матрица (n четное целое) из 1 и 0 со свойством, что один столбец отличается от любого другого столбца точно в $\frac{1}{2}n$ позициях. Один столбец матрицы нормально выбирается из одних нулей.

В любом из других столбцов половина элементов 0 и другая половина 1. Код Адамара с кодовыми словами размера $2(n-1)$ получается путем выбора этих $n-1$ столбцов и их дополнений. Выбирая $M = 2^k \le 2(n-1)$ из этих кодовых слов, мы получаем код Адамара, который мы обозначим $H(n,k)$, в котором каждое кодовое слово содержит k информационных символов. Результирующий код имеет постоянный вес $\frac{1}{2}n$ и минимальное расстояние $d_{\min} = \frac{1}{2}n$.

Поскольку n частотных ячеек используются для передачи k информационных символов, показатель расширения полосы для кода Адамара $H(n,k)$ определяется так:

$$B_e = \frac{n}{k} \text{ ячеек на информационный бит,}$$

что равно величине, обратной скорости кода. Среднее отношение сигнал-шум (ОСШ) на бит, обозначаемое $\bar{\gamma}_b$, связано со средним ОСШ на ячейку $\bar{\gamma}_c$ соотношением

$$\bar{\gamma}_c = \frac{k}{\frac{1}{2}n}\bar{\gamma}_b = 2\frac{k}{n}\bar{\gamma}_b = 2R_c\bar{\gamma}_b = \frac{2\bar{\gamma}_b}{B_c}. \qquad (14.6.16)$$

Сравним качество кода Адамара с постоянным весом при фиксированной ограниченной полосе частот с традиционным M-ичным ортогональным ансамблем сигналов, причем каждый сигнал имеет порядок разнесения L. M ортогональных сигналов с разнесением эквивалентны блоковому ортогональному коду, имеющему длину блока $n = LM$, а $k = \log_2 M$. Для примера, если $M = 4$ и $L = 2$, кодовые слова ортогонального кода равны

$$C_1 = [1\ 1\ 0\ 0\ 0\ 0\ 0\ 0]$$
$$C_2 = [0\ 0\ 1\ 1\ 0\ 0\ 0\ 0]$$
$$C_3 = [0\ 0\ 0\ 0\ 1\ 1\ 0\ 0]$$
$$C_4 = [0\ 0\ 0\ 0\ 0\ 0\ 1\ 1]$$

Для передачи этих кодовых слов используется ООК модуляция, требующая $n=8$ ячеек и поскольку каждое кодовое слово содержит $k=2$ бита информации, показатель расширения полосы $B_e=4$. В общем, мы обозначим ортогональный блоковый код $O(n,k)$. Показатель расширения полосы частот

$$B_e=\frac{n}{k}=\frac{LM}{k}. \tag{14.6.17}$$

Таким образом, ОСШ на бит связано с ОСШ на ячейку отношением

$$\bar{\gamma}_c=\frac{k}{L}\bar{\gamma}_b=M\frac{k}{n}\bar{\gamma}_b=M\frac{\bar{\gamma}_b}{B_e}. \tag{14.6.18}$$

Теперь обратим наше внимание на характеристику качества этих кодов. Во-первых, точное значение вероятности ошибки кодового слова (символа) для M-ичного ортогонального кода по каналу с релеевскими замираниями с разнесением было дано в замкнутой форме в разделе 14.4. Как показано раньше, это выражение скорее громоздкое для расчетов, особенно когда L или M или оба параметра большие. Вместо этого мы используем объединённую верхнюю границу, которая очень удобна. Это значит, для ансамбля из M ортогональных сигналов, вероятность ошибки на символ ограничена сверху

$$P_M\le(M-1)P_2(L)=\left(2^k-1\right)P_2(L)<2^kP_2(L), \tag{14.6.19}$$

где $P_2(L)$ – вероятность ошибки для двух ортогональных сигналов, каждый с разнесением порядка L, определяемая (14.6.12) с $p=1/(2+\bar{\gamma}_c)$. Вероятность ошибки на бит получается умножением P_M на $2^{k-1}/\left(2^k-1\right)$, как объяснено раньше.

Простая верхняя (объединенная) граница вероятности ошибки для кодового слова кода Адамара $H(n,k)$ можно получить, если учесть, что вероятность ошибки различения между переданным кодовым словом и любым другим кодовым словом ограничена сверху величиной $P_2\left(\frac{1}{2}d_{\min}\right)$, где $d_{\min}$ – минимальное расстояние кода. Следовательно, верхняя граница для P_M равна

$$P_M\le(M-1)P_2\left(\tfrac{1}{2}d_{\min}\right)<2^kP_2\left(\tfrac{1}{2}d_{\min}\right). \tag{14.6.20}$$

Таким образом, «эффективный порядок разнесения» кода ООК модуляции равен $\frac{1}{2}d_{\min}$. Вероятность ошибки на бит можно выразить как $\frac{1}{2}P_M$ или она неплотно ограничена сверху путем умножения P_M на множитель $2^{k-1}/\left(2^k-1\right)$, который является множителем, использованным выше для ортогональных кодов. Последний был выбран для расчетов вероятности ошибки, данных ниже.

Рис.14.6.6 и 14.6.7 иллюстрируют соответственно вероятность ошибки выбранного числа кодов Адамара и ортогональных блоковых кодов, соответственно для нескольких показателей расширения полосы частот. Выгода, полученная от увеличения объема M алфавита (или k, так как $k=\log_2 M$) и увеличения показателя расширения полосы частот очевидно из рассмотрения этих кривых. Заметим для примера, что двукратное повторение кода $H(20,5)$ приводит к коду, обозначенному ${}_2H(20,5)$ и имеющему показатель расширения полосы частот $B_e\doteq 8$. Рис.14.6.8 показывает качество двух типов кодов, сравниваемых при равенстве показателя расширения полосы частот.

Можно увидеть, что кривые вероятности ошибки для кодов Адамара идут круче, чем соответствующие кривые для блоковых ортогональных кодов. Это характерное поведение

объясняется просто тем фактом, что при том же показателе расширения полосы частот коды Адамара обеспечивают большее разнесение, чем ортогональные блоковые коды. Альтернативно можно сказать, что коды Адамара обеспечивают лучшую эффективность использования полосы частот, чем ортогональные блоковые коды. Необходимо напомнить, однако, что при малых ОСШ код с низким разнесением превосходит коды с большим разнесением, как следствие того факта, что в канале с релеевскими замираниями имеется оптимальное распределение суммарного принимаемого ОСШ по сигналам разнесения. Следовательно, кривые ортогональных блоковых кодов пересекают сверху кривые кодов Адамара в области малых ОСШ (высокая вероятность ошибки).

Метод 4: Каскадирование. В этом методе мы начнем с двух кодов: один двоичный и другой недвоичный. Двоичный код – внутренний блоковый код (n,k) с постоянным весом (нелинейный). Недвоичный код, который может быть линейным, является внешним. Чтобы отличать его от внутреннего кода, мы используем большие буквы, например (N,K) код, где N и K измеряются числом символов q-ичного алфавита. Объем q алфавита, через который определяется внешний код, не может быть больше, чем число слов внутреннего кода. Внешний код, если он определяется через кодовые слова двоичного внутреннего кода q-ичными символами, это новый код.

Важный частный случай получается, когда $q=2^k$ и объем внутреннего кода выбирается равным 2^k. Тогда число слов $M=2^{kK}$ и каскадная структура образует (nN,kK) код. Показатель расширения полосы частот этого каскадного кода определяется произведением показателей расширения полосы частот внутреннего и внешнего кодов.

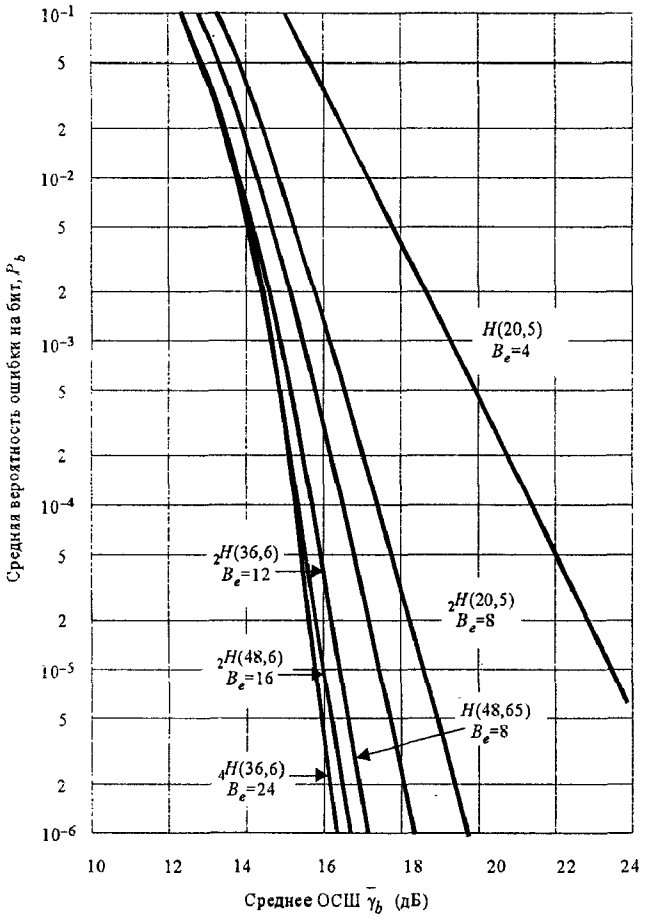


Рис. 14.6.6. Характеристики кода Адамара

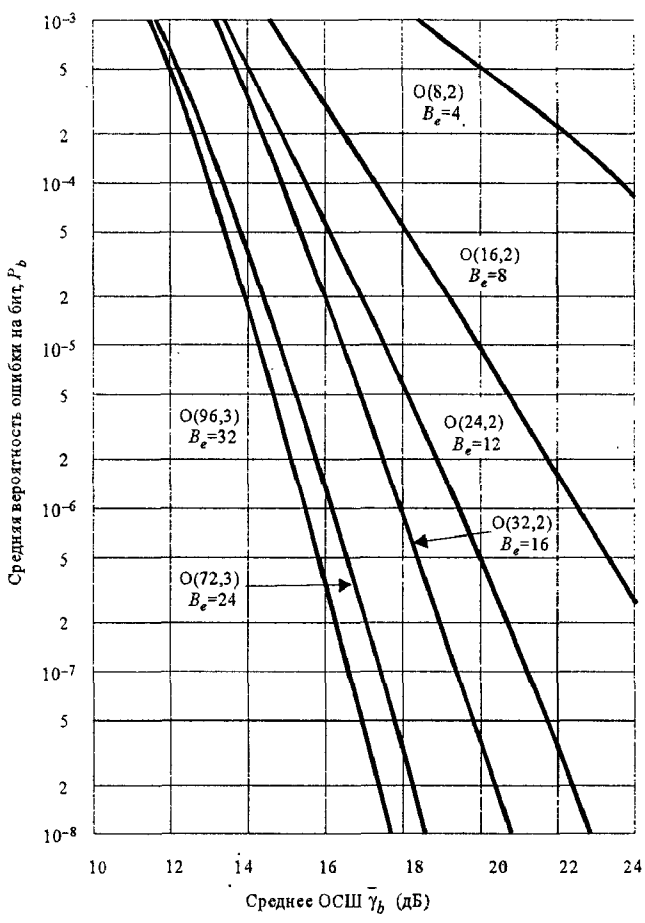


Рис. 14.6.7. Характеристики блоковых ортогональных кодов

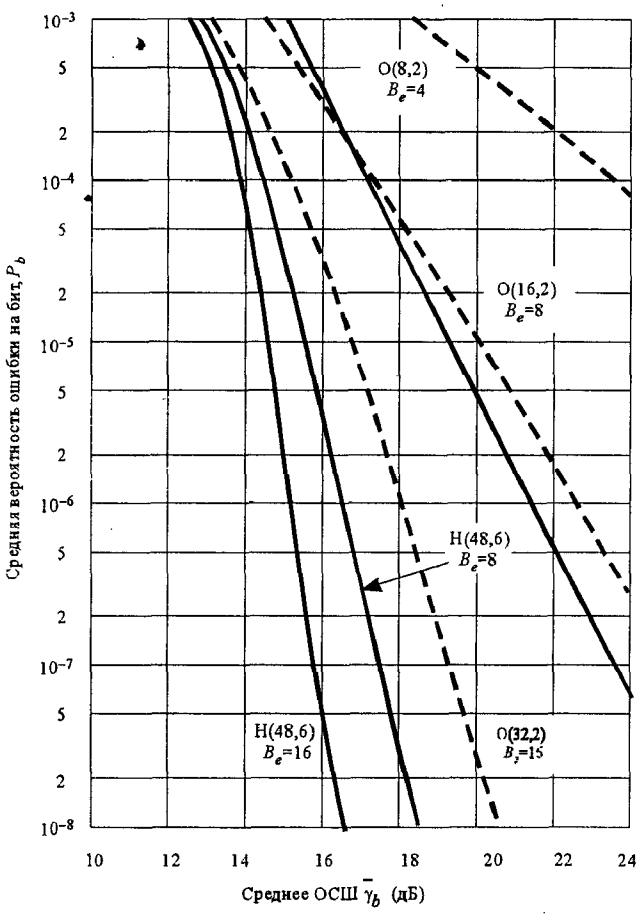


Рис. 14.6.8. Сравнение характеристик кодов Адамара и блоковых ортогональных кодов

Теперь мы покажем выгоду в качестве, получаемую в канале с релеевскими замираниями посредством каскадного кода. В частности мы сконструируем каскадный код, в котором внешний код является k-дуальным (недвоичным) свёрточным кодом, а внутренний код является или кодом Адамара, или блоковым ортогональным кодом. То есть мы рассматриваем k-дуальный код с M-ичными ($M = 2^k$) ортогональными сигналами для модуляции, как каскадный код. Во всех случаях при расчетах мы предполагаем декодирование по Витерби мягких решений.

Вероятность ошибки для k-дуального свёрточного кода получена при дифференцировании передаточной функции кода, определенной (8.2.39). Для k-дуального кода без повторений со скоростью 1/2 вероятность ошибки на бит (для случая, когда кодовый выходной k-битовый символ k-дуального кода отображается в один из $M = 2^k$ ортогональных кодовых слов) ограничена сверху

$$P_b < \frac{2^{k-1}}{2^k - 1} \sum_{m=4}^{\infty} \beta_m P_2(m), \qquad (14.6.21)$$

где $P_2(m)$ дано (14.6.12).

Для примера, 2-дуальный код со скоростью 1/2 может использовать 4-позиционный ортогональный код $O(4,2)$ в качестве внутреннего. Показатель расширения полосы частот для результирующего каскадного кода равен, конечно, произведению показателей расширения полосы частот внутреннего и внешнего кодов.

Так, в этом примере скорость внешнего и внутреннего кодов равна 1/2. Следовательно, $B_e = (4/2)(2) = 4$.

Заметим, что если каждый символ k-дуального кода повторить r раз, то это эквивалентно использованию ортогонального кода с разнесением $L = r$. Если мы выберем $r = 2$ в выше приведённом примере, результирующий ортогональный код обозначается $O(8,2)$ и показатель расширения полосы частот для 2-дуального кода со скоростью 1/2 равно $B_e = 8$. Следовательно, слагаемые $P_2(m)$ в (14.6.21) следует заменить на $P_2(mL)$, если ортогональный код имеет L-кратное разнесение. Поскольку код Адамара имеет «эффективное разнесение» $\frac{1}{2}d_{\min}$, то следует, что при использовании кода Адамара в качестве внутреннего кода с k-дуальным кодом в качестве внешнего, еще справедлива верхняя граница для вероятности ошибки на бит результирующего каскадного кода, определенного (14.6.21), если $P_2(m)$ заменить на $P_2\left(\frac{1}{2}md_{\min}\right)$. С этими изменениями была

рассчитана верхняя граница вероятности ошибки на бит, определяемая (14.6.21), для k-дуального свёрточного кода со скоростью 1/2 с кодом Адамара или блокового ортогонального кода,как внутреннего. Результирующий каскадный код имеет показатель расширения полосы в два раза больший, чем показатель расширения полосы внутреннего кода.

Сначала рассмотрим выигрышные качества, обусловленные каскадным объединением кодов. Рис.14.6.9 иллюстрирует качество k-дуального кода с ортогональным блоковым кодом в качестве внутреннего по сравнению с качеством ортогонального блокового кода для показателя расширения полосы частот $B_e = 4$, 8, 16 и 32. Выигрыш качества, обусловленный каскадированием, очень впечатляет. Например, при вероятности ошибки 10^{-6} и $B_e = 8$ k-дуальный код превосходит ортогональный блоковый код на 7,5 дБ. Этот выигрыш можно приписать увеличению разнесения (увеличению минимального расстояния), получаемому посредством каскадирования. Аналогично рис.14.6.10 иллюстрирует качество двух k-дуальных кодов с кодом Адамара в качестве внутреннего кода для $B_e = 8$ и 12.

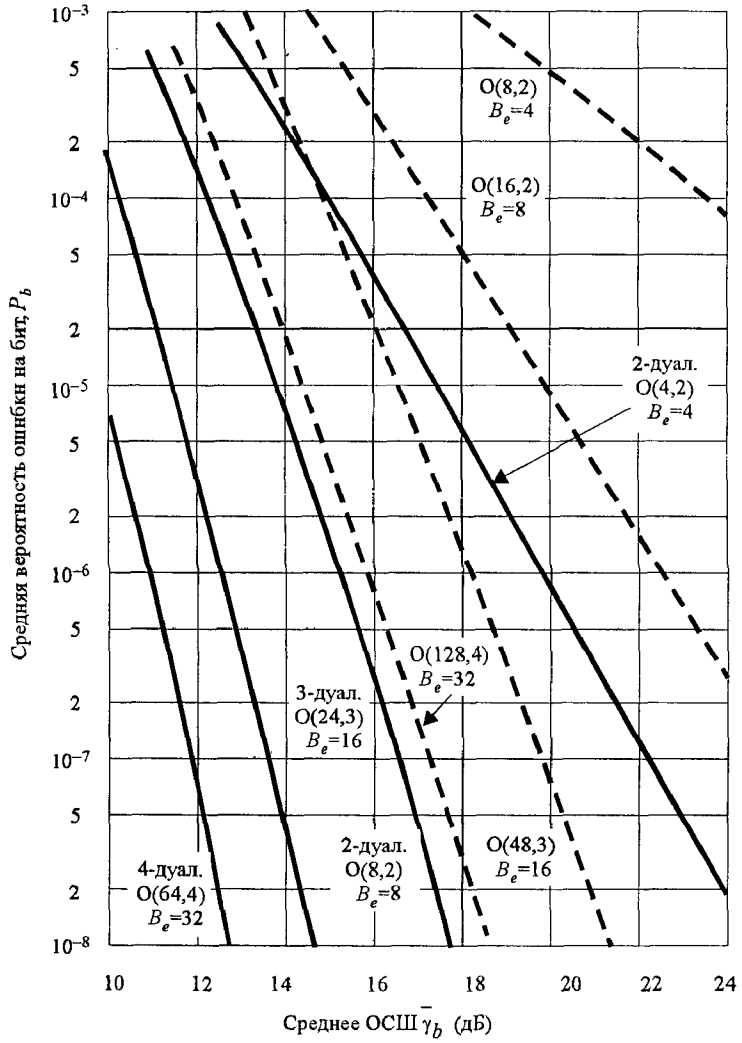


Рис. 14.6.9. Сравнение характеристик блоковых ортогональных кодов и каскадных кодов, содержащих k-дуальный код

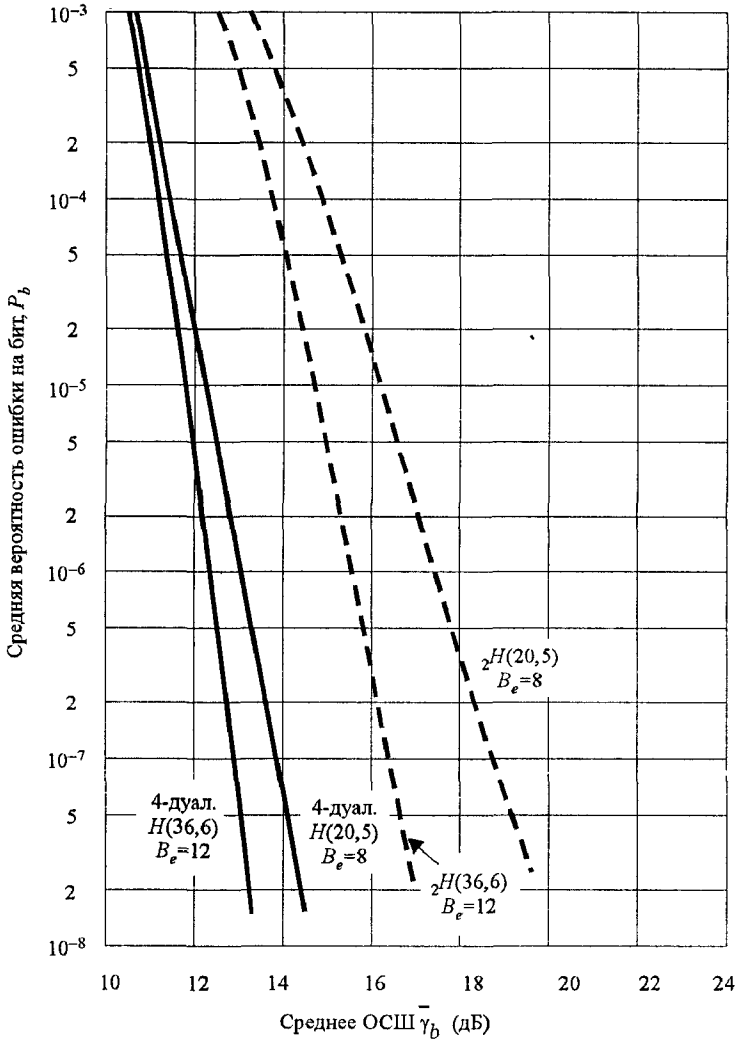


Рис. 14.6.10. Сравнение характеристик кодов Адамара и каскадных кодов, содержащих k-дуальный код

Видно, что выигрыш качества, обусловленный каскадированием, остаётся значительным, но, конечно, не таким впечатляющим, как тот, который иллюстрируется на рис.14.6.9. Объяснение в том, что сам код Адамара обеспечивает большое разнесение, так что рост разнесения, достигаемый каскадированием, не приводит к большому выигрышу качества для области ошибок, иллюстрируемых на рис.14.6.10.

Далее мы сравним качество для двух типов внутреннего кода, используемых с k-дуальным внешним кодом. Рис.14.6.11 показывает сравнение для $B_e = 8$.

Заметим, что внутренний код $_2H(4,2)$ имеет $d_{\min} = 4$ и, следовательно, он имеет эффективный порядок разнесения, равный 2. Но это двухкратное разнесение достигается передачей четырех частот на кодовое слово. Другими словами, ортогональный код $O(8,2)$ также даёт двухкратное разнесение, но это достигается передачей только двух частот на кодовое слово. Следовательно, код O(8, 2) на 3 дБ лучше, чем $_2H(4,2)$ код. Эта разница в качестве имеет место, когда два кода используются как внутренние коды в соединении с 2-дуальным кодом. С другой стороны, для Be=8 можно использовать код H (20, 5) как внутренний для 5-дуального кода и его качество значительно лучше, чем для 2-дуального кода при малых вероятностях ошибки. Это улучшение качества достигается ценой увеличения сложности декодирования. Аналогично на рис.14.6.12 мы сравниваем качество дуального кода с двумя типами внутренних кодов для Be=16.

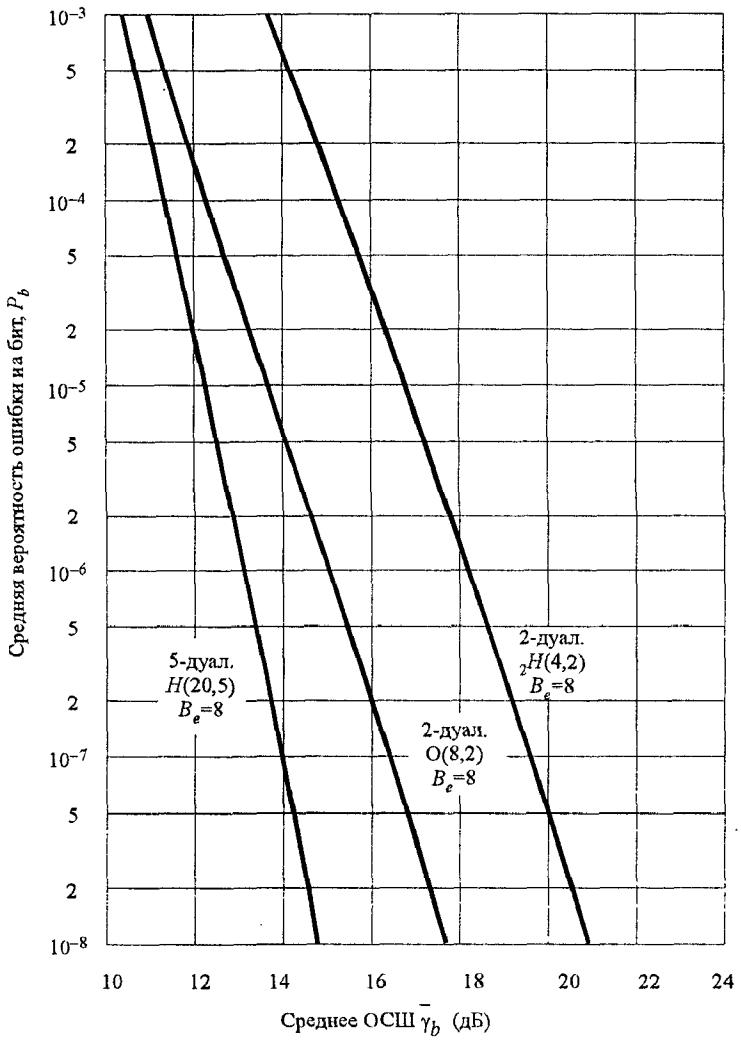


Рис. 14.6.11. Характеристика k-дуального кода с внутренним кодом Адамара или блоковым ортогональным кодом; B_e=8

Заметим, что внутренний код $_3H(8,3)$ имеет $d_{\min}=12$ и, следовательно, он даёт эффективное разнесение 6. Это разнесение достигается передачей 12 частот на кодовое слово. Ортогональный внутренний код $O(24,3)$ даёт только разнесение третьего порядка, которое достигается передачей трех частот на кодовое слово. Следовательно, внутренний код $O(24,3)$ более эффективен при низких ОСШ, т.е. для области вероятности ошибки, показанных на рис.14.6.12. При больших ОСШ, 3-дуальный код с кодом Адамара $_3H(8,3)$ в качестве внутреннего превосходит его другое объединение с кодом $O(24,3)$ как внутреннего, что обусловлено большим разнесением, даваемом кодом Адамара. Для того же показателя расширения полосы частот $B_e=16$ можно использовать 6-дуальный код с кодом $H(48,6)$ в качестве внутреннего для достижения улучшения по сравнению с 3-дуальным кодом с внутренним кодом $_3H(8,3)$. Снова, это улучшение качества (которое в этом случае не такое впечатляющее, как те, которые показаны на рис.14.6.11) должно быть взвешено против увеличения сложности декодирования, свойственные 6-дуальному коду.

Численные результаты, данные выше, иллюстрирует выигрыш качества при использовании кодов с хорошими дистанционными свойствами и декодировании мягких решений в канале с релеевскими замираниями как альтернатива обычным (традиционным) M-ичным ортогональным сигналам с разнесением. Дополнительно результаты

иллюстрируют выгоду каскадирования кодов в таком канале при использовании k-дуального свёрточного кода как внешнего кода и кода Адамара или блокового ортогонального кода как внутреннего кода. Хотя k-дуальные коды были использованы как внешние коды, аналогичные результаты получены, когда код Рида-Соломона используется как внешний код. Здесь имеется даже больший выбор для внутреннего кода.

Важным параметром при выборе как внешнего, так и внутреннего кода является минимальное расстояние результирующего каскадного кода, требуемое для достижения заданного уровня качества. Поскольку много кодов удовлетворяет требуемому качеству, окончательный выбор делается с учётом сложности декодирования и требований к полосе частот.

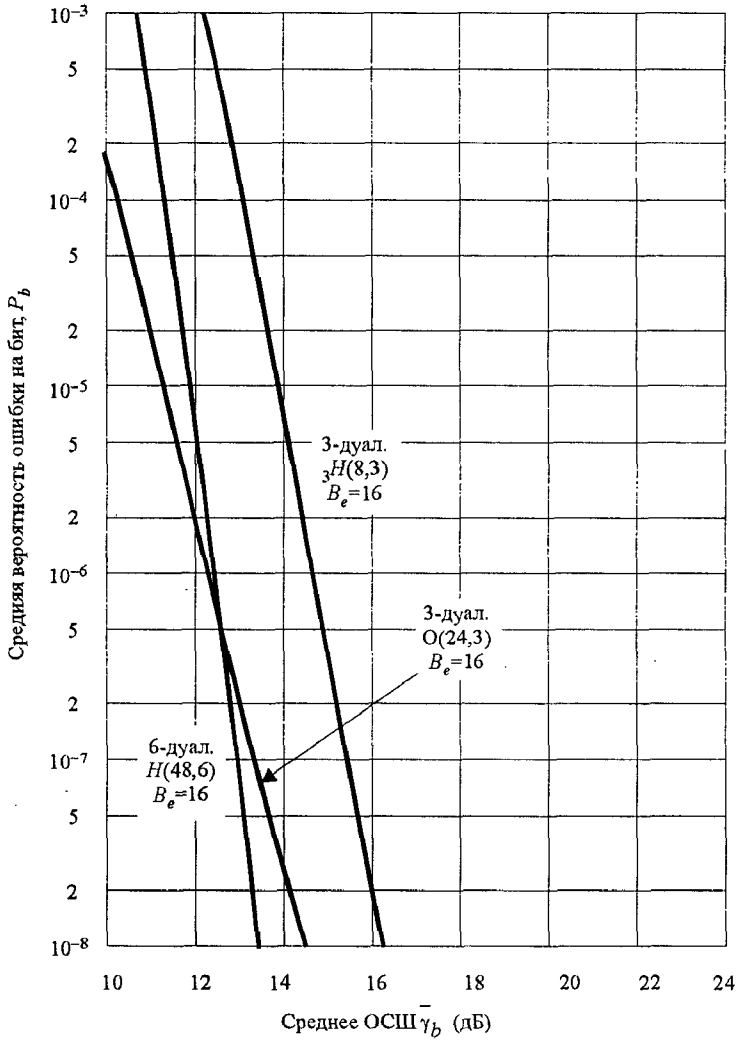


Рис. 14.6.12. Характеристика k-дуального кода с внутренним кодом Адамара или блоковым ортогональным кодом; B_e=16

14.6.5. Синтез систем, основанный на предельной скорости

В приведенном обсуждении кодированных сигналов мы продемонстрировали эффективность различных кодов для каналов с замираниями. В частности мы видели выгоду декодирования мягких решений и каскадного кодирования как средства для увеличения минимальных расстояний и следовательно, величину разнесения кодовых сигналов. В этом подразделе мы рассмотрим случайный выбор кодовых слов и определим верхнюю объединённую границу вероятности ошибки, которая зависит от параметра предельной скорости для канала с релеевскими замираниями.

Рассмотрим модель системы связи, показанную на рис.14.6.1.

Модулятор имеет q- ичный ортогональный ЧМ алфавит. Кодовые слова с длиной блока n отображаются в сигналы путем выбора n частот из алфавита q частот. Демодуляция выполняется пропусканием сигнала через банк из q согласованных фильтров, за которыми следуют квадратичные детекторы. Считается, что выполняется декодирование мягких решений. Выходы квадратичных детекторов демодулятора соответствующим образом комбинируются (суммируются) с равными весами для формирования M величин для решения, соответствующего M возможным переданным кодовым словам.

Чтобы рассчитать объединенную верхнюю границу для вероятности ошибки в канале с релеевскими замираниями и АБГШ, мы сначала рассчитаем вероятность ошибки на бит, включая расчёт величины для решения U_1, соответствующей переданному кодовому слову, и любых из остальных $M-1$ величин для решений, соответствующих остальным кодовым словам. Пусть U_2 – другая величина для решения и пусть U_1 и U_2 имеют l общих частот. Поскольку вклад в U_1 и U_2 этих l частот идентичен, он исчезает при формировании разности U_1-U_2 для принятия решения. Так как две величины для решения различаются в $n-l$ частотах, вероятность ошибки равно той, которая получается для двоичной ортогональной системы ЧМ с порядком разнесения $n-l$. Точное выражение для этой вероятности ошибки даётся (14.6.4), где $p=1/(2+\bar{\gamma}_c)$, а $\bar{\gamma}_c$ – среднее ОСШ на частоту. Для упрощения мы используем границу Чернова для этого двоичного перехода, ведущего к ошибке, и определяемого (14.6.2), т.е.

$$P_2(U_1,U_2|l)\le[4p(1-p)]^{n-l}. \qquad (14.6.22)$$

Теперь проведем усреднение по ансамблю двоичных систем связи. Имеется q^n возможных кодовых слов, из которых мы случайно выбираем два кодовых слова. Каждое кодовое слово выбирается с равной вероятностью. Вероятность того, что два случайных выбранных кодовых слов имеют вместе l частот, равна:

$$P(l)=\binom{n}{l}\left(\frac{1}{q}\right)^l\left(1-\frac{1}{q}\right)^{n-l}. \qquad (14.6.23)$$

Если усредним (14.6.22) по l с вероятностями (14.6.23), мы получим

$$\begin{aligned}\overline{P_2(U_1,U_2)}&=\sum_{l=0}^{n}P_2(U_1,U_2|l)P(l)\le\\ &\le\sum_{l=0}^{n}\binom{n}{l}\left(\frac{1}{q}\right)^l\left[4\left(1-\frac{1}{q}\right)p(1-p)\right]^{n-l}\le\\ &\le\left\{\frac{1}{q}\left[1+4(q-1)p(1-p)\right]\right\}^n\end{aligned} \qquad (14.6.24)$$

В заключении найдем объединенную границу вероятности ошибки системы связи, которая использует $M=2^k$ случайно выбираемых кодовых слов

$$\overline{P_M}\le(M-1)\overline{P_2(U_1,U_2)}<M\,\overline{P_2(U_1,U_2)}. \qquad (14.6.25)$$

Комбинируя (14.6.24) и (14.6.25), мы получаем верхнюю границу для средней вероятности ошибки на символ

$$\overline{P_M}<2^{-n(R_0-R_c)}, \qquad (14.6.26)$$

где $R_c=k/n$ – скорость кода, а R_0 – предельная скорость, определяемая так:

$$R_0 = \log_2 \frac{q}{1+4(q-1)p(1-p)}, \tag{14.6.27}$$

где

$$p = \frac{1}{2+\bar{\gamma}_c} \tag{14.6.28}$$

График R_0 как функции $\bar{\gamma}_c$, показан на рис.14.6.13 для $q=2$, 4 и 8.

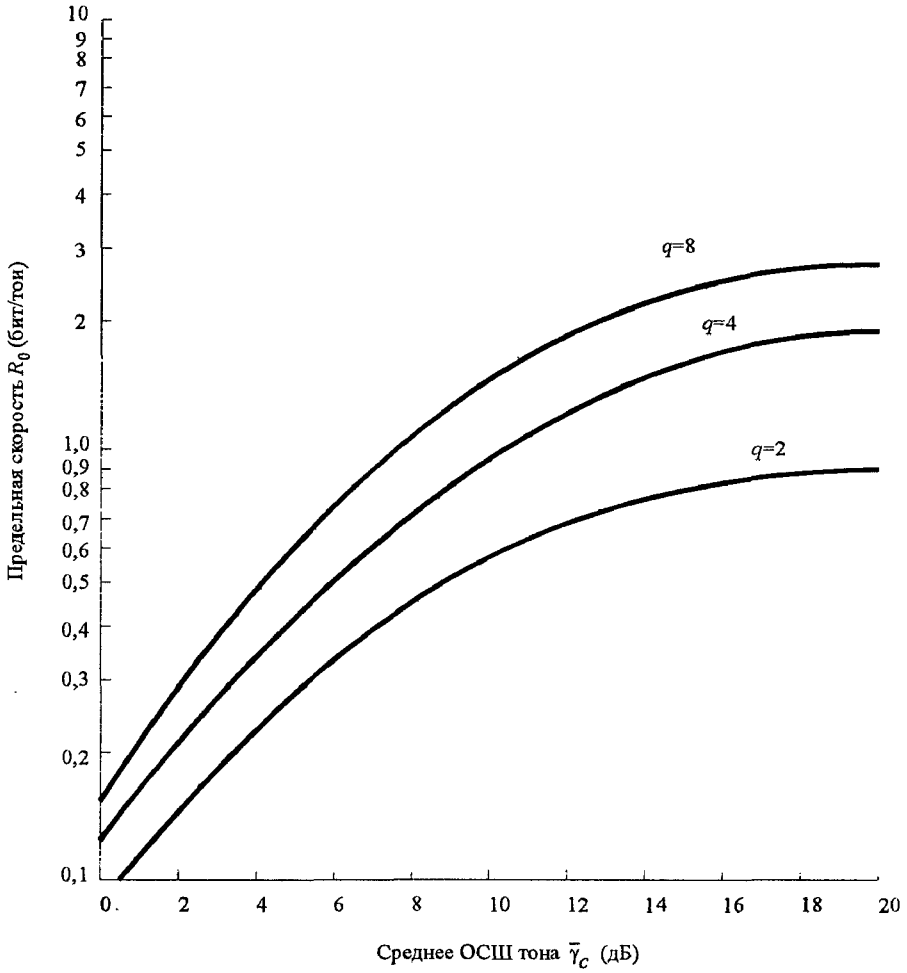


Рис. 14.6.13. Предельная скорость как функция от $\bar{\gamma}_c$ для канала с релеевскими замираниями

Более интересная форма (14.6.26) может получиться, если выразить $\overline{P_M}$ через ОСШ на бит. В частности (14.6.26) можно выразить так

$$\overline{P}_M < 2^{-k[\bar{\gamma}_b g(q,\bar{\gamma}_c)-1]}, \tag{14.6.29}$$

где по определению

$$g(q,\bar{\gamma}_c) = \frac{R_0}{\bar{\gamma}_c} = \frac{1}{\bar{\gamma}_c}\log_2\left[\frac{q}{1+4(q-1)p(1-p)}\right]. \tag{14.6.30}$$

Графики $g(q,\bar{\gamma}_c)$ с параметром q, как функция $\bar{\gamma}_c$ построены на рис.14.6.14.

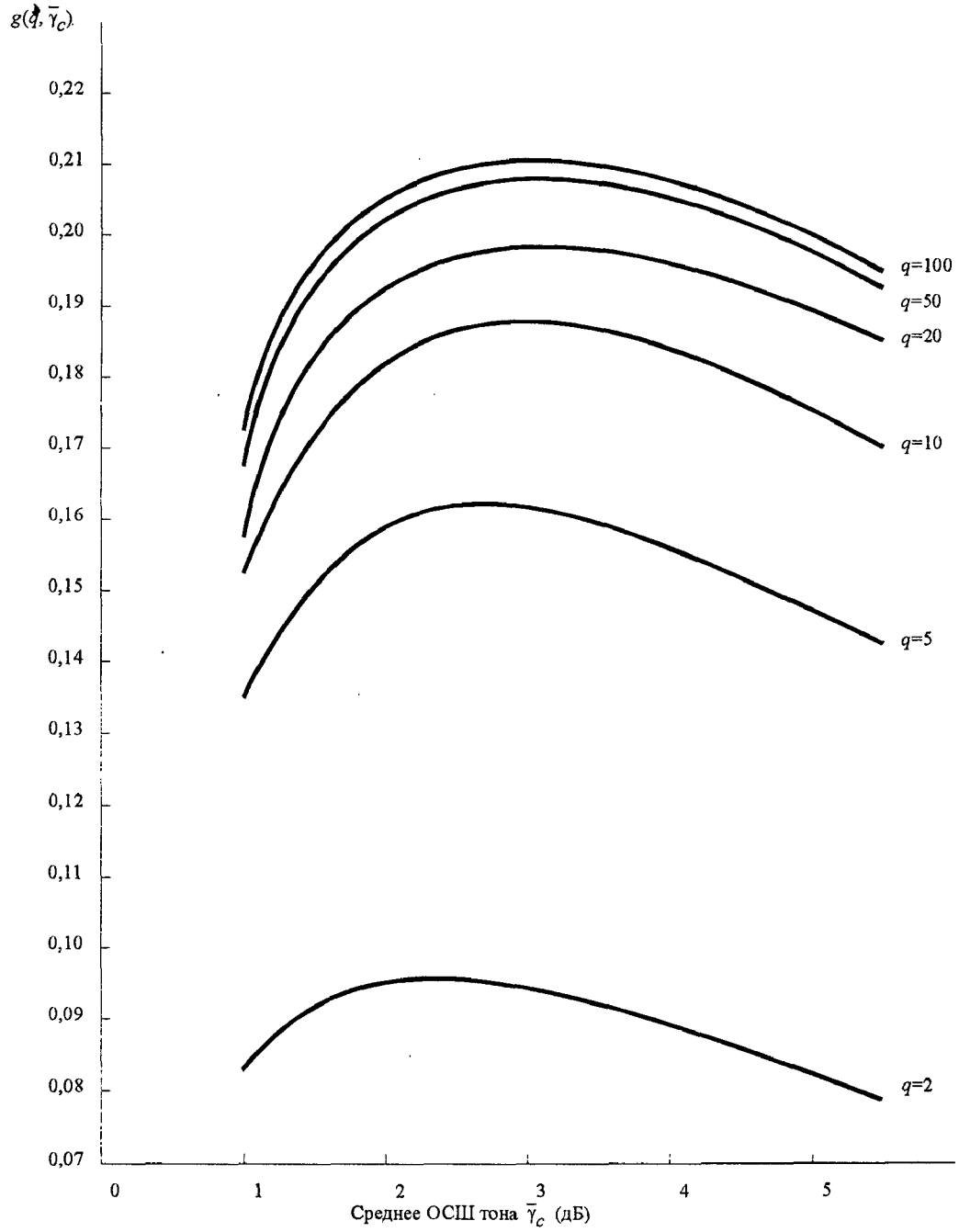


Рис. 14.6.14. График функции $g(q, \bar{\gamma}_c)$

Для начала заметим, что имеется оптимальное значение $\bar{\gamma}_c$ для каждой величины q, которая минимизирует вероятность ошибки. Для больших q эта величина примерно равна $\bar{\gamma}_c = 3$ (5 дБ), что согласуется с нашими прежними наблюдениями для обычного квадратичного сложения. Далее, если $q \to \infty$, функция $g(q, \bar{\gamma}_c)$ стремится к пределу, который равен

$$\lim_{q\to\infty} g(q, \bar{\gamma}_c) = g_\infty(\bar{\gamma}_c) = \frac{1}{\bar{\gamma}_c} \log_2 \left[\frac{(2+\bar{\gamma}_c)^2}{4(1+\bar{\gamma}_c)} \right]. \qquad (14.6.31)$$

Значение $g_\infty(\bar{\gamma}_c)$, рассчитанное при $\bar{\gamma}_c = 3$, равно

$$g_\infty(3) = \max_{\bar{\gamma}_c} g_\infty(\bar{\gamma}_c) = 0{,}215. \qquad (14.6.32)$$

Следовательно, вероятность ошибки в (14.6.29) для этого оптимального разбиения суммарного ОСШ определяется так:

$$\overline{P}_M < 2^{-0,215k(\bar{\gamma}_b - 4,65)}. \qquad (14.6.33)$$

Эти результаты показывают, что вероятность ошибки можно сделать сколь угодно малой с оптимальным ОСШ на кодовый чип, если среднее ОСШ на бит $\bar{\gamma}_b > 4,65$ (6,2 дБ). Даже относительно умеренные значения $q = 20$ приводят близко к этому минимальному значению. Как видно из рис.14.6.14 $g(20,3) = 0,2$, так что $P_M \to 0$ при условии, что $\bar{\gamma}_b > 5$ (7 дБ). С другой стороны, если $q = 2$, максимальная величина для $g(2,\bar{\gamma}_c) \approx 0,096$ и соответствующий минимум ОСШ на бит равен 10,2 дБ.

Для случая двоичных ЧМ символов $(q = 2)$ мы можем легко сравнить предельную скорость для неквантованного выхода (мягких решений) демодулятора с предельной скоростью при двоичном квантовании, для которого

$$R_Q = 1 - \log\left[1 + \sqrt{4p(1-p)}\right], \qquad Q = 2,$$

что дано (8.1.104). Рис.14.6.15 иллюстрирует графики R_0 и R_Q. Заметим, что разница между R_0 и R_Q примерно равна 3 дБ для скоростей ниже 0,3 и разница быстро возрастает при больших скоростях. Эту потерю можно значительно уменьшить увеличением числа уровней квантования до $Q = 8$ (три бита).

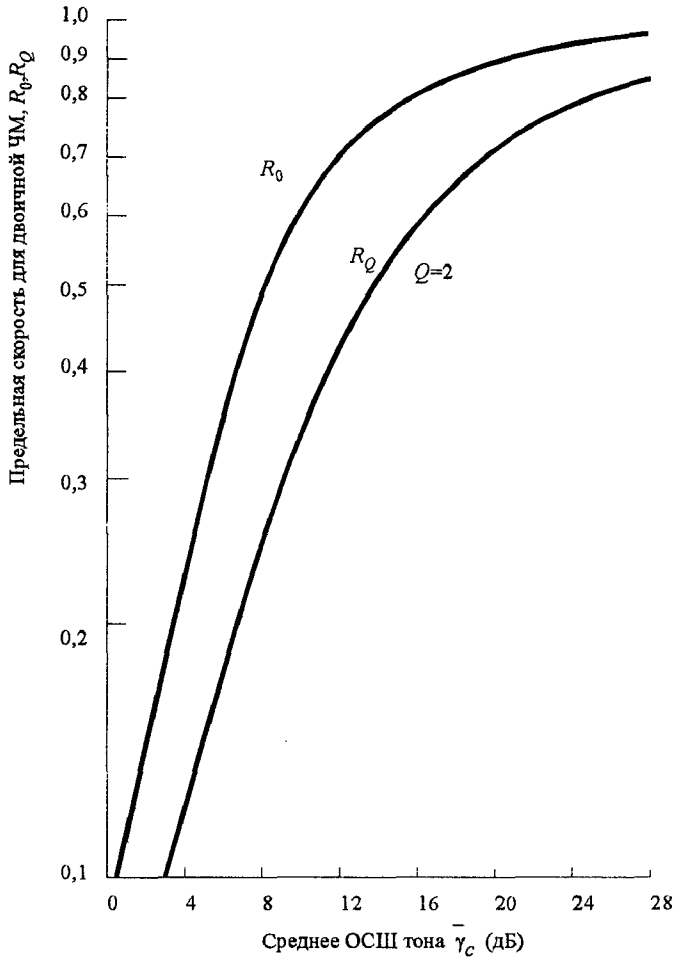


Рис. 14.6.15. Предельная скорость для декодирования мягких решений (неквантованного) и жёстких решений при двоичной ЧМ

Аналогичные сравнения сравнительного качества между не квантованным декодированием мягких решений и декодированием квантованных решений можно сделать при $q > 2$.

14.6.6. Решётчато-кодовая модуляция

Решётчато-кодовая модуляция была описана в разделе 8.3 как средство достижения выигрыша кодирования в частотно-ограниченных каналах, в которых мы хотим передать сигнал с отношением битовой скорости к полосе $R/W > 1$. Для таких каналов цифровые системы связи проектируются так, чтобы использовать частотно-эффективную многоуровневую или многофазную модуляцию (АМ, ФМ, ДФМ или КАМ) которые позволяют нам достичь $R/W > 1$. Если применяется кодирование для синтеза сигнала для частотно-ограниченного канала, то желателен выигрыш кодирования без расширения полосы частот канала. Эту цель можно достичь, как описано в разделе 8.3, путём увеличения числа сигнальных точек созвездия относительно соответствующей некодированной системы, чтобы компенсировать избыточность, введенную кодом, и такого синтеза решётчатого кода, чтобы евклидово расстояние цепочки переданных символов, соответствующих пути, который сливается в любом узле решётки с правильным путём, было бы больше, чем аналогичное расстояние в некодированной системе.

В противоположность этому, схемы кодирования, которые мы описали выше, в соединении с ЧМ расширяют полосу частот модулированного сигнала с целью достижения разнесения сигнала. Соединённые с ЧМ, которая по частоте не эффективна, схемы кодирования, которые мы описали выше, не подходят для использования в частотно-ограниченных каналах.

При синтезе решётчато-кодированных сигналов для каналов с замираниями мы можем использовать те же базовые принципы, которые мы изучили и применили при синтезе схем свёрточного кодирования. В частности, наиболее важная задача при любом синтезе сигналов для каналов с замираниями сводится к достижению наибольшего порядка разнесения сигнала. Это подразумевает, что соседние выходные символы кодера должны быть перемежены или достаточно разделены на передаче во времени, или по частоте, чтобы таким образом достичь независимых замираний в последовательности символов. Следовательно, мы можем представить такую систему решётчато-кодовой модуляции блок-схемой рис.14.6.16, в которой перемежитель рассматривается в широком смысле, как устройство, которое разделяет соседние кодовые символы так, чтобы обеспечить независимые замирания каждого символа (посредством временного или частотного разделения символов) последовательности. Приёмник состоит из демодулятора сигнала, выход которого после деперемежения подаётся на решётчатый декодер.

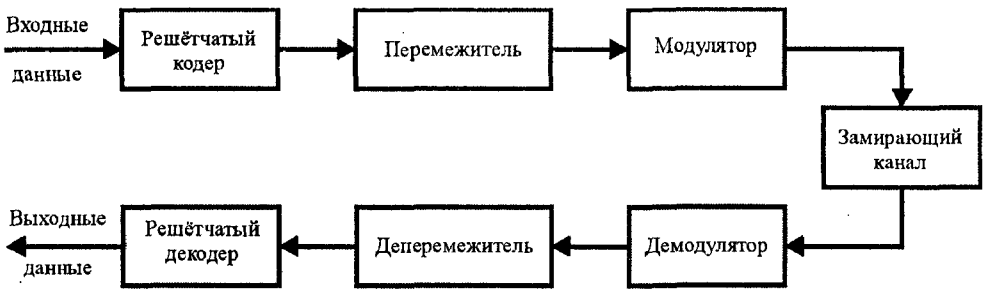


Рис. 14.6.16. Блок-схема системы решётчато-кодовой модуляции

Как показано выше, претендентами на методы модуляции, которые достигают высокую частотную эффективность является M-ичные ФМ, ДФМ, КАМ и АМ. Выбор зависит от большого набора характеристик канала. Если имеются быстрые амплитудные изменения принимаемого сигнала, то КАМ и АМ особенно уязвимы, поскольку потребуется использование широкополосного автоматического управления усилением (АРУ) для компенсации изменений в канале. В таком случае более подходящим являются ФМ или ДФМ поскольку информация содержится в фазе, а не в амплитуде сигнала. ДФМ обеспечивает дополнительную выгоду поскольку когерентность фазы несущей требуется только для двух соседних символов. Однако в ДФМ имеется ухудшение в ОСШ относительно ФМ.

При синтезе решётчатого кода наша цель сводится к достижению возможно большего свободного расстояния, поскольку этот параметр эквивалентен величине порядка разнесения принимаемого сигнала. В обычном решётчатом кодировании Унгербоека каждая ветвь решётки соответствует единственному M-ичному (ФМ, ДФМ, КАМ) выходу канального символа. Определим ошибочное событие с кратчайшим путём как путь при ошибочном событии с наименьшим числом ненулевых расстояний между ним самим и правильным путём, и пусть L – длина этого кратчайшего пути. Другими словами, L – это расстояние Хемминга между M-позиционными символами при ошибочном событии с кратчайшим путём и правильным путем. Если мы предполагаем, что передаваемая последовательность соответствует одним нулям в решётке, L – это число ветвей в кратчайшем пути с ненулевым M-ичным символом. В решётчатой диаграмме с параллельными путями пути ограничены так, чтобы иметь ошибочное событие с кратчайшим путём на одной ветви, так что $L=1$. Это означает, что такой решётчатый код не обеспечивает разнесение в канале с замираниями и, следовательно, вероятность ошибки обратно пропорциональна ОСШ на символ. Следовательно, при свёрточном решётчатом кодировании в канале с замираниями нежелательно синтезировать код, который имеет параллельные пути в решётке, поскольку такой код не даёт разнесения. Это случай свёрточного решётчатого кода со скоростью $m/(m+1)$, который мы заставили иметь параллельные пути, когда число состояний меньше 2^m.

Один довольно эффективный путь к увеличению минимального свободного расстояния и, как следствие, порядка разнесения кода, сводится к введению асимметрии в точках сигнального созвездия. Такой подход был разработан Саймоном и Дивсаларом (1985), Дивсаларом и Юэном (1989), и Дивсаларом и др. (1987).

Более эффективный путь для увеличения расстояния L и, как следствие, порядка разнесения сводится к использованию множественной решётчато-кодовой модуляции (МРКМ). В МРКМ, иллюстрируемой на рис.14.6.17, b входных бит кодируются в c выходных символов, которые затем делятся на k групп, каждая из m бит, так что $c=km$.

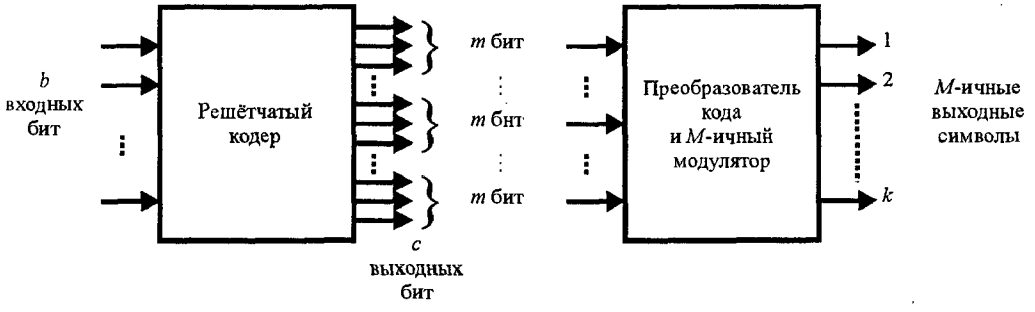


Рис. 14.6.17. Блок-схема МРКМ передатчика

Каждая m-символьная группа отображается в M-ичный символ. Таким образом, мы получаем M-ичный выходной символ. Частный случай $k=1$ соответствует свёрточному коду Унгербоека При помощи k M-ичных выходных символов возможно синтезировать решётчатый код с параллельными путями, имеющий расстояние $L=k$. Таким образом, мы можем достичь вероятность ошибки, которая уменьшается обратно пропорционально $(\mathscr{E}/N_0)^k$.

Важное обстоятельство в синтезе декодера для решётчатого кода сводится к использованию сторонней информации относительно ослабления в канале для каждого символа. В случае ЧМ с квадратичным сложением в декодере для образования метрики не требуется знания ослабления канала для демодулированных символов. Однако при когерентном детектировании, оптимальная метрика евклидового расстояния для каждого демодулируемого символа определяется в форме $|r_n-\alpha_n s_n|^2$, где α_n – ослабление в канале для переданного символа s_n, а r_n – выход демодулятора. Следовательно, сумма метрик путей для любого заданного пути по решётке определится в форме

$$D(\mathbf{r},\mathbf{s}^{(i)})=\sum_n |r_n-\alpha_n s_n^{(i)}|^2,$$

где соответствующий верхний индекс (i) указывает i-й путь по решётке. Следовательно, надо выполнить оценивание канальных ослаблений, чтобы реализовать оптимальный решётчатый декодер. Оценивание канального ослабления и фазовых сдвигов рассматривается в приложении С для случая фазовой модуляции и демодуляции. Влияние качества оценивания ослаблений и фазовых сдвигов на качество ФМ (не кодированной) также оценивается в приложении С.

14.7. БИБЛИОГРАФИЧЕСКИЕ ЗАМЕЧАНИЯ И ССЫЛКИ

В этой главе мы рассмотрели совокупность тем, связанных с цифровой связью по многопутевому каналу с замираниями. Мы начали с характеристики статистики канала и затем описали набор канальных характеристик для синтеза цифровых сигналов и анализа их качества. Мы видели, что надёжность систем связи увеличивается при использовании разнесенной передачи и приема. В заключении мы показали, что канальное кодирование и декодирование мягких решений обеспечивает частотно-эффективный способ для получения разнесения по таким каналам.

Основополагающие работы по характеристике многопутевых каналов с замираниями и по синтезу сигналов и приёмников для реализуемых цифровых систем по таким каналам были выполнены Прайсом (1954, 1956). За этим работами последовали значительные вклады Прайса и Грина (1958, 1960), Кайлата (1960, 1961) и Грина (1962). Техника разнесенной передачи и техника суммирования при разнесении при изменений условий в канале были рассмотрены в статьях Пирса (1958), Бренанна (1959), Турина (1961, 1962), Пирса и Стейна (1960), Барроу (1963) Белло и Нелина (1962a,b, 1963), Прайса (1962a,b) и Линдсея (1964)[1].

[1] В 1957 г. Л.М. Финк определил потенциальную помехоустойчивость при разнесённом приёме в однопутевом канале с АБГШ и релеевскими замираниями. В 1958 г. Д.Д. Кловский предложил адаптивную систему радиосвязи (для каналов с МСИ) с периодическим зондированием канала СИИП (систему с испытательным импульсом и предсказанием), а в 1959 г. им же определена потенциальная помехоустойчивость при разнесённом приёме в однопутевом канале с АБГШ при райсовских замираниях в канале и при распределении амплитуд по преобразованному хи-квадрат распределению. Начиная с 1959 г., когда Д.Д. Кловский предложил оптимальный (по правилу МП) алгоритм демодуляции для каналов с МСИ (с

Наша трактовка кодирования для каналов с замираниями основывается на вкладе большого числа исследователей. В частности, использование k-дуальных кодов и M-ичной ортогональной ЧМ были предложены в публикациях Витерби и Джекобса (1975) и Оденвальдера (1976). Важность кодирования для цифровой связи по каналам с замираниями было также подчеркнуто в статье Чейза (1976). Выигрыш, достигаемый каскадными кодами с декодированием мягких решений для канала с замираниями, был продемонстрирован Пайпером и др. (1978). Здесь код Рида-Соломона был использован в качестве внешнего кода, а код Адамара был выбран как внутренний код. Качество k-дуальных кодов с блоковыми ортогональными кодами или кодами Адамара в качестве внутреннего кода было исследовано Прокисом и Раманом (1979). Вероятности ошибок для двоичных свёрточных кодов с максимальным свободным расстоянием были рассчитаны Раманом (1981). Наконец, расчёт предельно достижимой скорости для каналов с релеевскими замираниями выполнено Возенкрафтом и Джекобсом (1965)[1].

Решётчато-кодовая модуляция для каналов с замираниями была разработана многими исследователями, чьи работы были мотивированы большим числом приложений к мобильной и сотовой связи. Книга Биглиери и др. (1991) даёт учебную трактовку этой темы и содержит большое число ссылок на техническую литературу.

Наша трактовка цифровой связи по каналам с замираниями ориентирована прежде всего на модель канала с релеевскими замираниями. В большей части это сделано из-за широкого распространения этой модели для описания влияния замираний на многих радиоканалах и вследствие простоты её исследования. Хотя другие статистические модели, такие, как модель Райса или модель Накагами для каналов с замираниями более подходят для характеристики замираний по некоторым реальным каналам, общий подход в синтезе надёжной связи, представленный в этой главе, их покрывает.

ЗАДАЧИ

14.1. Функция рассеяния $S(\tau,\lambda)$ для многопутевого канала с замираниями не равна нулю в области $0 \le \tau \le 1$ мс и $-0{,}1$ Гц $\le \lambda \le 0{,}1$ Гц. Предположим, что функция рассеяния приближенно равномерна по двум переменным.

a Дайте численные значения для следующих параметров:
- (i) многопутевое рассеяние канала;
- (ii) доплеровский сдвиг в канале;
- (iii) интервал временной когерентности канала;
- (IV) полоса частотной когерентности канала;
- (V) фактор рассеяния канала.

b Объясните значение следующих ответов, взяв во внимание ответы, данные в (a):
- (i) канал неселективен по частоте;
- (ii) канал с медленными замираниями;
- (iii) канал селективен по частоте;

эхо-сигналами) и АБГШ с ОСР при анализе на тактовом интервале, им исследовалась сравнительная помехоустойчивость алгоритма в составе системы СИИП и других цифровых систем радиосвязи, предложенных к тому времени [25, 74, 75, 66]. АКН и его сравнительная помехоустойчивость, а также вычислительная сложность по сравнению с АВ в многопутевых каналах с замираниями рассматривались в [66, 68 и др.]. Для радиоканалов с замираниями и коррелированными помехами весьма эффективной оказывается пространственная (ПВ) обработка сигналов. Одна из первых монографий по оптимальной ПВ обработке сигнала принадлежит Д.Д. Кловскому и В.А. Сойферу [69]. В дальнейшем эта тематика разрабатывалась, в том числе и для каналов с памятью (МСИ) в работах Д.Д. Кловского, В.Г. Карташевского, С.М. Широкова и В.Я. Конторовича [66, 70, 76, 77].

[1] Модификация АКН для декодирования мягких решений в канале с МСИ и АБГШ при свёрточном кодировании и отсутствии перемежения кодовых символов рассматривалась в [78], а при перемежении кодовых символов в [79].

c Предположим, что мы имеем в канале полосу частот 10 кГц и мы желаем передавать по этому каналу информацию со скоростью 100 Бит/с. Синтезируйте двоичную систему связи с частотным разнесением. В частности для случая (i) тип модуляции, для (ii) число подканалов, для (iii) частотный разнос между соседними несущими и (IV) – сигнальный интервал, используемый в Вашем синтезе. Объясните Ваш выбор параметров.

14.2. Рассмотрите двоичную систему связи для передачи двоичных последовательностей по каналу с замираниями. Модулятор выдаёт ортогональные сигналы ЧМ и обеспечивает частотное разнесение третьего порядка $(L=3)$. Демодулятор состоит из согласованных фильтров, а за ними следуют квадратичные детекторы. Предположите, что несущие ЧМ замирают независимо и по одинаковому закону с релеевским распределением огибающей. Аддитивный шум в каналах разнесения гауссовской с нулевым средним и автокорреляционной функцией $\frac{1}{2}E\left[z_k^*(t)z_k(t+\tau)\right]=N_0\delta(\tau)$. Шумовые процессы в отдельных каналах взаимно-независимые.

a Передаваемый сигнал можно рассматривать как двоичную ЧМ с квадратичным детектированием, который генерируется кодом с повторением вида

$$1\rightarrow C_1=[1\ 1\ 1],\qquad 0\rightarrow C_0=[0\ 0\ 0]$$

Определите вероятность ошибки P_{2h} для декодера жёстких решений, следующего после квадратичного детектирования сигналов.

b Рассчитайте P_{2h} для $\bar{\gamma}_c$ = =100 и 1000.

c Рассчитайте вероятность ошибки P_{2s} для $\bar{\gamma}_c=100$ и 1000, если используется декодирование мягких решений.

d Рассмотрите обобщение результата (a). Если используется код с повторением с длиной блока L (L четно) определите вероятность ошибки P_{2h} для декодера жёстких решений и сравните с вероятностью ошибки P_{2s} декодера мягких решений. Предположите, что $\gamma_c>>1$.

14.3. Предположите, что двоичный сигнал $s_1(t)$ передаётся по каналу с замираниями, а принимаемый сигнал

$$r(t)=\pm as_1(t)+z(t),\qquad 0\le t\le T,$$

где $z(t)$ – белый гауссовский шум с нулевым средним и автокорреляционной функцией

$$\phi_{zz}(\tau)=N_0\delta(\tau).$$

Энергия переданного сигнала $\mathscr{E}=\frac{1}{2}\int_0^T|s_1(t)|^2\,dt$. Канальное ослабление a определяется ФПВ

$$p(a)=0{,}1\delta(a)+0{,}9\delta(a-2)$$

a Определите среднюю вероятность ошибки P_2 для демодулятора, который использует фильтр, согласованный с $s_1(t)$.

b Какое значение примет P_2, когда $\mathscr{E}/N_0$ стремится к бесконечности.

c Предположите, что один и тот же информационный сигнал передаётся по двум статистически независимым каналам с замираниями и ослаблениями a_1 и a_2, где

$$p(a_k)=0{,}1\delta(a)+0{,}9\delta(a-2),\qquad k=1,\ 2.$$

Шум в двух каналах статистически независим и одинаково распределен. Демодулятор использует согласованный фильтр в каждом канале и просто суммирует выходы двух фильтров для формирования величины для решения. Определите среднее значение P_2.

d Для случая c найдите значение для P_2, когда $\mathscr{E}/N_0$ стремится к бесконечности.

14.4. Многопутевой канал с замираниями имеет временное рассеяние $T_m=1$ с и доплеровское рассеяние $B_d=0{,}01$ Гц. Суммарная полоса частот, используемая сигналом для передачи сигнала, равна $W=5$ Гц. Чтобы уменьшить влияние МСИ проектировщик выбирает сигнальный импульс длительностью $T=10$ с.

a Определите полосу частотной когерентности и интервал временной когерентности.

b Является ли канал селективным по частоте? Объясните.

c Являются ли замирания в канале медленными или быстрыми? Объясните.

d Предположим, что канал используется для передачи двоичных данных посредством когерентного детектирования ФМ при частотном разнесении. Объясните, как Вы используете доступную полосу частот канала для получения частотного разнесения и определите какой порядок разнесения возможен.

e Для случая (d), какое приближенное значение ОСШ требуется на канал разнесения, чтобы достичь вероятности ошибки 10^{-6} ?

f Предположите, что широкополосный сигнал используется для передачи, а для демодуляции используется приёмник типа Rake. Сколько ячеек Вы используете в приёмнике Rake?

g Объясните, можно или нельзя реализовать приёмник Rake как когерентный приёмник с суммированием максимальных отношений.

h Если используются ортогональные сигналы для широкополосного сигнала с квадратичным последетекторным суммированием в приёмнике Rake, какие приближённые значения ОСШ требуются для достижения вероятности ошибки 10^{-6}? (предположите, что во всех отводах одинаковое ОСШ).

14.5. В двоичной системе связи, показанной на рис. Р.14.5, $z_1(t)$ и $z_2(t)$ – статистически независимые гауссовские шумовые процессы с нулевым средним и одинаковой функцией автокорреляции $\phi_{zz}(\tau)=N_0\delta(\tau)$.

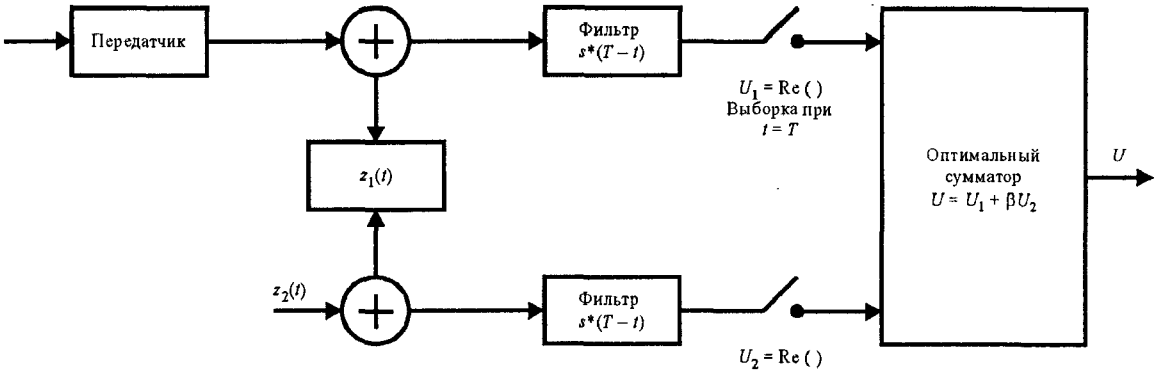


Рис. Р14.5

Отсчетные величины U_1 и U_2 представляют реальные части выходов согласованных фильтров. Для примера, если передаётся $s_l(t)$, тогда имеем

$$U_1 = 2\mathscr{E} + N_1$$
$$U_2 = N_1 + N_2,$$

где $\mathscr{E}$ – энергия, переданная сигналом, и

$$N_k = \mathrm{Re}\left[\int_0^T s_l^*(t) z_k(t)\,dt\right], \qquad k = 1, 2.$$

Очевидно, что U_1 и U_2 являются коррелированными гауссовскими величинами, в то время как N_1 и N_2 – независимые гауссовские величины с распределением

$$p(n_1)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{n_1^2}{2\sigma^2}\right)$$
$$p(n_2)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{n_2^2}{2\sigma^2}\right),$$

где дисперсия N_k равна $\sigma^2 = 2\mathscr{E}N_0$

a Покажите, что совместная ФПВ для U_1 и U_2 равна

$$p(U_1,U_2)=\frac{1}{2\pi\sigma^2}\exp\left\{-\frac{1}{2\sigma^2}\left[(U_1-2\mathscr{E})^2-U_2(U_1-2\mathscr{E})+\frac{1}{2}U_2^2\right]\right\},$$

если передаётся $s(t)$ и

$$p(U_1,U_2)=\frac{1}{2\pi\sigma^2}\exp\left\{-\frac{1}{2\sigma^2}\left[(U_1+2\mathscr{E})^2-U_2(U_1+2\mathscr{E})+\frac{1}{2}U_2^2\right]\right\}$$

если передаётся $-s(t)$.

b Основываясь на отношении правдоподобия покажите, что оптимальное сложение U_1 и U_2 приводит к величине для решения

$$U = U_1 + \beta U_2,$$

где β – константа. Каково оптимального значения β ?

c Предположите, что передаётся $s(t)$. Какова плотность вероятности U ?

d Какова вероятность ошибки при предположении, что было передано $s(t)$? Выразите Ваш ответ, как функцию от ОСШ $\mathcal{E}/N_0$.

e Какова потеря качества в случае, когда $U=U_1$ является величиной для решения?

14.6. Рассмотрите модель двоичной системы связи с разнесением, показанной на рис. Р.14.6.

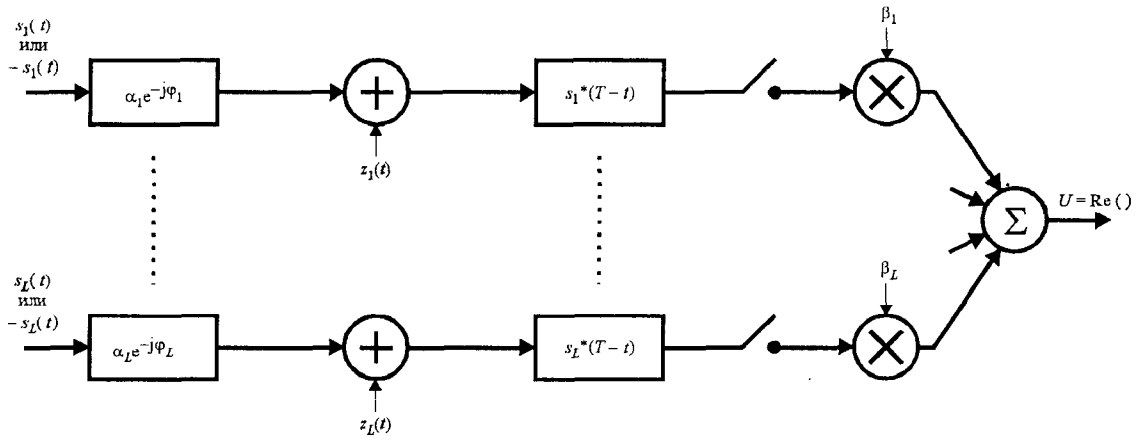


Рис. Р14.6

Каналы имеют фиксированные ослабления и фазовые сдвиги. $\{z_k(t)\}$ – комплексные белые гауссовские шумовые процессы с нулевыми средними и корреляционными функциями

$$\phi_{zz}(t) = \tfrac{1}{2} E\left[z_k^*(t) z_k(t+\tau)\right] = N_{0k}\delta(\tau).$$

(Заметим, что спектральные плотности $\{N_{0k}\}$ различны). Шумовые процессы $\{z_k(t)\}$ статистически взаимно независимы. $\{\beta_k\}$ – комплексные взвешивающие множители, которые должны быть определены. Величина решения на выходе схемы сложения

$$U = \mathrm{Re}\left(\sum_{k=1}^{L} \beta_k U_k\right) \underset{-1}{\overset{1}{\gtrless}} 0$$

a определите ФВП $p(U)$, когда передача $+1$.

b определите вероятность ошибки P_2, как функцию от весов $\{\beta_k\}$.

c определите величины $\{\beta_k\}$, которые минимизируют P_2.

14.7. Определите вероятность ошибки для двоичной ортогональной системы сигналов с разнесением L порядка в канале с релеевскими замираниями. ФПВ двух величин для решения даны (14.4.31) и (14.4.32).

14.8. Двоичный свёрточный код с передаточной функцией, определенной (8.2.5), со скоростью 1/3, кодовым ограничением $\nu = 3$ используется для передачи данных по каналу с релеевскими замираниями посредствам двоичной ФМ.

a Определите и постройте вероятность ошибки при декодировании жёстких решений. Предположите, что передаваемые сигналы, соответствующие кодовым символам замирают независимо.

b Определите и постройте вероятность ошибки при декодировании мягких решений. Предположите что сигналы, соответствующие кодовым символам, замирают независимо.

14.9. Двоичная последовательность передаётся двоичными противоположными сигналами по каналу с релеевскими замираниями с разнесением L-го порядка. Когда передаётся $s_l(t)$ принимаемые эквивалентные низкочастотные сигналы равны

$$r_k(t) = \alpha_k e^{-j\phi_k} s_l(t) + z_k(t), \qquad k = 1, 2, \ldots, L.$$

Замирания в L подканалах статистически независимы. Слагаемые аддитивных шумов $\{z_k(t)\}$ являются независимыми и одинаково распределенными белыми гауссовскими процессами с нулевыми средними

значениями и автокорреляционной функцией $\phi_{zz}(t) = N_0\delta(\tau)$. На приёме каждый из L сигналов проходит через фильтр, согласованный с $s_l(t)$, а выход корректируется по фазе, чтобы получить

$$U_k = \mathrm{Re}\left[e^{j\phi_k} \int_0^T r_k(t) s_l^*(t) dt \right], \qquad k = 1, 2, ..., L.$$

$\{U_k\}$ суммируются линейно для формирования величины для решения

$$U = \sum_{k=1}^{L} U_k .$$

a Определите ФПВ для U при условии фиксированных величин $\{a_k\}$.

b Определите выражение для вероятности ошибки, когда $\{a_k\}$ статистически независимые и одинаково распределенные релеевские случайные величины.

14.10. Было показано, что граница Чернова для вероятности ошибки двоичной ЧМ при L кратном разнесении в канале с релеевскими замираниями можно выразить так

$$P_2(L) < [4p(1-p)]^L = \left[4\frac{1+\bar{\gamma}_c}{(2+\bar{\gamma}_c)^2} \right]^L < 2^{-\bar{\gamma}_b g(\bar{\gamma}_c)},$$

где

$$g(\bar{\gamma}_c) = \frac{1}{\bar{\gamma}_c} \log_2 \left[\frac{(2+\bar{\gamma}_c)^2}{4(1+\bar{\gamma}_c)} \right].$$

a Нарисуйте $g(\bar{\gamma}_c)$ и определите приближенно его максимальное значение и величину $\bar{\gamma}_c$, при которой наступает максимум.

b Для заданного $\bar{\gamma}_b$ определите оптимальный порядок разнесеиия.

c Сравните P_2(L) при условии максимизации $g(\bar{\gamma}_c)$ (оптимальное разнесение) с вероятностью ошибки двоичной ЧМ в канале с АБГШ баз замираний, которое равно

$$P_2 = \tfrac{1}{2} e^{-\gamma_b/2}$$

и определите потери в ОСШ, обусловленные замираниями и некогерентным (квадратичным) суммированием.

14.11. ПП система используется для разделения многопутевых компонент в двухпутевом радиоканале. При условии, что длина второго пути на 300 м длиннее прямого пути, определите минимально необходимую скорость чипов для разделения многопутевых компонент.

14.12. Низкочастотная (базовая) цифровая система связи использует сигналы, показанные на рис. P14.12(a), для передачи двух равновероятных сообщений. Предполагается, что проблема связи, которая изучается здесь, это проблема связи «одного выстрела». Это значит, указанные сообщения передаются именно один раз, а после этого и передача не имеет места. Канал не имеет ослабления $(\alpha = 1)$, а шум АБГШ со спектральной плотностью мощности $\frac{1}{2}N_0$.

a Найдите подходящий ортонормированный базис для представления сигналов

b На блок-схеме дайте подробную спецификацию оптимального приёмника с согласованными фильтрами. Тщательно обозначьте блоки.

c Найдите вероятность ошибки оптимального приёмника.

d Покажите, что оптимальный приёмник можно реализовать, используя только один фильтр (смотрите блок-схему рис. P14.12(в)). Каковы характеристики согласованного фильтра, стробирующего устройства и решающего блока?

e Теперь предположите, что канал не идеальный, но имеет импульсную характеристику $c(t) = \delta(t) + \frac{1}{2}\delta\left(t - \frac{1}{2}T\right)$. Используя тот же согласованный фильтр, что в (d), синтезируйте оптимальный приёмник.

f Предположите, что импульсная характеристика равна $c(t) = \delta(t) + a\delta\left(t - \frac{1}{2}T\right)$, где a – случайная величина с равномерным распределением на отрезке $[0, 1]$. Используя тот же фильтр, что в (d), синтезируйте оптимальный приёмник.

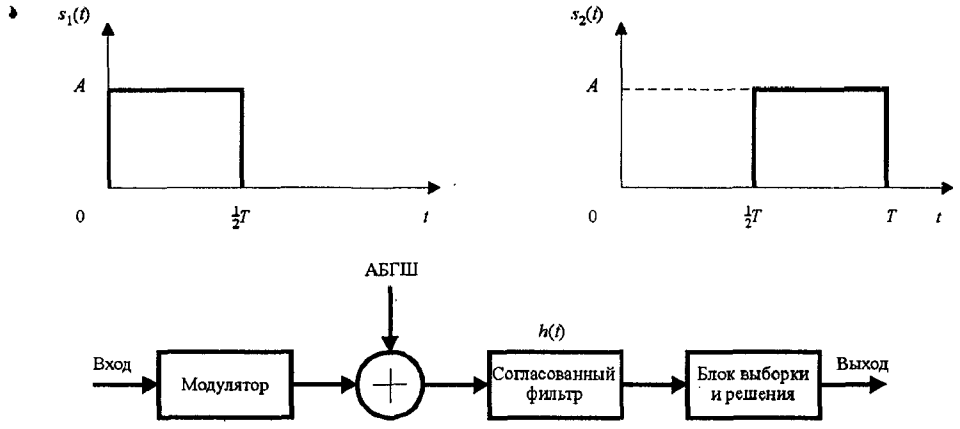


Рис. Р14.12

14.13. Система связи использует две разнесённые антенны и двоичные ортогональные сигналы ЧМ. Принимаемые сигналы в двух антеннах равны

$$r_1(t) = \alpha_1 s(t) + n_1(t)$$
$$r_2(t) = \alpha_2 s(t) + n_2(t),$$

где α_1 и α_2 – статистически независимые случайные величины с распределением Релея, $n_1(t)$ и $n_2(t)$ – статистически независимые белые гауссовские случайные процессы с нулевым средним и спектральной плотностью мощности $\frac{1}{2}N_0$. Два сигнала демодулируются, квадратируются и затем суммируются до детектирования.

a Нарисуйте функциоиальную блок-схему полного приёмника, включающую демодулятор, устройство сложения и детектор.

b Нарисуйте график вероятности ошибки детектора и сравните этот результат со случаем отсутствия разнесения.

14.14. Два эквивалентных низкочастотных сигнала, показанных на рис. Р14.14, используются для передачи двоичной последовательности.

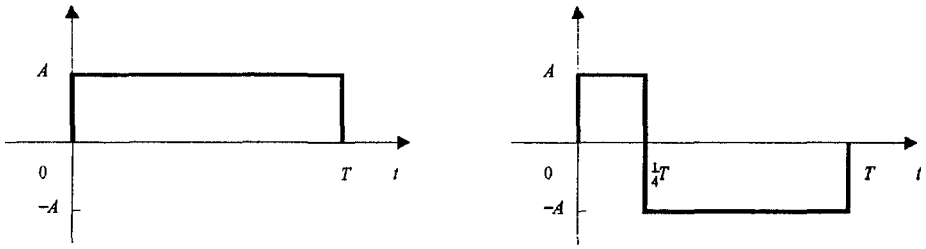


Рис. Р14.14

Эквивалентная низкочастотная импульсная характеристика канала $h(t) = 4\delta(t) - 2\delta(t-T)$. Чтобы избежать перекрытия импульсов между соседними передачами, скорость передачи выбирается как $R = 1/2T$. Передаваемые сигналы равновероятны и искажаются аддитивным белым гауссовским шумом с нулевым средним, имеющий эквивалентное низкочастотное представление $z(t)$ с автокорреляционной функцией

$$\phi_{zz}(t) = \tfrac{1}{2}E\left[z_k^*(t)z_k(t+\tau)\right] = N_0\delta(\tau).$$

a Нарисуйте два возможных эквивалентных принимаемых низкочастотных сигнала (без помех).

в Определите оптимальный приёмник и нарисуйте эквивалентные низкочастотные отклики всех фильтров, используемых в оптимальном приёмнике. Предположите когерентное детектирование сигналов.

14.15. Подтвердите соотношение (14.3.14) путём замены переменной $\gamma = \alpha^2 \mathcal{E}_b / N_0$ в m-распределении Накагами.

15

СИСТЕМЫ СВЯЗИ СО МНОГИМИ ПОЛЬЗОВАТЕЛЯМИ

Наша трактовка систем связи до сих пор была сфокусирована на единственную линию связи, включающую передатчик и приёмник. В этой главе внимание уделяется многим пользователям и многим линиям связи. Мы исследуем различные пути, посредством которых многие пользователи получают доступ в общий канал для передачи информации. Методы множественного доступа, которые описываются в этой главе, образуют основу для современных и будущих проводных и беспроводных сетей связи, таких как сети космической связи, сети сотовой и мобильной связи и сетей подводной акустической связи.

15.1. ВВЕДЕНИЕ В ТЕХНИКУ МНОЖЕСТВЕННОГО ДОСТУПА

Полезно различать несколько типов систем связи со многими пользователями. Один тип – это система множественного доступа, в которой большое число пользователей занимает общий канал связи для передачи информации к приёмнику. Такая система изображена на рис.15.1.1.

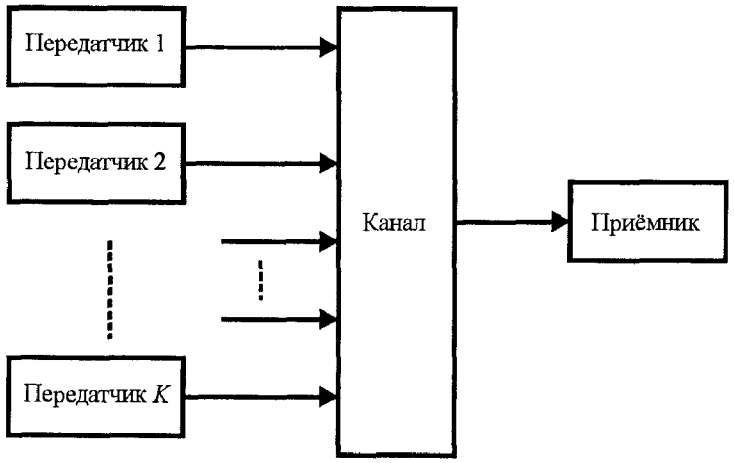


Рис. 15.1.1. Система с множественным доступом

Общий канал может быть средой в космической системе связи, или кабелем, к которому присоединен ряд терминалов, которые имеют доступ к центральному компьютеру, или некоторая полоса частот в радиоспектре, которая используется многими пользователями для связи с радиоприёмником. Для примера, в мобильной сотовой системе связи пользователями являются мобильные передатчики в некоторой частной соте системы, а приёмник находится на базовой станции частной соты.

Второй тип системы связи со многими пользователями является сеть вещания, в которой отдельный передатчик передает информацию многим приёмникам, как показано на рис.15.1.2. Различные виды вещательных систем включают общий радиоканал и системы телевизионного вещания, как и в космических системах.

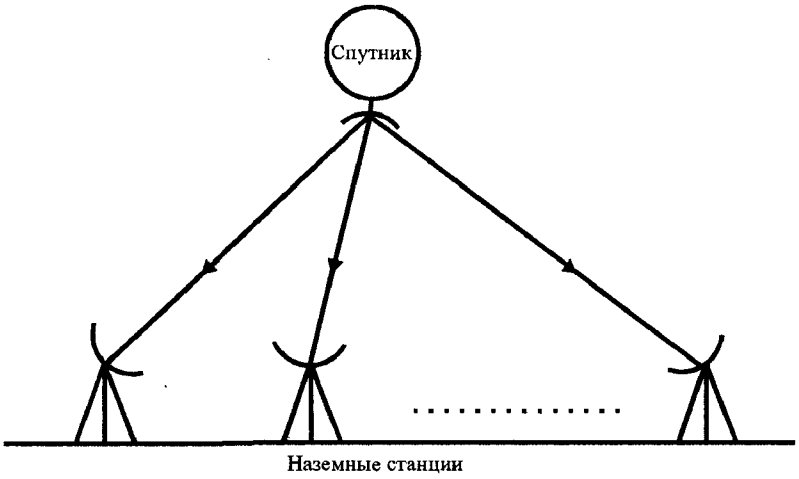


Рис. 15.1.2. Сеть вещания

Множественный доступ и сети вещания образуют, вероятно, наиболее общую систему связи со многими пользователями.

Третий тип системы со многими пользователями – это сети накопления-передачи, как показано на рис.15.1.3. Четвертым типом являются двусторонние (дуплексные) системы связи, показанные на рис.15.1.4.

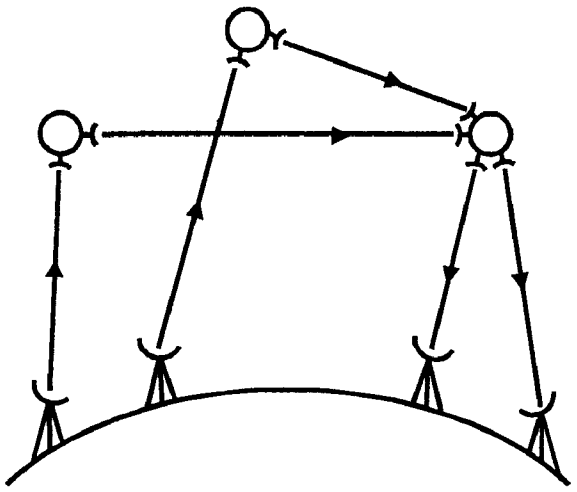


Рис. 15.1.3. Сеть сбора и передачи информации
со спутниковыми ретрансляторами

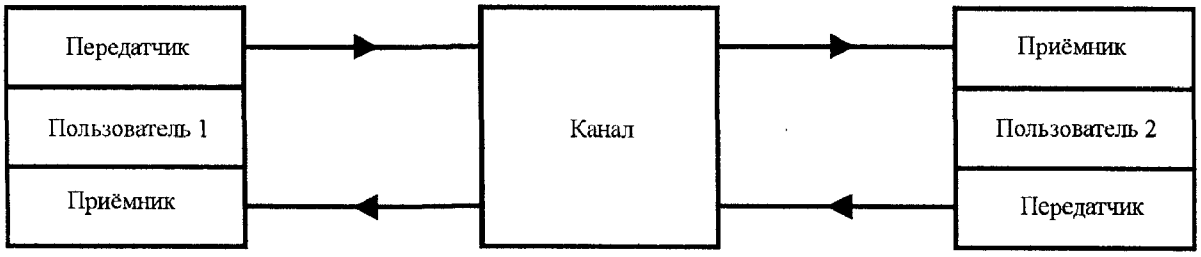


Рис. 15.1.4. Двусторонний канал связи

В этой главе мы сконцентрируем внимание на методы множественного доступа для связи со многими пользователями. В общем, имеется несколько различных путей,

посредством которых многие пользователи могут посылать информацию через канал связи на приёмник. Один простой метод сводится к разделению доступной полосы частот канала на определенное число N неперекрываемых частотных подканалов, как показано на рис.15.1.5, и назначению подканала каждому пользователю по его требованию. Этот метод в общем называется *множественный доступ с частотным разделением* [МДЧР-FDMA], и он обычно используется в непроводных каналах, чтобы разместить многих пользователей передачи речи и данных.

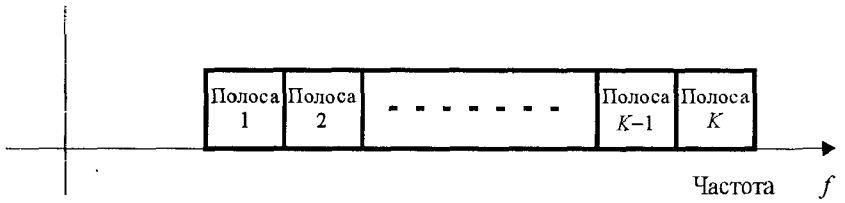


Рис. 15.1.5. Разбиение канала на неперекрывающиеся частотные полосы

Другой метод для создания многих подканалов для множественного доступа сводится к делению длительности T_f, называемой рамочной (каркасной) длительностью, на, скажем, N неперекрывающихся интервалов, каждый длительностью T_f/N. Каждому пользователю, который желает передать информацию, выделяется частный временной интервал внутри каждого каркаса. Этот метод множественного доступа *назван множественным доступом с временным разделением* [МДВР-TDMA] и он часто используется при передаче данных и речи.

Мы видим, что в FDMA и TDMA каналы принципиально подразделяются на независимые подканалы, выделяемые отдельным пользователям. В этом смысле методы синтеза систем связи, которые мы описали для единственного пользователя системы, непосредственно применимы и не возникают новые проблемы в обстановке множественного доступа, исключая дополнительной задачи закрепления пользователей к возможным каналам.

Интересная проблема возникает, когда данные пользователей, поступая в сеть, образуют по своей природе вспышки. Другими словами, информация, переданная от одного пользователя, отделяется периодами, когда передачи информации от этого пользователя нет, причём периоды молчания могут быть дольше периода передачи. Так обычно обстоят дела с пользователями различных терминалов компьютерной сети связи, которая содержит центральный компьютер. В определенной степени это также имеет место в мобильных сотовых системах связи, передающих оцифрованный голос, поскольку для сигналов речи типично наличие длинных пауз.

В обстановке, когда передача от различных пользователей образует вспышки, низкоскоростной цикл FDMA и TDMA может быть неэффективным, поскольку определенный процент предоставляемых частотных и временных интервалов не переносят информацию. В конечном счёте, неэффективный синтез систем множественного доступа ограничивает число одновременных пользователей для канала.

Альтернативой FDMA и TDMA заключается в том, чтобы позволить больше чем одному пользователю владеть совместно каналом или подканалом путем использования прямых последовательностей (ПП) широкополосных сигналов. В этом методе каждому пользователю присваивается уникальная кодовая последовательность *или адресная последовательность (последовательность подписи)*, которая позволяет пользователю рассеять информационный сигнал по выделенной полосе частот. Сигналы различных

пользователей разделяются на приёме посредством взаимной корреляции принимаемого сигнала с адресной последовательностью каждого из пользователей. Синтезируя эти кодовые последовательности с относительно малыми взаимными корреляциями, можно минимизировать переходную помеху, возникающую при демодуляции сигналов, принимаемых от множества передатчиков. Этот метод множественного доступа назван множественным доступом с кодовым разделением [МДКР–CDMA].

В CDMA пользователи поступают в канал случайным образом. Следовательно, передаваемые сигналы от многих пользователей полностью перекрываются во времени и частоте. Демодуляция и разделение этих сигналов на приёме облегчается тем фактом, что каждый сигнал рассеян по частоте посредством псевдослучайной кодовой последовательности. CDMA называют множественным доступом с рассеянным спектром [МДРС-SSMA].

Альтернатива CDMA – случайный доступ без рассеяния сигнала по спектру. В этом случае, когда два пользователя пытаются использовать общий канал одновременно, их передачи сталкиваются и интерферируют друг с другом. Если это случается, информация теряется и должна быть снова передана. Чтобы обсудить возникающие ситуации, следует устанавливать протоколы для повторных передач (ретрансляции) сообщений, которые сталкиваются. Протокол для расписания прохождения сталкивающихся сообщений описывается ниже.

15.2. ПРОПУСКНАЯ СПОСОБНОСТЬ МЕТОДОВ МНОЖЕСТВЕННОГО ДОСТУПА

Интересно сравнить FDMA, TDMA и CDMA по информационной скорости, которую каждый из методов множественного доступа достигает в идеальном канале с полосой частот W и АБГШ. Сравним пропускную способность K пользователей, где каждый пользователь имеет среднюю мощность $P_i = P$ для всех $1 \le i \le K$. Напомним, что в идеальном частотно-ограниченном канале с полосой W и АБГШ пропускная способность одного пользователя равна

$$C' = W \log_2\left(1 + \frac{P}{WN_0}\right), \qquad (15.2.1)$$

где $\frac{1}{2}N_0$ – спектральная плотность аддитивного шума. В FDMA каждый пользователь локализован в полосе W/K. Следовательно, пропускная способность каждого пользователя равна

$$C'_k = \frac{W}{K}\log_2\left[1 + \frac{P}{(W/K)N_0}\right], \qquad (15.2.2)$$

а суммарная пропускная способность для K пользователей равна

$$KC'_k = W \log_2\left(1 + \frac{KP}{WN_0}\right) \qquad (15.2.3)$$

Следовательно, суммарная пропускная способность эквивалентна пропускной способности одного пользователя со средней мощностью $P_{ср} = KP$.

Интересно отметить, что при фиксированной полосе W суммарная пропускная способность становится неограниченной, если число пользователей линейно возрастает с K. С другой стороны, когда K возрастает, каждый пользователь занимает меньшую полосу (W/K) и, как следствие, пропускная способность на пользователя уменьшается. Рис.15.2.1 иллюстрирует пропускную способность C'_K на пользователя, нормированную

полосой канала W, как функцию от $\mathscr{E}_b / N_0$ с параметром K. Это выражение определяется так[1]

$$\frac{C_k'}{W} = \frac{1}{K}\log_2\left[1 + K\frac{C_k'}{W}\left(\frac{\mathscr{E}_b}{N_0}\right)\right]. \qquad (15.2.4)$$

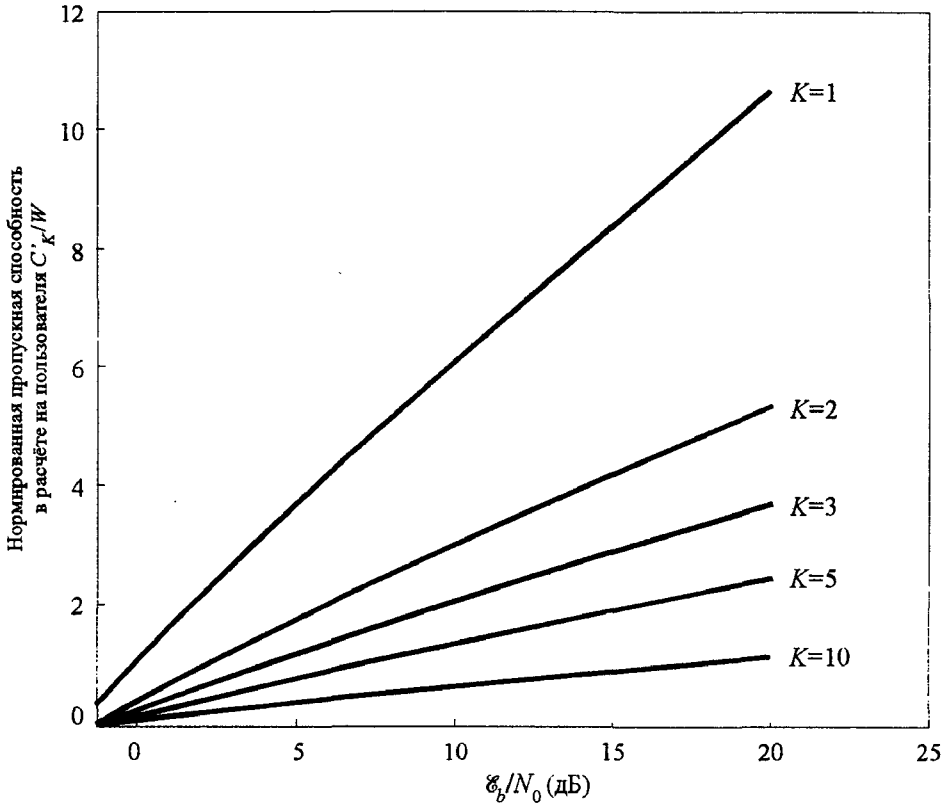


Рис. 15.2.1. Нормированная пропускная способность как функция от $\mathscr{E}_b/N_0$ для FDMA

Более компактная форма (15.2.4) получается путем определения нормированной суммарной пропускной способности $C_н = KC_k'/W$, которая определяет суммарную битовую скорость всех K пользователей на единицу полосы частот. Таким образом, (15.2.4) можно выразить так

$$C_н = \log_2\left(1 + C_н\frac{\mathscr{E}_b}{N_0}\right), \qquad (15.2.5)$$

или, эквивалентно

$$\frac{\mathscr{E}_b}{N_0} = \frac{2^{C_н} - 1}{C_н}. \qquad (15.2.6)$$

Зависимость $C_н$ от $\mathscr{E}_b / N_0$ показана на рис.15.2.2. Мы видим, что $C_н$ растет с ростом $\mathscr{E}_b / N_0$, если оно больше минимального значения ln 2.

[1] Предполагается, что $C_k'\mathscr{E}_b = P$ (прп).

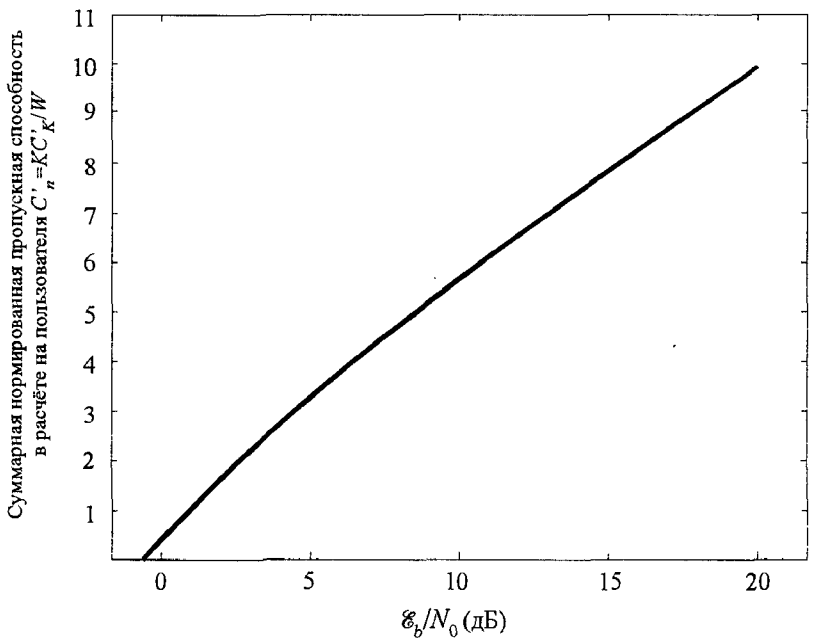


Рис. 15.2.2. Суммарная нормированная пропускная способность как функция от $\mathscr{E}_b/N_0$ для FDMA

В системе TDMA каждый пользователь передает по каналу в полосе W на интервале времени $1/K$ со средней мощностью KP. Следовательно, пропускная способность на пользователя равна

$$C_k = \frac{1}{K} W \log_2\left(1 + \frac{KP}{WN_0}\right), \tag{15.2.7}$$

что равно пропускной способности TDMA системы. Однако с практической точки зрения мы должны подчеркнуть, что в TDMA передатчики не всегда могут поддерживать мощность передачи KP, если K очень велико. Следовательно имеется практический предел, выше которого мощность передатчика нельзя увеличить с ростом K.

В CDMA системе каждый пользователь передает псевдослучайный сигнал с полосой W и средней мощностью P. Пропускная способность системы зависит от уровня сотрудничества между K пользователями. В экстремальном случае имеем CDMA без сотрудничества, когда приёмник для каждого сигнала пользователя не знает рассеянный сигнал других пользователей или выбирается с игнорированием этого знания в процессе демодуляции. Тогда сигналы других пользователей проявляются как интерференция на приёме у каждого пользователя. В этом случае приёмник многих пользователей состоит из банка K приёмников отдельных пользователей. Если предположим, что псевдослучайный сигнал каждого пользователя гауссовский, тогда сигнал каждого пользователя поражается гауссовской интерференцией мощностью $(K-1)P$ и аддитивным гауссовским шумом мощности WN_0. Следовательно, пропускная способность на пользователя

$$C'_k = W \log_2\left[1 + \frac{P}{WN_0 + (K-1)P}\right], \tag{15.2.8}$$

или, что эквивалентно

$$\frac{C'_k}{W} = \log_2\left[1 + \frac{C'_k}{W}\frac{\mathscr{E}_b / N_0}{1 + (K-1)\left(C_k / W\right)\mathscr{E}_b / N_0}\right] \tag{15.2.9}$$

Рис.15.2.3 иллюстрирует зависимость C'_k / W от $\mathscr{E}_b / N_0$ с параметром K.

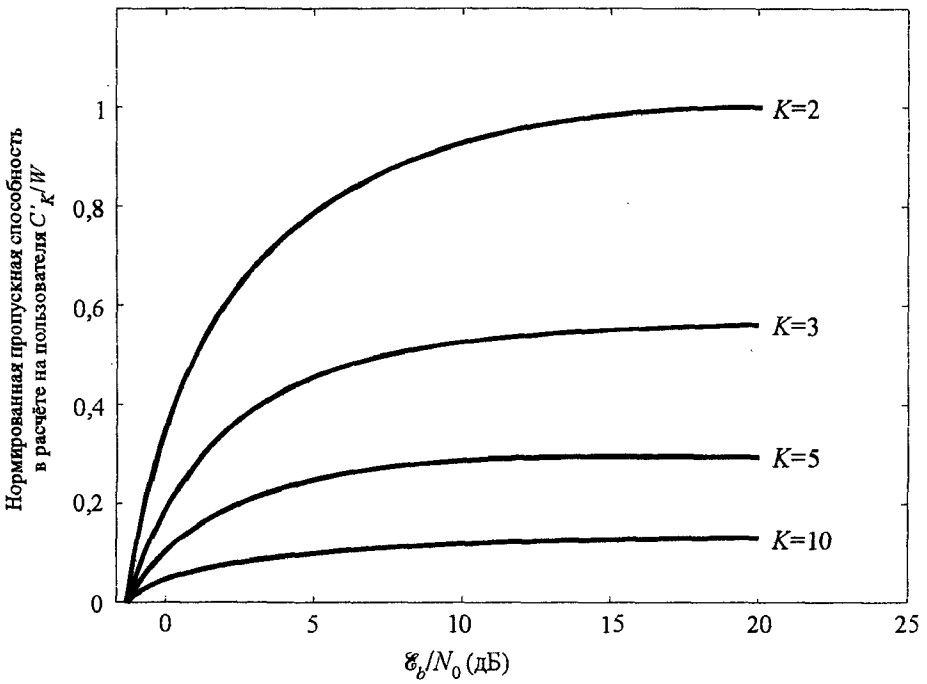


Рис. 15.2.3. Нормированная пропускная способность как функция от $\mathscr{E}_b/N_0$ для несогласованной CDMA

При большом числе пользователей мы можем использовать аппроксимацию $\ln(1+x) \le x$. Следовательно,

$$\frac{C'_k}{W} \le \frac{C'_k}{W} \frac{\mathscr{E}_b / N_0}{1+K\left(C'_k/W\right)\left(\mathscr{E}_b/N_0\right)} \log_2 \mathrm{e}, \tag{15.2.10}$$

или, что эквивалентно,

$$C_{\text{н}} \le \log_2 \mathrm{e} - \frac{1}{\mathscr{E}_b/N_0} = \frac{1}{\ln 2} - \frac{1}{\mathscr{E}_b/N_0} < \frac{1}{\ln 2}. \tag{15.2.11}$$

В этом случае мы видим, что суммарная пропускная способность не увеличивается с ростом K, как при FDMA и TDMA. С другой стороны, предположим, что K пользователей сотрудничают посредством синхронной передачи во времени и приёмник многих пользователей знает рассеяние сигналов всех пользователей и совместно демодулирует и детектирует все сигналы пользователей. Пусть каждый пользователь имеет скорость передачи R_i, $1 \le i \le K$, и кодовый словарь, содержащий набор из 2^{nR_i} кодовых слов мощностью P. На каждом сигнальном интервале каждый пользователь выбирает произвольное кодовое слово, скажем, $\mathbf{X}_i$ из своего собственного кодового словаря и все пользователи передают их кодовые слова одновременно. Таким образом, декодер на приёме наблюдает

$$\mathbf{Y} = \sum_{i=1}^{K} \mathbf{X}_i + \mathbf{Z}, \tag{15.2.12}$$

где $\mathbf{Z}$ – вектор аддитивного шума. Оптимальный декодер выносит решение по K кодовым словам, одно по каждому кодовому словарю, в пользу слов, которые образуют векторную сумму, которая наиболее близка по Евклиду к принимаемому вектору $\mathbf{Y}$.

Достигаемый K-мерный диапазон скоростей для K пользователей в канале с АБГШ, при условии равенства мощностей каждого пользователя, дается следующим уравнением:

$$R_i < W \log_2\left(1+\frac{P}{WN_0}\right), \quad 1 \le i \le K \tag{15.2.13}$$

$$R_i + R_j < W\log_2\left(1+\frac{2P}{WN_0}\right), \quad 1 \le i, j \le K \tag{15.2.14}$$

$$\vdots$$

$$\sum_{i=1}^{K} R_i < W\log_2\left(1+\frac{KP}{WN_0}\right) \tag{15.2.15}$$

Для частного случая, когда все скорости одинаковы, неравенство (15.2.15) доминирует относительно других $K-1$ неравенств. Отсюда следует, что если скорости $\{R_i,\ 1 \le i \le K\}$ для K сотрудничающих синхронных пользователей выбираются так, чтобы вместиться в область пропускной способности, определенную вышеприведенными неравенствами, тогда вероятность ошибки для K пользователей стремится к нулю, когда длина кодового блока n стремится к бесконечности.

Из приведенного обсуждения мы заключаем, что сумма скоростей K пользователей становится неограниченной с ростом K. Следовательно, при сотрудничающих синхронных пользователей пропускная способность CDMA имеет форму похожую на форму FDMA и TDMA. Заметим, что если все скорости пользователей CDMA системы выбраны одинаковыми и равными R, тогда (15.2.15) дает

$$R < \frac{W}{K}\log_2\left(1+\frac{KP}{WN_0}\right), \tag{15.2.16}$$

что идентично ограничению скорости для FDMA и TDMA. В этом случае CDMA не обеспечивает большую скорость, чем FDMA и TDMA. Однако, если скорости K пользователей выбираются неравными так, чтобы неравенства (15.2.13)–(15.2.15) выполнялись, тогда возможно найти такие точки в достижимой области скоростей, что сумма скоростей K пользователей CDMA превосходит пропускную способность FDMA и TDMA.

Пример 15.2.1. Рассмотрим случай двух пользователей в системе CDMA, которые используют кодированные сигналы, описанные выше. Скорости двух пользователей должны удовлетворять неравенствам

$$R_1 < W\log_2\left(1+\frac{P}{WN_0}\right),$$

$$R_2 < W\log_2\left(1+\frac{P}{WN_0}\right),$$

$$R_1 + R_2 < W\log_2\left(1+\frac{2P}{WN_0}\right),$$

где P – средняя переданная мощность каждого пользователя, а W – полоса частот сигнала. Определим область пропускной способности, для системы CDMA с двумя пользователями.

Область пропускной способности для CDMA с двумя пользователями с кодированными сигналами имеет форму, иллюстрированную на рис.15.2.4, где

$$C_i' = W\log_2\left(1+\frac{P_i}{WN_0}\right), \quad i = 1, 2$$

– это пропускные способности, соответствующие двум пользователям с $P_1 = P_2 = P$.

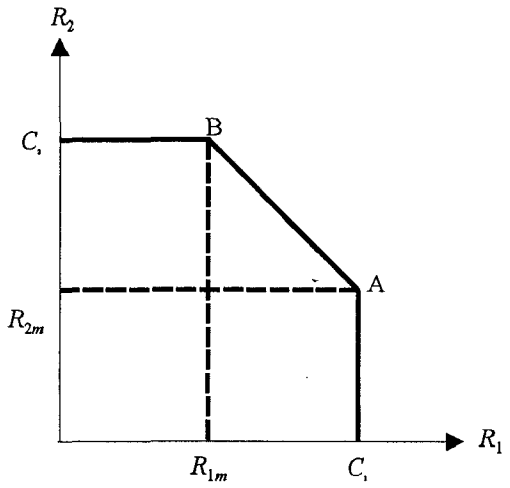


Рис. 15.2.4. Область пропускной способности гауссовского канала с CDMA с двумя пользователями

Заметим, что если пользователь 1 передает с пропускной способностью C_1, то пользователь 2 может передавать с максимальной скоростью

$$R_{2m} = W\log_2\left(1+\frac{2P}{WN_0}\right) - C_1' = W\log_2\left(1+\frac{P}{P+WN_0}\right), \qquad (15.2.17)$$

что иллюстрируется на рис.15.2.4 точкой А. Этот результат имеет интересную интерпретацию. Мы видим, что R_{2m} соответствует случаю, когда сигнал пользователя 1 рассматривается как эквивалентный аддитивный шум при детектировании сигнала пользователя 2. С другой стороны, пользователь 1 может передавать с пропускной способностью C_1', поскольку приёмник знает передаваемый сигнал пользователя 2 и, следовательно, он может ограничить его влияние при детектировании сигнала пользователя 1.

Вследствие симметрии аналогичная ситуация существует если пользователь 2 передает с пропускной способностью C_2'. Тогда пользователь 1 может передавать с максимальной скоростью $R_{1m} = R_{2m}$, что иллюстрируется на рис.15.2.4 точкой В. В этом случае мы имеем аналогичную интерпретацию, как выше с заменой ролей пользователей 1 и 2.

Точки А и В соединяются прямой линией. Легко видеть, что эта прямая линия является границей достижимой области скоростей, поскольку любая точка линии соответствует максимальной скорости $W\log_2\left(1+2P/WN_0\right)$, которую можно достичь простым делением во времени канала между двумя пользователями.

В следующем разделе мы рассмотрим проблему детектирования сигнала для систем CDMA со многими пользователями и оценим качество и вычислительную сложность нескольких структур приёмника.

15.3. МНОЖЕСТВЕННЫЙ ДОСТУП С КОДОВЫМ РАЗДЕЛЕНИЕМ

Как мы видели ТДМА и FDMA являются методами множественного доступа, при которых канал разделяется на независимые, используемые одним пользователем подканалы, т.е. неперекрывающиеся интервалы времени или частоты, соответственно. В CDMA каждому пользователю предназначается различная адресная последовательность (или сигнал), которую получатель использует для модуляции с рассеиванием информации по всему сигналу. Адресная последовательность также позволяет приёмнику

демодулировать сообщения, переданные многими пользователями канала, которые передают сигналы одновременно и в общем асинхронно.

В этом разделе мы рассмотрим демодуляцию и детектирование CDMA сигналов от многих пользователей. Мы увидим, что оптимальный детектор максимального правдоподобия имеет вычислительную сложность, которая растёт экспоненциально с числом пользователей. Такая высокая сложность служит мотивацией для разработки субоптимальных детекторов, имеющих более низкую вычислительную сложность. В заключении, мы рассмотрим характеристики качества различных детекторов.

15.3.1. Сигналы CDMA и модели канала

Рассмотрим CDMA канал, который делят K одновременных пользователей. Каждому пользователю предназначается адресный сигнал $g_k(t)$ длительностью T, где T – символьный интервал. Адресный сигнал можно выразить так

$$g_k(t) = \sum_{n=0}^{L-1} a_k(n) p\left(t - nT_c\right), \quad 0 \le t \le T, \tag{15.3.1}$$

где $\left\{a_k(n),\ 0 \le n \le L-1\right\}$ псевдошумовая (ПШ) кодовая последовательность, содержащая L чипов, которые принимают значение $\{\pm 1\}$, $p(t)$ – импульс длительности T_c, а T_c – интервал чипа. Таким образом, мы имеем L чипов на символ $T = LT_c$. Без потери общности мы предположим, что все K адресных сигнала имеют единичную энергию:

$$\int_0^T g_k^2(t)dt = 1. \tag{15.3.2}$$

Взаимная корреляция между парой адресных сигналов играет важную роль для метрик детектора сигнала и его качества. Мы определим следующие взаимные корреляции:

$$\rho_{ij}(\tau) = \int_0^T g_i(t) g_j(t-\tau)dt, \quad i \le j, \tag{15.3.3}$$

$$\rho_{ji}(\tau) = \int_0^T g_i(t) g_j(t+T-\tau)dt, \quad i \le j. \tag{15.3.4}$$

Для простоты предположим, что для передачи информации от каждого пользователя используются двоичные противоположные сигналы. Далее, пусть информационная последовательность от k-го пользователя обозначается $\left\{b_k(m)\right\}$, где величина каждого информационного символа может быть ± 1. Удобно рассмотреть передачу блока символов одинаковой произвольной длины, скажем N. Тогда блок данных от k-го пользователя

$$\mathbf{b}_k = \left[b_k(1)\ \ldots\ b_k(N)\right]^{\mathrm{T}}, \tag{15.3.5}$$

и соответствующий эквивалентный низкочастотный сигнал можно выразить так:

$$s_k(t) = \sqrt{\mathscr{E}_k} \sum_{i=1}^{N} b_k(i) g_k(t - iT), \tag{15.3.6}$$

где $\mathscr{E}_k$ – энергия сигнала на бит. Суммарный передаваемый сигнал от K пользователей можно записать

$$s(t) = \sum_{k=1}^{K} s_k\left(t - \tau_k\right) = \sum_{k=1}^{K} \sqrt{\mathscr{E}_k} \sum_{i=1}^{N} b_k(i) g_k\left(t - iT - \tau_k\right), \tag{15.3.7}$$

где $\left\{\tau_k\right\}$ – задержки передачи, удовлетворяющие условию $0 \le \tau_k < T$ для $1 \le k < K$. Без потери общности предположим, что $0 \le \tau_1 \le \tau_2 \le \ldots \le \tau_k < T$. Это модель переданного

сигнала от многих пользователей в асинхронном режиме. В специальном случае синхронной передачи, $\tau_k = 0$ для $1 \le k < K$. Величины τ, фигурирующие в выражениях взаимной корреляции, даваемых (15.3.3) и (15.3.4) также можно, без потери общности, ограничить областью $0 \le \tau < T$.

Считается, что передаваемый сигнал искажается АБГШ. Следовательно, принимаемый сигнал можно записать так.

$$r(t) = s(t) + n(t), \tag{15.3.8}$$

где $s(t)$ определяется (15.3.7), а $n(t)$ – АБГШ со спектральной плотностью мощности $\frac{1}{2}N_0$.

15.3.2. Оптимальный приёмник

Оптимальный приёмник определяется как приёмник, который выбирает наиболее правдоподобную последовательность символов $\{b_k(n), 1 \le n \le N, 1 \le k \le K\}$ по принимаемому сигналу $r(t)$, наблюдаемому на интервале $0 \le t \le NT + 2T$. Сначала мы рассмотрим случай синхронной передачи; позже мы рассмотрим асинхронную передачу.

Синхронная передача. При синхронной передаче каждый пользователь производит точно один символ на сигнальном интервале рассматриваемого пользователя. При белом гауссовском шуме в канале достаточно рассмотреть принимаемый сигнал на одном сигнальном интервале, $0 \le t \le T$ и определять оптимальный приёмник. Таким образом $r(t)$ можно выразить так

$$r(t) = \sum_{k=1}^{K} \sqrt{\mathscr{E}_k}\, b_k(1) g_k(t) + n(t), \quad 0 \le t \le T. \tag{15.3.9}$$

Оптимальный максимально правдоподобный приёмник вычисляет функцию правдоподобия

$$\Lambda(\mathbf{b}) = \int_0^T \left[r(t) - \sum_{k=1}^{K} \sqrt{\mathscr{E}_k}\, b_k(1) g_k(t) \right]^2 dt \tag{15.3.10}$$

и выбирает информационную последовательность $\{b_k(1), 1 \le k \le K\}$, которая минимизирует $\Lambda(\mathbf{b})$. Если мы раскроем интеграл в 15.3.10, то получим

$$\begin{aligned} \Lambda(\mathbf{b}) = \int_0^T r^2(t)dt - 2\sum_{k=1}^{K} \sqrt{\mathscr{E}_k}\, b_k(1) \int_0^T r(t) g_k(t) dt + \\ + \sum_{j=1}^{K}\sum_{k=1}^{K} \sqrt{\mathscr{E}_j \mathscr{E}_k}\, b_k(1) b_j(1) \int_0^T g_k(t) g_j(t) dt. \end{aligned} \tag{15.3.11}$$

Видим, что интеграл от $r^2(t)$ является общим для всех возможных последовательностей $\{b_k(1)\}$ и он не влияет в определении того, какая последовательность передана. Следовательно его можно упустить. Слагаемое

$$r_k = \int_0^T r(t) g_k(t) dt, \quad 1 \le k \le K \tag{15.3.12}$$

представляет собой взаимную корреляцию принимаемого сигнала с каждым из K адресными последовательностями. Вместо взаимных корреляторов мы можем использовать согласованные фильтры. Наконец интеграл, включающий $g_k(t)$ и $g_j(t)$ даёт

$$\rho_{jk}(0) = \int_0^T g_j(t) g_k(t) dt. \qquad (15.3.13)$$

Следовательно, (15.3.11) можно выразить в форме корреляционных метрик

$$C(\mathbf{r}_k, \mathbf{b}_k) = 2\sum_{k=1}^{K} \sqrt{\mathscr{E}_k} b_k(1) r_k - \sum_{j=1}^{K}\sum_{k=1}^{K} \sqrt{\mathscr{E}_j \mathscr{E}_k} b_k(1) b_j(1) \rho_{jk}(0). \qquad (15.3.14)$$

Эти корреляционные метрики можно также выразить в векторной форме

$$C(\mathbf{r}_k, \mathbf{b}_k) = 2\mathbf{b}_k^{\mathrm{T}} \mathbf{r}_k - \mathbf{b}_k^{\mathrm{T}} \mathbf{R}_s \mathbf{b}_k, \qquad (15.3.15)$$

где

$$\mathbf{r}_k = \left[r_1\ r_2\ \ldots\ r_K\right]^{\mathrm{T}},\ \mathbf{b}_k = \left[\sqrt{\mathscr{E}_1} b_1(1)\ \ldots\ \sqrt{\mathscr{E}_k} b_k(1)\right]^{\mathrm{T}},$$

и $\mathbf{R}_s$ – матрица корреляции с элементами $\rho_{jk}(0)$. Видно, что оптимальный детектор должен знать энергии принимаемых сигналов для расчёта корреляционных метрик.

Имеется 2^K возможных выборов символов в информационных последовательностях K пользователей. Оптимальный детектор вычисляет корреляционные метрики для каждой возможной последовательности и выбирает последовательность, которая имеет наибольшую корреляционную метрику. Видно, что оптимальный детектор имеет сложность, которая растёт экспоненциально с числом пользователей K.

В целом оптимальный приёмник при синхронной передаче символов состоит из банка из K корреляторов или согласованных фильтров, за ними детектор , который вычисляет 2^K корреляционных метрик, определяемых (15.3.15), соответствующих 2^K возможным передаваемым информационным последовательностям. Затем детектор выбирает последовательность, соответствующую наибольшей корреляционной метрике.

Асинхронная передача. В этом случае имеются точно два соседних символа от каждого пользователя, которые перекрывают символ на сигнальном интервале рассматриваемого пользователя. Мы предполагаем, что приёмник знает энергии принимаемых символов $\{\mathscr{E}_k\}$ для K пользователей и задержки передачи $\{\tau_k\}$. Ясно, что эти параметры должны быть измерены на приёме или переданы приёмнику как сторонняя информация от пользователей посредством некоторого сигнала управления.

Оптимальный максимально правдоподобный приёмник вычисляет функцию правдоподобия

$$\begin{aligned}\Lambda(\mathbf{b}) &= \int_0^{NT+2T} \left[r(t) - \sum_{k=1}^{K} \sqrt{\mathscr{E}_k} \sum_{i=1}^{N} b_k(i) g_k(t - iT - \tau_k) \right]^2 dt = \\ &= \int_0^{NT+2T} r^2(t) dt - 2\sum_{k=1}^{K} \sqrt{\mathscr{E}_k} \sum_{i=1}^{N} b_k(i) \int_0^{NT+2T} r(t) g_k(t - iT - \tau_k) dt + \\ &+ \sum_{k=1}^{K}\sum_{l=1}^{K} \sqrt{\mathscr{E}_k \mathscr{E}_l} \sum_{i=1}^{N}\sum_{j=1}^{N} b_k(i) b_l(j) \int_0^{NT+2T} g_k(t - iT - \tau_k) g_l(t - jT - \tau_l) dt,\end{aligned} \qquad (15.3.16)$$

где $\mathbf{b}$ представляет последовательность данных от K пользователей. Интеграл от $r^2(t)$ можно упустить поскольку он общий для всех возможных последовательностей. Интеграл

$$r_k(i) \equiv \int_{iT+\tau_k}^{(i+1)T+\tau_k} r(t) g_k(t - iT - \tau_k) dt, \quad 1 \le i \le N \qquad (15.3.17)$$

представляет выход коррелятора или согласованного фильтра для k-го пользователя на каждом сигнальном интервале. Наконец, интеграл

$$\int_0^{NT+2T} g_k(t-iT-\tau_k)g_l(t-jT-\tau_l)dt = \\ = \int_{-iT-\tau_k}^{NT+2T-iT-\tau_k} g_k(t)g_l(t+iT-jT+\tau_k-\tau_l)dt \qquad (15.3.18)$$

можно легко разложить на слагаемые, включающие взаимную корреляцию $\rho_{kl}(\tau)=\rho_{kl}(\tau_k-\tau_l)$, для $k \le l$ и $\rho_{lk}(\tau)$ для $k>l$. Следовательно, мы видим что функция правдоподобия можно выразить через слагаемые корреляционных метрик, которые включают выходы $\{r_k(i), 1\le k\le K, 1\le i\le N\}$ K корреляторов или согласованных фильтров – один для каждой из K адресных последовательностей. Используя векторные обозначения, можно показать, что выходы NK корреляторов или согласованных фильтров – $\{r_k(i)\}$ можно выразить в форме

$$\mathbf{r} = \mathbf{R}_N\mathbf{b}+\mathbf{n} \qquad (15.3.19)$$

где по определению

$$\mathbf{r} = \left[\mathbf{r}^{\mathrm{T}}(1)\ \mathbf{r}^{\mathrm{T}}(2)\ \ldots \mathbf{r}^{\mathrm{T}}(N)\right]^{\mathrm{T}}$$
$$\mathbf{r}(i) = \left[r_1(i)\ r_2(i)\ \ldots r_K(i)\right]^{\mathrm{T}} \qquad (15.3.20)$$

$$\mathbf{b} = \left[\mathbf{b}^{\mathrm{T}}(1)\ \mathbf{b}^{\mathrm{T}}(2)\ \ldots \mathbf{b}^{\mathrm{T}}(N)\right]^{\mathrm{T}}$$
$$\mathbf{b}(i) = \left[\sqrt{\mathscr{E}_1}b_1(i)\ \sqrt{\mathscr{E}_2}b_2(i)\ \ldots\ \sqrt{\mathscr{E}_K}b_K(i)\right]^{\mathrm{T}} \qquad (15.3.21)$$

$$\mathbf{n} = \left[\mathbf{n}^{\mathrm{T}}(1)\ \mathbf{n}^{\mathrm{T}}(2)\ \ldots \mathbf{n}^{\mathrm{T}}(N)\right]^{\mathrm{T}}$$
$$\mathbf{n}(i) = \left[n_1(i)\ n_2(i)\ \ldots n_K(i)\right]^{\mathrm{T}} \qquad (15.3.22)$$

$$\mathbf{R}_N = \begin{bmatrix} \mathbf{R}_a(0) & \mathbf{R}_a^{\mathrm{T}}(1) & 0 & \cdots & \cdots & 0 \\ \mathbf{R}_a(1) & \mathbf{R}_a(0) & \mathbf{R}_a^{\mathrm{T}}(1) & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \mathbf{R}_a(1) & \mathbf{R}_a(0) & \mathbf{R}_a^{\mathrm{T}}(1) \\ 0 & 0 & 0 & 0 & \mathbf{R}_a(1) & \mathbf{R}_a(0) \end{bmatrix}, \qquad (15.3.23)$$

а $\mathbf{R}_a(m)$ – матрица размером $K\times K$ с элементами

$$R_{kl}(m) = \int_{-\infty}^{\infty} g_k(t-\tau_k)g_l(t+mT-\tau_l)dt \qquad (15.3.24)$$

Векторы гауссовского шума $\mathbf{n}(i)$ имеют нулевые средние и автокорреляционную матрицу

$$E\left[\mathbf{n}(k)\mathbf{n}^{\mathrm{T}}(j)\right] = \tfrac{1}{2}N_0\mathbf{R}_a(k-j). \qquad (15.3.25)$$

Заметим, что вектор $\mathbf{r}$, определённый (15.3.19), образует достаточную статистику для оценивания передаваемых символов $b_k(i)$.

Если мы хотим проследить схему обработки то оптимальный МП детектор должен вычислить 2^{NK} корреляционных метрик и выбрать K последовательностей длины N, которые соответствуют наибольшей корреляционной метрике. Ясно, что такой подход слишком сложен в вычислительном отношении, чтобы быть реализуемом на практике, особенно, когда K и N велико. Альтернативный подход – МП оценивание

последовательностей, использующее алгоритм Витерби. Чтобы сконструировать детектор последовательного типа, мы должны использовать то обстоятельство, что каждый передаваемый символ перекрывается с $2K-2$ символами. Таким образом, получается существенное уменьшение вычислительной сложности с учётом параметра длины блока N, но экспоненциальная зависимость от K остаётся.

Важно, что оптимальный МП приёмник, использующий алгоритм Витерби, предполагает такую большую вычислительную сложность, что его использование на практике ограничено системами связи, в которых число пользователей крайне мало, например $K<10$. Для больших значений K следует рассматривать детектор последовательного типа, который схож или последовательному детектированию, или стек алгоритмам, описанным в гл. 8. Ниже мы рассмотрим некоторые субоптимальные детекторы, сложность которых растёт линейно с K.

15.3.3. Субоптимальные детекторы

В выше приведённом обсуждении мы видели, что оптимальный детектор для K пользователей CDMAимеет вычислительную сложность, измеряемую числом арифметических операций (сложений и умножений/делений) на модулированный символ, которые увеличиваются экспоненциально с K. В этом подразделе мы опишем субоптимальные детекторы с вычислительной сложностью, которая растёт линейно с числом пользователей K. Мы начнём с простейшего субоптимального детектора, который мы назовём общепринятым (для одного пользователя) детектором.

Общепринятый детектор для одного пользователя. В общепринятом детектировании сигнала одного пользователя приёмник для каждого пользователя состоит из демодулятора, который коррелирует (согласованно фильтрует) принимаемый сигнал с адресной последовательностью пользователя и подаёт выход коррелятора на детектор, который выносит решение, основываясь на выход единственного коррелятора. Таким образом, общепринятый детектор пренебрегает присутствием других пользователей канала или, что эквивалентно, предполагает, что аппаратурный шум вместе с интерференцией является белым и гауссовским.

Рассмотрим синхронную передачу. Тогда выход коррелятора для k-го пользователя на сигнальном интервале $0 \le t \le T$ равен

$$r_k = \int_0^T r(t) g_k(t) dt, \tag{15.3.26}$$

$$r_k = \sqrt{\mathscr{E}_k} b_k(1) + \sum_{\substack{j=1 \\ j \neq k}}^{K} \sqrt{\mathscr{E}_j} b_j(1) \rho_{jk}(0) + n_k(1), \tag{15.3.27}$$

где компонента шума $n_k(1)$ равна

$$n_k(1) = \int_0^T n(t) g_k(t) dt. \tag{15.3.28}$$

Поскольку $n(t)$ – белый гауссовский шум со спектральной плотностью мощности $\frac{1}{2} N_0$ дисперсия $n_k(1)$ равна

$$E\left[n_k^2(1)\right] = \tfrac{1}{2} N_0 \int_0^T g_k^2(t) dt = \tfrac{1}{2} N_0. \tag{15.3.29}$$

Ясно, если адресные последовательности ортогональны, интерференция от других пользователей, определённая средним слагаемым в (15.3.22), исчезает и общепринятый детектор одного пользователя оптимален. С другой стороны, если одна или больше адресных последовательностей не ортогональны к адресной последовательности данного

пользователя, интерференция от других пользователей может оказаться чрезмерной, если уровни мощности сигналов (или энергии принимаемых сигналов) от одного или больше других пользователей существенно больше, чем уровень мощности k-го пользователя. Эта ситуация обычно называется проблема разных дальностей (near-far problem) в системе связи со многими пользователями и делает необходимым некоторые виды контроля мощности при общепринятом детектировании.

При асинхронной передаче общепринятый детектор более уязвим к интерференции от других пользователей. Это потому, что невозможно синтезировать адресные последовательности для любой пары пользователей, которые ортогональны для всех сдвигов во времени. Следовательно, интерференция от других пользователей неустранима при асинхронной передаче с общепринятым детектированием одного пользователя. В этом случае проблема разных дальностей, возникшая из-за неравной мощности сигналов, переданных различными пользователями, особенно серьёзная. Её практическое решение обычно требует регулировки мощности, осуществляемой приёмником посредством отдельного канала связи, который все пользователи непрерывно перехватывают. Другой выбор сводится к использованию одного из многопользовательских детекторов, описываемых ниже.

Детектор с декорреляцией. Мы видели, что общепринятый детектор имеет сложность, которая растёт линейно с числом пользователей, но его уязвимость к проблеме разных дальностей требует некоторого вида контроля мощности. Мы хотим придумать другой тип детектора, который также имеет линейную вычислительную сложность, но не проявляет уязвимость к интерференции других пользователей.

Сначала рассмотрим случай синхронной передачи символов. В этом случае, вектор, принимаемого сигнала $\mathbf{r}_K$, который представляет собой выход K согласованных фильтров равен

$$\mathbf{r}_K = \mathbf{R}_s \mathbf{b}_K + \mathbf{n}_K, \tag{15.3.30}$$

где $\mathbf{b}_K = \left[\sqrt{\mathscr{E}_1}\, b_1(1)\ \sqrt{\mathscr{E}_2}\, b_2(1) \ldots \sqrt{\mathscr{E}_K}\, b_K(1)\right]^{\mathrm{T}}$, а вектор шума с элементами $\mathbf{n}_K = \left[n_1(1)\ n_2(1) \ldots n_K(1)\right]^{\mathrm{T}}$ имеет ковариацию

$$E\left(\mathbf{n}_K \mathbf{n}_K^{\mathrm{T}}\right) = \mathbf{R}_s. \tag{15.3.31}$$

Поскольку шум гауссовский, $\mathbf{r}_K$ описывается K-мерной гауссовской ФПВ со средним $\mathbf{R}_s\mathbf{b}_K$ и ковариацией $\mathbf{R}_s$. То есть

$$p\left(\mathbf{r}_K \middle| \mathbf{b}_K\right) = \frac{1}{\sqrt{(2\pi)^K \det \mathbf{R}_s}} \exp\left[-\frac{1}{2}\left(\mathbf{r}_k - \mathbf{R}_s\mathbf{b}_k\right)^{\mathrm{T}} \mathbf{R}_s^{-1}\left(\mathbf{r}_k - \mathbf{R}_s\mathbf{b}_k\right)\right]. \tag{15.3.32}$$

Наилучшая линейная оценка $\mathbf{b}_K$ это величина $\mathbf{b}_K$, которая минимизирует логарифм функции правдоподобия

$$\Lambda\left(\mathbf{b}_K\right) = \left(\mathbf{r}_k - \mathbf{R}_s\mathbf{b}_k\right)^{\mathrm{T}} \mathbf{R}_s^{-1}\left(\mathbf{r}_k - \mathbf{R}_s\mathbf{b}_k\right). \tag{15.3.33}$$

Результат такой минимизации даёт

$$\mathbf{b}_K^0 = \mathbf{R}_s^{-1}\mathbf{r}_k. \tag{15.3.34}$$

Тогда детектируемые символы (оценки) определяются знаком каждого элемента $\mathbf{b}_K^0$, то есть

$$\hat{\mathbf{b}}_K = \mathrm{sgn}\left(\mathbf{b}_K^0\right). \tag{15.3.35}$$

Рис.15.3.1 иллюстрирует структуру приёмника. Из (15.3.34) и (15.3.35) видно, что декоррелятор требует знания относительной задержки $\mathbf{R}_s$, но не требуется знания сигнальных амплитуд.

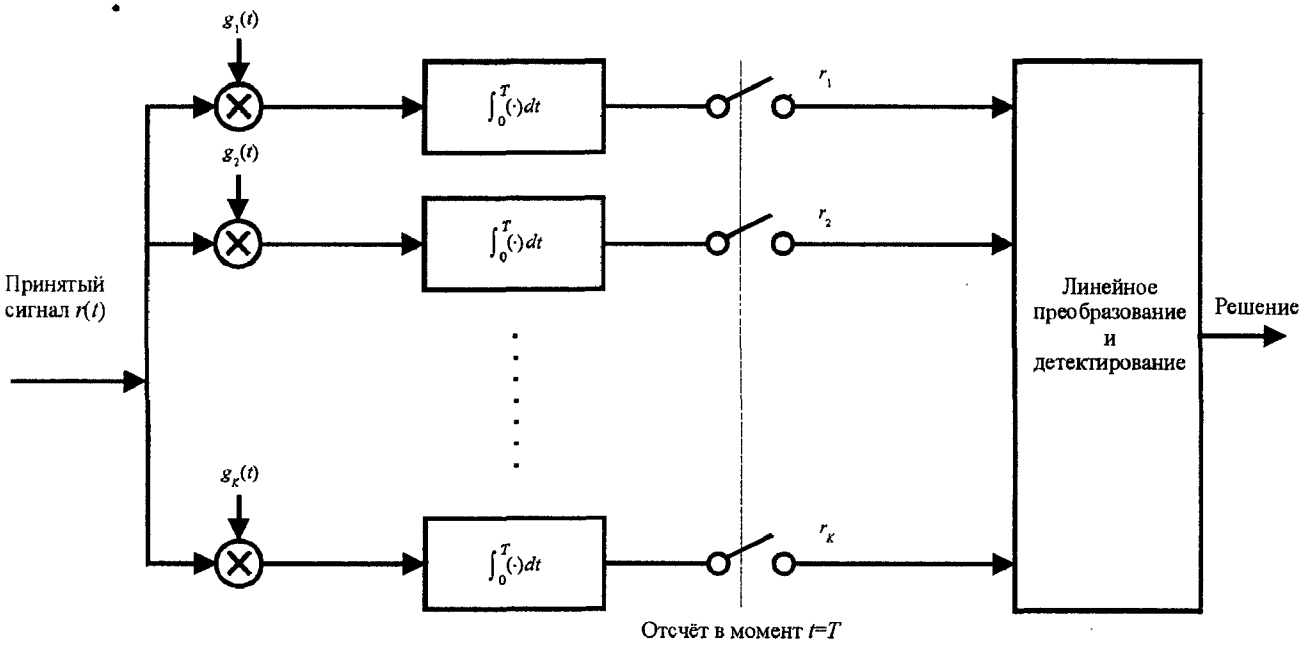


Рис. 15.3.1. Структура приёмника с декорреляцией

Поскольку оценка $\mathbf{b}_K^0$ получена формированием линейного преобразования вектора выходов корреляторов, вычислительная сложность линейно растёт[1] с K.

Читатель может видеть, что наилучшая (максимально правдоподобная) оценка $\mathbf{b}_K$, даваемая (15.3.34), отличается от оптимального нелинейного МП детектора последовательности, который находит наилучшую дискретную последовательность с величинами $\{\pm 1\}$, которая максимизирует функцию правдоподобия. Интересно также отметить, что оценка $\mathbf{b}_K^0$ – это наилучшая линейная оценка, которая максимизирует корреляционные метрики, определяемые (15.3.15).

Интересную интерпретацию детектора, который вычисляет $\mathbf{b}_K^0$ по (15.3.34) и выносит решения согласно (15.3.35), можно получить, рассмотрев случай $K=2$. В этом случае

$$\mathbf{R}_S = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \tag{15.3.36}$$

$$\mathbf{R}_S^{-1} = \frac{1}{1-\rho^2}\begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix}, \tag{15.3.37}$$

где

$$\rho = \int_0^T g_1(t)g_2(t)dt\,. \tag{15.3.38}$$

Затем, если мы коррелируем принимаемый сигнал

[1] Замена различения гипотез оцениванием непрерывных параметров в канале с МСИ и АБГШ исследована Д.Д. Кловским и С.М. Широковым [80]. При большой МСИ в канале такая замена ведёт к заметной потере качества (прп).

$$r(t) = \sqrt{\mathscr{E}_1}b_1 g_1(t) + \sqrt{\mathscr{E}_2}b_2 g_2(t) + n(t) \tag{15.3.39}$$

с $g_1(t)$ и $g_2(t)$, мы получим

$$\mathbf{r}_2 = \begin{bmatrix} \sqrt{\mathscr{E}_1}b_1 + \rho\sqrt{\mathscr{E}_2}b_2 + n_1 \\ \rho\sqrt{\mathscr{E}_1}b_1 + \sqrt{\mathscr{E}_2}b_2 + n_2 \end{bmatrix}, \tag{15.3.40}$$

где n_1 и n_2 являются шумовыми компонентами на выходе корреляторов. Следовательно

$$\mathbf{b}_2^0 = \mathbf{R}_s^{-1}\mathbf{r}_2 = \begin{bmatrix} \sqrt{\mathscr{E}_1}b_1 + \left(n_1 - \rho n_2\right)/\left(1-\rho^2\right) \\ \sqrt{\mathscr{E}_2}b_2 + \left(n_2 - \rho n_1\right)/\left(1-\rho^2\right) \end{bmatrix}. \tag{15.3.41}$$

Это очень интересный результат, т.к. преобразование $\mathbf{R}_S^{-1}$ ограничивает интерференционные компоненты между двумя последовательностями. Следовательно, проблема разных дальностей ограничивается и нет нужды в контроле мощности.

Интересно отметить, что результат (15.3.41) получается также, если мы коррелируем $r(t)$, определяемое (15.3.39), с двумя модифицированными адресными сигналами

$$g_1'(t) = g_1(t) - \rho g_2(t), \tag{15.3.42}$$

$$g_2'(t) = g_2(t) - \rho g_1(t). \tag{15.3.43}$$

Это означает, что путем коррелирования принимаемого сигнала с модифицированными адресными сигналами, мы можем отстроить или декоррелировать интерференцию многих пользователей. Поэтому детектор, основанный на (15.3.34), называют *детектором с декорреляцией.*

При асинхронной передаче принимаемый сигнал на выходе корреляторов определяется (15.3.19). Следовательно, логарифм функции правдоподобия определяется как

$$\Lambda(\mathbf{b}) = \left(\mathbf{r} - \mathbf{R}_N\mathbf{b}\right)^{\mathrm{T}}\mathbf{R}_N^{-1}\left(\mathbf{r} - \mathbf{R}_N\mathbf{b}\right), \tag{15.3.44}$$

где $\mathbf{R}_N$ определяется (15.3.23), а $\mathbf{b}$ определяется (15.3.21). Относительно легко показать, что вектор $\mathbf{b}$, который минимизирует $\Lambda(\mathbf{b})$, равен

$$\mathbf{b}^0 = \mathbf{R}_N^{-1}\mathbf{r} \tag{15.3.45}$$

Это МП оценка $\mathbf{b}$, и она опять получена путем формирования линейного преобразования выходов блока корреляторов или согласованных фильтров.

Поскольку $\mathbf{r} = \mathbf{R}_N\mathbf{b} + \mathbf{n}$, то следует из (15.3.45), что

$$\mathbf{b}^0 = \mathbf{b} + \mathbf{R}_N^{-1}\mathbf{n} \tag{15.3.46}$$

Следовательно, $\mathbf{b}^0$ – несмещенная оценка $\mathbf{b}$. Это значит, что интерференция многих пользователей ограничена, как в случае синхронной передачи символов. Поэтому детектор для асинхронной передачи также называется *детектор с декорреляцией.*

Эффективным в вычислительном отношении методом для получения решения (15.3.45) является метод факторизации матрицы системы, описанный в приложении D. Конечно, имеется много других методов, которые можно использовать для обращения матрицы $\mathbf{R}_N$. Были также исследованы итеративные методы для декорреляции сигналов.

Детектирование по минимуму среднего квадрата ошибки. В приведенном выше обсуждении мы показали, что линейная МП оценка $\mathbf{b}$ получается минимизацией квадратичной функции логарифма правдоподобия в (15.3.44). Так мы получили результат, определяемый (15.3.45). Он является оценкой, получаемая путем формирования линейного преобразования выходов банка корреляторов или согласованных фильтров.

Другое, в определенной степени иное, решение можно получить, если мы осуществим линейное преобразование $\mathbf{b}^0 = \mathbf{Ar}$, где матрица $\mathbf{A}$ определяется так, чтобы минимизировать средний квадрат ошибки (СКО)

$$J(\mathbf{b}) = E\left[\left(\mathbf{b}-\mathbf{b}^0\right)^{\mathrm{T}}\left(\mathbf{b}-\mathbf{b}^0\right)\right] = E\left[(\mathbf{b}-\mathbf{Ar})^{\mathrm{T}}(\mathbf{b}-\mathbf{Ar})\right] \qquad (15.3.47)$$

Легко показать, что оптимальный выбор $\mathbf{A}$, который минимизирует $J(\mathbf{b})$ равен

$$\mathbf{A}^0 = \left(\mathbf{R}_N + \tfrac{1}{2}N_0\mathbf{I}\right)^{-1} \qquad (15.3.48)$$

и, следовательно

$$\mathbf{b}^0 = \left(\mathbf{R}_N + \tfrac{1}{2}N_0\mathbf{I}\right)^{-1}\mathbf{r}. \qquad (15.3.49)$$

Затем выход детектора определяется как $\hat{\mathbf{b}} = \mathrm{sgn}\left(\mathbf{b}^0\right)$.

Оценка, определяемая (15.3.49), называется оценкой по минимуму СКО (МСКО) для $\mathbf{b}$. Заметим, что когда $\frac{1}{2}N_0$ мало по сравнению с диагональными элементами матрицы $\mathbf{R}_N$ решение МСКО приближается к МП решению, определяемому (15.3.45). С другой стороны, когда уровень шума большой по сравнению с уровнем сигнала в диагональных элементах матрицы $\mathbf{R}_N$ матрица $\mathbf{A}^0$ приближается к единичной матрице (со скаляром $\frac{1}{2}N_0$). В этом случае (низкого ОСШ) детектор принципиально игнорирует интерференцию от других пользователей, поскольку аддитивный шум является доминирующим слагаемым.

Можно также отметить, что критерий МСКО дает смещённую оценку $\mathbf{b}$. Следовательно, остаётся некоторая остаточная интерференция многих пользователей.

Чтобы выполнить вычисления, ведущие к $\mathbf{b}$, мы решаем систему линейных уравнений

$$\left(\mathbf{R}_N + \tfrac{1}{2}N_0\mathbf{I}\right)\mathbf{b} = \mathbf{r}. \qquad (15.3.50)$$

Это решение можно эффективно выполнить, используя факторизацию матрицы $\mathbf{R}_N + \frac{1}{2}N_0\mathbf{I}$, как указано выше. Таким образом, детектирование NK символов требует $3NK^2$ умножений. Следовательно, вычислительная сложность определяется как $3K$ умножений на символ (бит), причём она не зависит от длины блока N и линейно связана с K.

Другие типы детекторов. Детектор с декорреляцией и МСКО детектор, описанные выше, включают формирование линейных преобразований блока данных от K корреляторов или согласованных фильтров. МСКО детектор подобен линейному СКО эквалайзеру, описанному в главе 10. Следовательно, МСКО многопользовательское детектирование можно реализовать, используя линию задержки с отводами и с настраивающимися коэффициентами для каждого пользователя и выбирая коэффициенты фильтра для минимизации СКО для сигнала каждого пользователя. Таким образом, принимаемые информационные символы оцениваются последовательно с постоянной задержкой, вместо оценки блока символов.

Оценку $\mathbf{b}^0$, определенную (15.3.46), получаемую обработкой блока из N символов детектором с декорреляцией, можно также вычислить последовательно. Хай и др. (1990) продемонстрировали, что передаваемые символы могут быть восстановлены последовательно от принимаемого сигнала путем использования разновидности эквалайзера с обратной связью по решению и с постоянной задержкой. Таким образом, здесь имеется похожесть между детектированием сигналов, искаженных МСИ в системе с одним пользователем и детектированием сигналов в системе со многими пользователями при ас синхронной передаче.

15.3.4. Характеристики качества детекторов

Вероятность ошибки на бит является обычно желательной мерой качества в системах со многими пользователями. Для расчета влияния интерференции многих пользователей на качество детектора одного пользователя, мы можем использовать в качестве исходной вероятность ошибки на бит для приёмника одного пользователя в отсутствие других пользователей канала, равную

$$P_k(\gamma_k) = Q\left(\sqrt{2\gamma_k}\right), \tag{15.3.51}$$

где $\gamma_k = \mathscr{E}_k / N_0$, $\mathscr{E}_k$ – энергия сигнала на бит, а $\frac{1}{2}N_0$ – спектральная плотность мощности АБГШ.

Для случая оптимального детектора, как при синхронной, так и асинхронной передаче, вероятность ошибки предельно рассчитать трудно. В этом случае мы можем использовать (15.3.51), как нижнюю границу, а качество субоптимального детектора как верхнюю границу.

Рассмотрим сначала субоптимальный общепринятый детектор для одного пользователя. Для синхронной передачи выход коррелятора для k-го пользователя определяется (15.3.27). Следовательно вероятность ошибки для k-го пользователя, при условии наличия последовательности символов $\mathbf{b}_i$ от других пользователей равна

$$P_k(\mathbf{b}_i) = Q\left(\sqrt{2\left[\sqrt{\mathscr{E}_k} + \sum_{j=1, j\neq k}^{K} \sqrt{\mathscr{E}_j} b_j(1)\rho_{jk}(0)\right]^2 \Big/ N_0}\right). \tag{15.3.52}$$

Затем средняя вероятность ошибки

$$P_k = \left(\tfrac{1}{2}\right)^{K-1} \sum_{\substack{i=1 \\ i\neq k}}^{K} P_k(\mathbf{b}_i). \tag{15.3.53}$$

Вероятность (15.3.53) в основном определяется слагаемым, которое имеет наименьший аргумент Q-функции. Наименьший аргумент получается при ОСШ

$$(\text{ОСШ})_{\min} = \frac{1}{N_0}\left[\sqrt{\mathscr{E}_k} - \sum_{j=1, j\neq k}^{K} \sqrt{\mathscr{E}_j}\left|\rho_{jk}(0)\right|\right]^2. \tag{15.3.54}$$

Следовательно,

$$\left(\tfrac{1}{2}\right)^{K-1} Q\left(\sqrt{2(\text{ОСШ})_{\min}}\right) < P_k < \left(\tfrac{1}{2}\right)^{K-1} (K-1) Q\left(\sqrt{2(\text{ОСШ})_{\min}}\right). \tag{15.3.55}$$

Аналогичный подход можно использовать для получения границ вероятности ошибки при асинхронной передаче. В случае детектора с декорреляцией, интерференция от других пользователей в целом ограничена. Следовательно, вероятность ошибки можно выразить так

$$P_k = Q\left(\mathscr{E}_k / \sigma_k^2\right), \tag{15.3.56}$$

где σ_k^2 – дисперсия шума для k-го элемента оценки $\mathbf{b}^0$.

Пример 15.3.1. Рассмотрим случай синхронной передачи двух пользователей, где $\mathbf{b}_2^0$ определяется (15.3.41). Определим вероятность ошибки.

Сигнальная компонента для первого слагаемого (15.3.4) равна $\sqrt{\mathscr{E}_1}$. Компонента шума равна $n = \dfrac{n_1 - \rho n_2}{1-\rho^2}$, где ρ – коэффициент корреляции между двумя адресными сигналами. Дисперсия этого шума равна

$$\sigma_1^2 = \frac{E\left[\left(n_1 - \rho n_2\right)\right]^2}{\left(1-\rho^2\right)^2} = \frac{1}{1-\rho^2}\frac{N_0}{2} \tag{15.3.57}$$

и

$$P_1 = Q\left(\sqrt{\frac{2\mathscr{E}_1}{N_0}\left(1-\rho^2\right)}\right). \tag{15.3.58}$$

Аналогичный результат можно получить для качества второго пользователя. Таким образом дисперсия шума увеличивается на множитель $\left(1-\rho^2\right)^{-1}$. Это повышение шума является платой за ограничение интерференции многих пользователей детектором с декорреляцией.

Вероятность ошибки для МСКО детектора равна (или похожа) на ту, которая определяет детектор с декорреляцией, когда уровень шума низок. Для примера, из (15.3.49) мы видим, что когда N_0 мала относительно диагональных элементов матрицы корреляции сигналов $\mathbf{R}_N$,

$$\mathbf{b}^0 \approx \mathbf{R}_N^{-1}\mathbf{r}, \tag{15.3.59}$$

что является решением детектора с декорреляцией. Для малой интерференции других пользователей МСКО детектор приводит к малому увеличению шума по сравнению с детектором с декорреляцией, но имеет некоторое остаточное смещение, обусловленное другими пользователями. Так МСКО детектор стремится достичь баланс между остаточной интерференцией и увеличением шума.

Альтернативой вероятности ошибки, как меры качества, используемой для характеристики системы со многими пользователями, является ОСШ при наличии и отсутствии интерференции. В частности, (15.3.51) дает вероятность ошибки для *k*-го пользователя при отсутствии интерференции других пользователей. В этом случае ОСШ равно $\gamma_k = \mathscr{E}_k / N_0$. При наличии интерференции других пользователей пользователь, который передает сигнал с энергией $\mathscr{E}_k$, будет иметь вероятность ошибки, превосходящей $P_k\left(\gamma_k\right)$. Эффективное ОСШ γ_{ke} определяются как ОСШ, требуемые для заданной вероятности ошибки

$$P_k = P_k\left(\gamma_{ke}\right) = Q\left(\sqrt{2\gamma_{ke}}\right). \tag{15.3.60}$$

Эффективность определяется отношением γ_{ke}/γ_k и представляет потерю качества, обусловленную интерференцией других пользователей. Желательным мерилом совершенства является асимптотическая эффективность, определяемая как

$$\eta_k = \lim_{N_0 \to 0} \frac{\gamma_{ke}}{\gamma_k}. \tag{15.3.61}$$

Эту величину часто легче сосчитать, чем вероятность ошибки.

Пример 15.3.2. Рассмотрим случай синхронной передачи символов от двух пользователей с энергией сигналов $\mathscr{E}_1$ и $\mathscr{E}_2$. Определим асимптотическую эффективность общепринятого детектора.

В этом случае вероятность ошибки легко найти из (15.3.52) и (15.3.53) как

$$P_1 = \tfrac{1}{2}Q\left(\sqrt{2\left(\sqrt{\mathscr{E}_1} + \rho\sqrt{\mathscr{E}_2}\right)^2 / N_0}\right) + \tfrac{1}{2}Q\left(\sqrt{2\left(\sqrt{\mathscr{E}_1} - \rho\sqrt{\mathscr{E}_2}\right)^2 / N_0}\right).$$

Однако асимптотическую эффективность можно рассчитать значительно легче. Она следует из определения (15.3.61) и (15.3.52):

$$\eta_1 = \left[\max\left(0,\ 1-\sqrt{\frac{\mathscr{E}_2}{\mathscr{E}_1}}|\rho|\right)\right]^2.$$

Аналогичное выражение можно получить для η_2.

Асимптотическую эффективность для оптимального и субоптимального детекторов, которые мы описали, были рассчитаны Верду (1986), Лупасом и Верду (1989) и Хаем и др. (1990). Рис.15.3.2 иллюстрирует асимптотическую эффективность этих детекторов, когда $K = 2$ пользователя передают информацию синхронно.

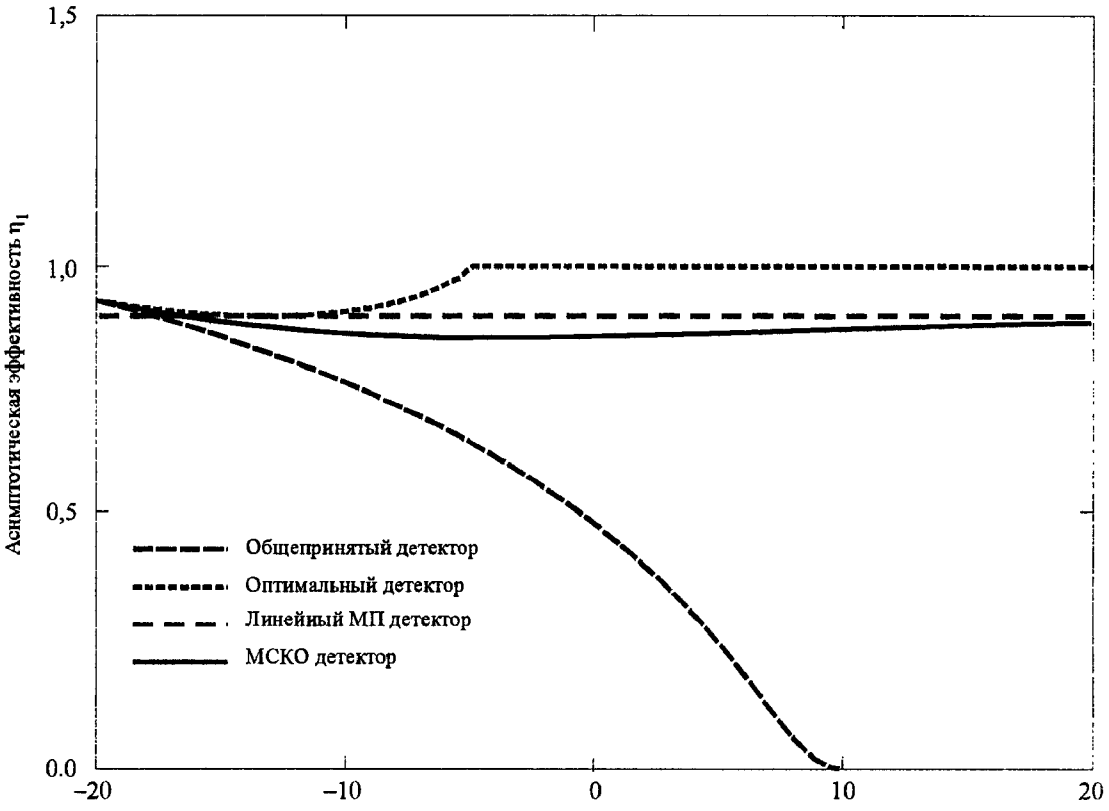


Рис. 15.3.2. Асимптотическая эффективность оптимального (Витерби) детектора, обычного детектора, МСКО детектора и линейного МП детектора в синхронной системе DS/SSMA с 2 пользователями [*Xie* и др. (1990), © *IEEE*]

Эти графики показывают, что когда интерференция мала $(\mathscr{E}_2 \to 0)$, асимптотическая эффективность этих детекторов относительно велика (близка к единице) и сопоставима. Когда $\mathscr{E}_2$ возрастает, асимптотическая эффективность общепринятого детектора быстро ухудшается. Однако остальные линейные детекторы работают относительно хорошо по сравнению с оптимальным детектором. Схожие заключения достигнуты при вычислении вероятности ошибки, но эти вычисления часто более утомительны.

15.4. Методы случайного доступа

В этом разделе мы рассмотрим систему связи со многими пользователями, в которой пользователи передают информацию пакетами по общему каналу. В противоположность CDMA методу, описанному в разделе 15.3, информационный сигнал пользователей теперь не рассеивается по частоте, т.е. является узкополосным. Как следствие, при одновременной передаче сигналы от многих пользователей нельзя разделить в приёмнике. Методы доступа, описанные ниже, являются принципиально случайными, поскольку

пакеты генерируются согласно некоторой статистической модели. Пользователи занимают канал, когда у них имеется для передачи один или более пакетов. Когда больше чем один пользователь пытаются передать пакеты одновременно, пакеты перекрываются во времени, т.е. сталкиваются и, следовательно, возникает конфликт, который должен быть разрешён путем использования некоторого канального протокола для повторной передачи пакетов. Ниже мы опишем несколько канальных протоколов для случайного доступа, которые разрешают конфликт при пакетной передаче.

15.4.1. АЛОХА. Системы и протоколы

Предположим, что используется схема случайного доступа, когда каждый пользователь передает пакет по мере его генерации. Когда пакет передан пользователем и никакой другой пользователь не передает пакет на этом интервале времени, тогда пакет считается успешно переданным. Однако, если один или более других пользователей передают пакет, который перекрывается во времени с пакетом первого пользователя, возникает столкновение и передача неуспешна. Рис.15.4.1 иллюстрирует этот сценарий. Если пользователи знают, когда их пакеты переданы успешно и когда они сталкиваются с другими пакетами, возможно разработать схему, которую мы назовем *протокол доступа в канал*, для ретрансляции столкнувшихся пакетов.

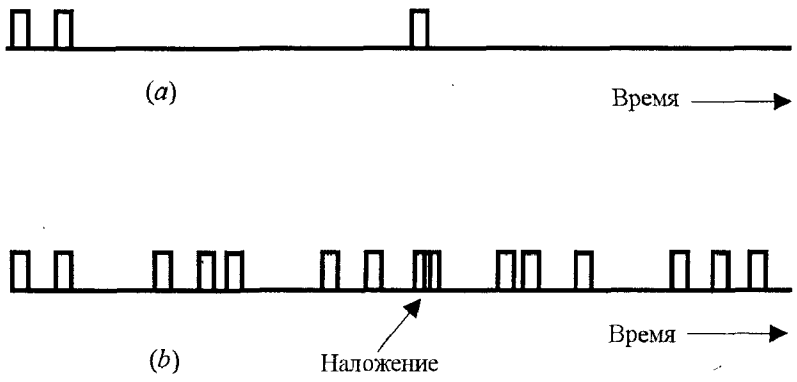


Рис. 15.4.1. Пакетная передача со случайным доступом:
(*a*) пакеты от типичного пользователя
(*b*) пакеты от нескольких пользователей

Обратная связь к пользователям об успешной или неуспешной передаче пакетов необходима и может быть обеспечена различными путями. В радиовещательной системе, такой как скажем в спутниковой ретрансляции, как изображено на рис.15.4.2, пакеты – это сигналы вещания от многих станций (передатчиков) ко всем пользователям по линии вниз. Все передатчики могут отслеживать их передачи, и, таким образом, получить следующую информацию: ни один пакет не был передан, или пакет был передан успешно, или возникло столкновение. Этот тип обратной связи к передатчикам обычно обозначается как $(0, 1, c)$ обратная связь.

В системах, использующих проводные или волоконно-оптические каналы приёмник может послать сигнал обратной связи по отдельному каналу.

В системе ALOHA (АЛОХА), изобретенной Абрамсоном (1973, 1977) и др. в университете на Гавайях, использовался спутниковый повторитель, который передавал пакеты от различных пользователей, которые имели доступ к спутнику. В этом случае все пользователи могли отслеживать передачи спутника и, таким образом, установить, переданы ли их пакеты успешно или нет.

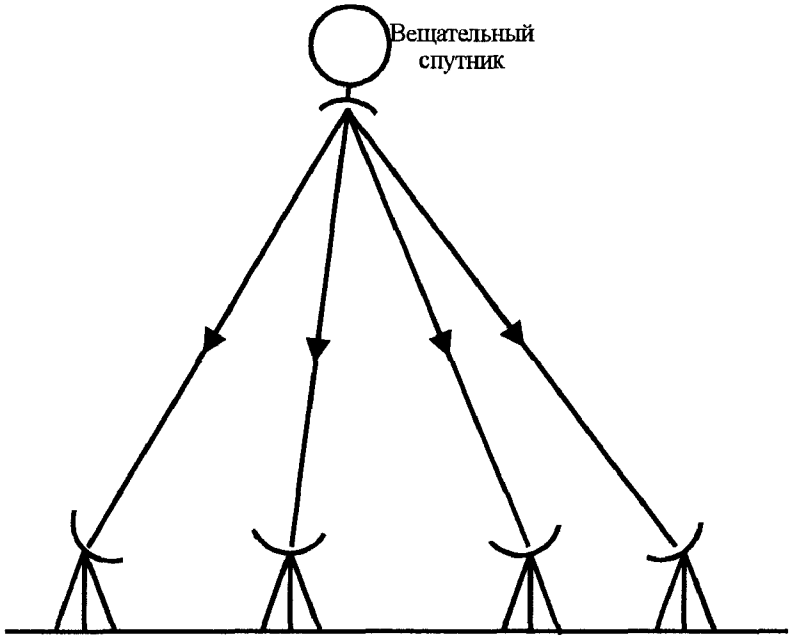


Рис. 15.4.2. Система спутникового вещания

В своей основе имеются два типа систем Алоха: синхронизированная (щелевая) и не синхронизированная (бесщелевая). В бесщелевой системе Алоха («чистая» Алоха) пользователь может начинать передачу пакета в любое произвольное время. В щелевой системе пакеты передаются во временные щели, которые начинаются и кончаются в определенное время.

Мы предположим, что время начала переданного пакета является точечным пуассоновским процессом, имеющим среднюю скорость (интенсивность) λ пакетов/с. Пусть T_p – длительность пакета. Тогда нормированный канальный трафик G определяется так:

$$G = \lambda T_p. \qquad (15.4.1)$$

Имеется много протоколов доступа в канал, которых можно использовать для разрешения конфликтов Рассмотрим один, принадлежащий Абрамсону (1971). В его протоколе пакеты, которые сталкиваются, ретранслируются с некоторой задержкой τ, которая выбирается случайно с ФПВ

$$p(\tau) = \alpha e^{-\alpha\tau}, \qquad (15.4.2)$$

где α – расчётный параметр. Случайная задержка τ прибавляется ко времени первоначальной передачи и пакет ретранслируется на новом интервале времени. Если снова возникает столкновение, новая величина τ случайно выбирается и пакет ретранслируется с новой задержкой относительно времени второй передачи. Этот процесс продолжается до тех пор, пока пакет не передастся успешно. Расчетный параметр α определяет среднюю задержку между ретрансляциями. Чем меньше α, тем длиннее задержка между ретрансляциями.

Теперь пусть λ', где $\lambda' < \lambda$ является скоростью, с которой пакеты передаются успешно. Тогда нормированная канальная проходимость равна

$$S = \lambda' T_p. \qquad (15.4.3)$$

Мы можем связать канальную проходимость S с предложенным канальным трафиком G путем использования предположенного распределения времени старта. Вероятность того, что какой-либо пакет не будет перекрывать данный пакет, равна вероятности того,

что ни один пакет не появится раньше точки старта на время меньшее T_p и ни один пакет не появится позже точки старта на время, меньшее T_p. Поскольку точка старта для всех пакетов имеет распределение Пуассона, вероятность того, что пакет не будет перекрываться, равна $\exp(-2\lambda T_p) = \exp(-2G)$. Следовательно,

$$S = Ge^{-2G}. \qquad (15.4.4)$$

Эта зависимость показана на рис.15.4.3. Видим, что максимальная проходимость равна $S_{max} = 1/2\,e = 0{,}184$ пакетов на щель, которая возникает при $G = \frac{1}{2}$. Если $G > \frac{1}{2}$, проходимость S уменьшается. Вышеприведенное исследование показывает, что несинхронизированный или бесщелевой метод доступа имеет относительно малую проходимость и не эффективен.

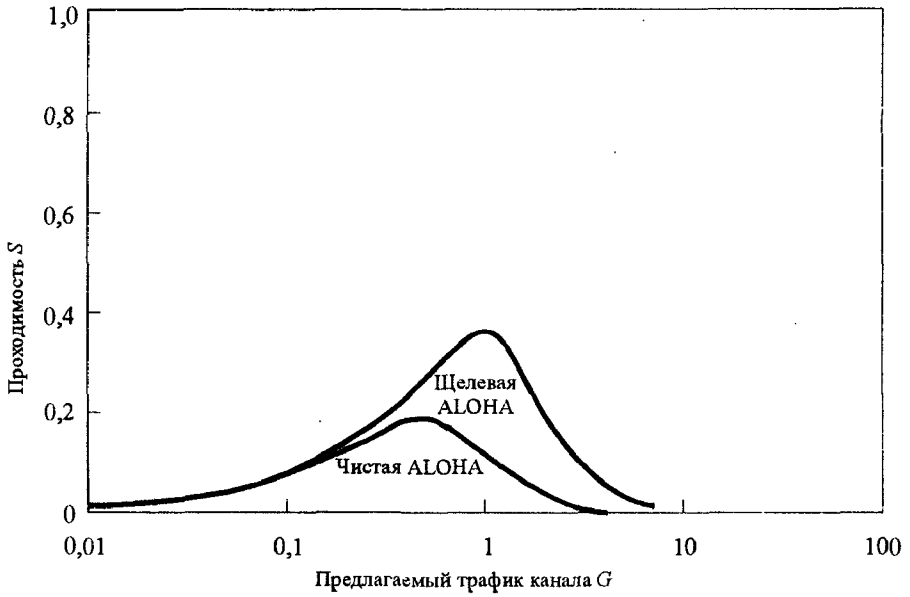


Рис. 15.4.3. Проходимость в системе ALOHA

Щелевая АЛОХА. Чтобы определить проходимость в щелевой системе АЛОХА, положим G_i – вероятность того, что i-й пользователь будет передавать пакет в некоторой щели. Если все K пользователей работают независимо и нет статистической зависимости между передачей пакетов пользователя в текущей щели и передачей пакета пользователя в предыдущей по времени щели, общий (нормализованный) предоставляемый каноном трафик равен

$$G = \sum_{i=1}^{K} G_i. \qquad (15.4.5)$$

Заметим, что в этом случае G может быть больше единицы.

Теперь, пусть $S_i \le G_i$ является вероятностью того, что пакет, переданный во временной щели, принимается без столкновения. Тогда нормированная проходимость канала равна

$$S = \sum_{i=1}^{K} S_i. \qquad (15.4.6)$$

Вероятность того, что пакет от i-го пользователя не будет иметь столкновения с другим пакетом, равна

$$Q_i = \prod_{\substack{j=1 \\ j \neq i}}^{K} \left(1 - G_j\right). \qquad (15.4.7)$$

Следовательно,

$$S_i = G_i Q_i. \qquad (15.4.8)$$

Простое выражение для канальной проходимости получается при рассмотрении K идентичных пользователей. Тогда

$$S_i = \frac{S}{K}, \quad G_i = \frac{G}{K}.$$

Далее, если предположим $K \to \infty$, мы получим проходимость

$$S = G\mathrm{e}^{-G}. \qquad (15.4.10)$$

Этот результат также изображен на рис.15.4.3. Видим, что S достигает максимум проходимости $S_{\max} = 1/\mathrm{e} = 0{,}368$ пакетов на щель при $G = 1$, что в два раза больше проходимости бесщелевой системы АЛОХА.

Качество щелевой системы АЛОХА, определенное выше, основывается на протоколе Абрамсона для конфликтных ситуаций. Большая проходимость возможна при разработке лучшего протокола.

Базовая слабость протокола Абрамсона заключается в том, что он не берет во внимание информацию о величине трафика канала, который можно наблюдать при возникающих столкновениях. Улучшение проходимости щелевой системы АЛОХА можно получить, используя древовидный протокол, разработанный Капетанакисом (1979). В этом алгоритме, пользователям не разрешается передавать новые пакеты, которые они генерируют, до тех пор, пока все предыдущие столкновения не будут разрешены. Пользователь может передавать новый пакет во временной щели немедленно за его генерацией, при условии, что все предыдущие пакеты, которые сталкивались были переданы успешно. Если создан новый пакет в то время, когда канал проясняет предыдущие столкновения, пакет сохраняется в буфере. Когда новый пакет сталкивается с другим, каждый пользователь относит свой соответствующий пакет к одному из двух ансамблей, скажем А и В, с равной вероятностью (бросанием монеты). Затем, если пакет помещен в ансамбль А, пользователь передает его в следующей временной щели. Если он столкнется снова, пользователь снова случайно относит пакет к одному из двух ансамблей и процесс передачи повторяется. Этот процесс продолжается до тех пор, пока все пакеты, содержащиеся в ансамбле А, не будут переданы успешно. Затем передаются все пакеты ансамбля В, следуя такой же процедуре. Все пользователи отслеживают состояние канала и, следовательно, они знают, когда все столкновения разрешены.

Когда канал оказывается в состоянии передавать новые пакеты, наиболее ранние созданные пакеты передаются первыми. Чтобы установить очередь, шкала времени разделяется на достаточно короткие подынтервалы, так что на подынтервале пользователями генерируется не более чем один пакет. Таким образом, каждый пакет имеет «временную этикетку», которая связана с подынтервалом, в котором он создан. Затем новый пакет, относящийся к первому подынтервалу, передается в первой возможной временной щели. Если нет столкновений, то передается пакет из второго подынтервала и так далее. Эта процедура продолжается, пока генерируются новые пакеты и так долго, пока существуют невыполненные заказы для передачи пакетов. Капетанакис показал, что этот протокол доступа в канал достигает максимальную проходимость из 0,43 пакета на щель.

В дополнение к проходимости, другая важная мера качества в системах со случайным доступом – это среднее время задержки при передаче пакета. В системе АЛОХА среднее число передач на пакет равно G/S. К этому числу мы можем прибавить среднее время

ожидания между передачами и таким образом получить среднюю задержку для успешной передачи.

Мы напомним из предыдущего обсуждения, что в протоколе Абрамсона параметр α определяет среднюю задержку между ретрансляциями. Если мы выбираем α малым, мы получаем ожидаемый эффект сглаживания (успокоения) канальной нагрузки во время пиковых нагрузок, но результатом является большее время ретрансляции. Это «профессиональный недостаток» при выборе α в (15.4.2). С другой стороны, можно показать, что протокол Капетанакиса имеет малую среднюю задержку в передаче пакетов. Следовательно, он превосходит протокол Абрамсона как по средней задержке, так и по проходимости.

Другим важным исследованием для проектирования протоколов в системе случайного доступа – это стабильность протокола. В нашей трактовке протоколов доступа в канал системы типа АЛОХА, мы безоговорочно предположили, что при заданной предоставляемой загрузке достигается точка равновесия, когда среднее число пакетов, поступающих в канал, равно среднему числу успешно переданных пакетов. Действительно, можно показать, что такой протокол доступа в канал, как протокол Абрамсона, который не берёт во внимание число предыдущих неуспешных передач при установлении режима ретрансляций, является по существу нестабильным. С другой стороны, алгоритм Капетанакиса отличается от протокола Абрамсона в этом отношении и может обеспечить стабильность. Полное обсуждение исследований стабильности протоколов случайного доступа можно найти в статье Месси (1988).

15.4.2. Системы, которые обнаруживают несущую, и протоколы

Как мы видели, протоколы случайного доступа для систем типа АЛОХА (щелевых и бесщелевых) обеспечивает относительно низкую проходимость. Далее, щелевая система АЛОХА требует, чтобы пользователи передавали сигналы в синхронизированных щелях времени. В каналах, в которых задержки передачи относительно малы, возможно синтезировать протоколы случайного доступа, которые дадут более высокую проходимость. Пример такого протокола – протокол обнаружения несущей с детектированием столкновений, который используется как стандарт эфирного протокола в локальных сетях в свободном пространстве. Этот протокол известен как протокол обнаружения несущей при случайном доступе с декодированием столкновений (CSMA/CD – carrier sense multiple access with collusion detection).

Протокол (CSMA/CD) очень прост. Все пользователи прислушиваются к передачам по каналу. Пользователь, который желает передавать пакет, захватывает канал, когда он обнаруживает, что канал не занят. Столкновения могут возникнуть, когда два или больше пользователей обнаруживает, что канал не занят и начинает передачу. Когда пользователи, которые передают одновременно, обнаруживают столкновение, они посылают специальный сигнал, называемый *сигналом столкновения* (jam), который служит для указания всем пользователям о столкновении, и прекращают свои передачи. И обнаружение несущей, и прекращения передач, когда возникает столкновение, ведут к минимизации времени занятия канала и, следовательно, дают большую проходимость.

Для детальной разработки эффективности CSMA/CD рассмотрим локальную сеть в свободном пространстве, имеющую трассовую (bus) структуру, показанную на рис.15.4.4.

Рассмотрим двух пользователей U_1 и U_2 с максимальным разделением, т.е. на двух концах трассы, и пусть τ_d определяет время задержки сигнала при его распространении по длине трассы. Тогда, время (максимальное), требуемое для обнаружения незанятого канала равно τ_d. Предположим, что U_1 передает пакет длительностью T_p.

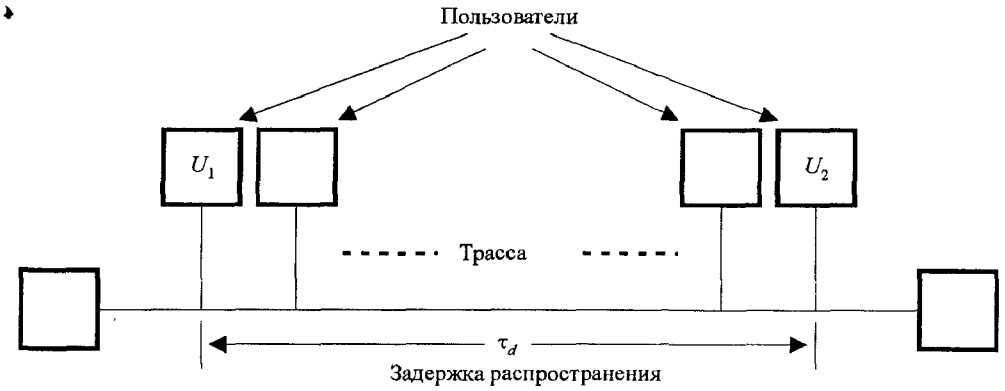


Рис. 15.4.4. Локальная сеть в свободном пространстве с трассовой архитектурой

Пользователь U_2 может захватить канал на τ_d позже, используя обнаружение несущей, и начинать передачу. Однако пользователь U_1 не знает об этой передаче до момента τ_d после начала передачи U_2. Таким образом, мы можем определить интервал времени $2\tau_d$ как временный интервал (максимальный) для обнаружения столкновения. Если мы предположим, что время, требуемое для передачи сигнала столкновения, пренебрежимо мало, то протокол CSMA/CD даст высокую проходимость, когда $2\tau_d << T_p$.

Имеется несколько возможных протоколов, которые можно использовать для повторения передач при возникновении столкновений. Один протокол называется ненастойчивый (nonpersistent) CSMA, другой называется 1-настойчивый CSMA, а обобщение последнего называется p-настойчивый CSMA.

Ненастойчивый CSMA. В этом протоколе пользователь, который имеет пакет для передачи, обнаруживает (отслеживает) канал и действует согласно следующему правилу:

a) Если канал не занят, пользователь передает пакет,

b) Если канал обнаружен занятым, пользователь переносит передачу пакета на более позднее время, согласно некоторому распределению задержек. На конце интервала задержки пользователь снова отслеживает канал и повторяет шаги a) и b).

1-настойчивый CSMA. Этот протокол проектируется для достижения высокой проходимости посредством того, что не позволяет каналу быть незанятым, если несколько пользователей имеют пакеты к передаче. Здесь пользователи отслеживают канал и работают согласно следующему правилу:

a) Если обнаружен незанятый канал, пользователь передает пакет с вероятностью 1;

b) Если канал обнаружен занятым, пользователь ждет, пока канал окажется незанятым и передает пакет с некоторой вероятностью.

Заметим, что в этом протоколе столкновение будет возникать, когда больше чем один пользователь имеют пакеты для передачи.

***p*-настойчивый CSMA.** Чтобы сократить скорость столкновений в 1-настойчивом CSMA и увеличить проходимость, мы можем рандомизировать время начала передачи пакетов. В частности, отследив незанятость канала, пользователь, имеющий пакет для передачи, посылает его с вероятностью p или задерживает его на время τ с вероятностью $1-p$. Вероятность p выбирается таким путем, чтобы сократить вероятность столкновений, в то время когда период незанятости между последовательными (неперекрывающимися) передачами сохраняется малым. Это выполняется путем деления оси времени на мини-щели длительностью τ и выбора начала передачи пакета в начале

мини-щели. В целом в p-настойчивом протоколе пользователь, имеющий пакет для передачи, поступает следующим образом:

a) Если канал обнаружен незанятым, пакет посылается с вероятностью p или с вероятностью $(1-p)$ передача задерживается на τ секунд.

b) Если при $t=\tau$ канал еще обнаруживается незанятым, шаг a) повторяется. Если возникает столкновение, пользователи переносят ретрансляцию пакетов согласно некоторого выбранного до передачи распределения задержек.

c) Если при $t=\tau$ канал обнаружен занятым, пользователи ждут, пока он окажется незанятым и затем поступают согласно a) и b).

Можно также конструировать щелевые версии описанного выше протокола.

Анализ проходимости для протоколов ненастойчивой и p-настойчивой CSMA/CD был выполнено Клейнроком и Тобаджи (1975), основываясь на следующих предположениях:

1. Среднее время ретрансляции велико по сравнению с длительностью пакета T_p.

2. Интервалы пребывания точечного процесса, определённые временем старта всех пакетов плюс времени ретрансляций, независимы и распределены экспоненциально.

Для ненастойчивой CSMA проходимость равна

$$S=\frac{Ge^{aG}}{G(1+2a)+e^{-aG}}, \tag{15.4.11}$$

где параметр $a=\tau_d/T_p$. Заметим, что когда $a\to 0$, $S\to G/(1+G)$. Рис.15.4.5 иллюстрирует зависимость прохождения от предоставляемого трафика G с параметром a. Видим, что $S\to 1$, когда $G\to\infty$ для $a=0$. Для $a>0$ величина $S_{\max}$ уменьшается.

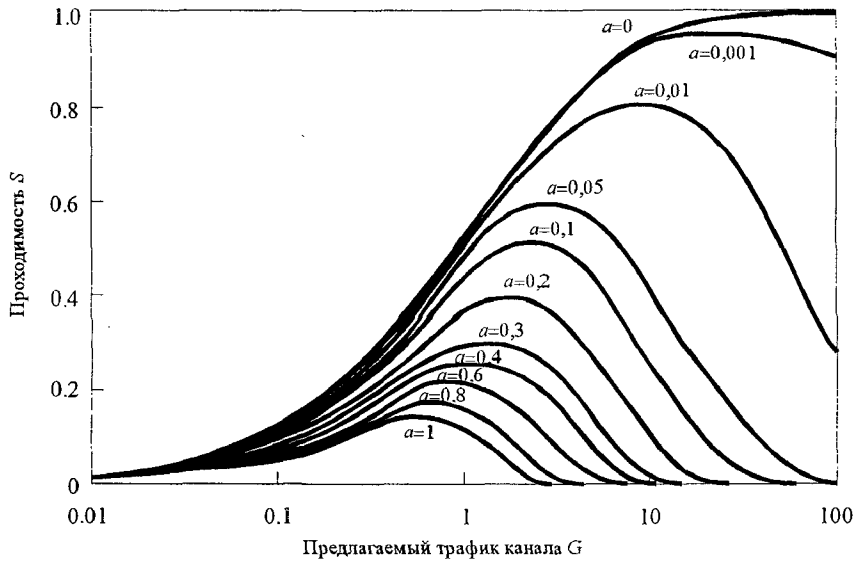


Рис. 15.4.5. Проходимость в ненастойчивой системе CSMA [*Kleinrock* и *Tobagi* (1975), © *IEEE*]

Для 1-настойчивого протокола проходимость, полученная Клейнроком и Тобаджи (1975), равна

$$S=\frac{G\left[1-G+aG\left(1+G+\frac{1}{2}aG\right)\right]e^{-G(1+2a)}}{G(1+2a)-\left(1-e^{-aG}\right)+(1+aG)e^{-G(1+a)}}. \tag{15.4.12}$$

В этом случае

$$\lim_{a\to 0}S=\frac{G(1+G)e^{-G}}{G+e^{-G}}, \tag{15.4.13}$$

что дает меньшую величину пика, чем при ненастойчивом протоколе.

При использовании p-настойчивого протокола возможно увеличить проходимость относительно 1-настойчивой схемы. Для примера, рис.15.4.6 иллюстрирует проходимость в зависимости от предоставляемого трафика с фиксированным значением $a = \tau_d / T_p$ и с параметром p. Видим, что по мере стремления p к единице максимальная проходимость уменьшается.

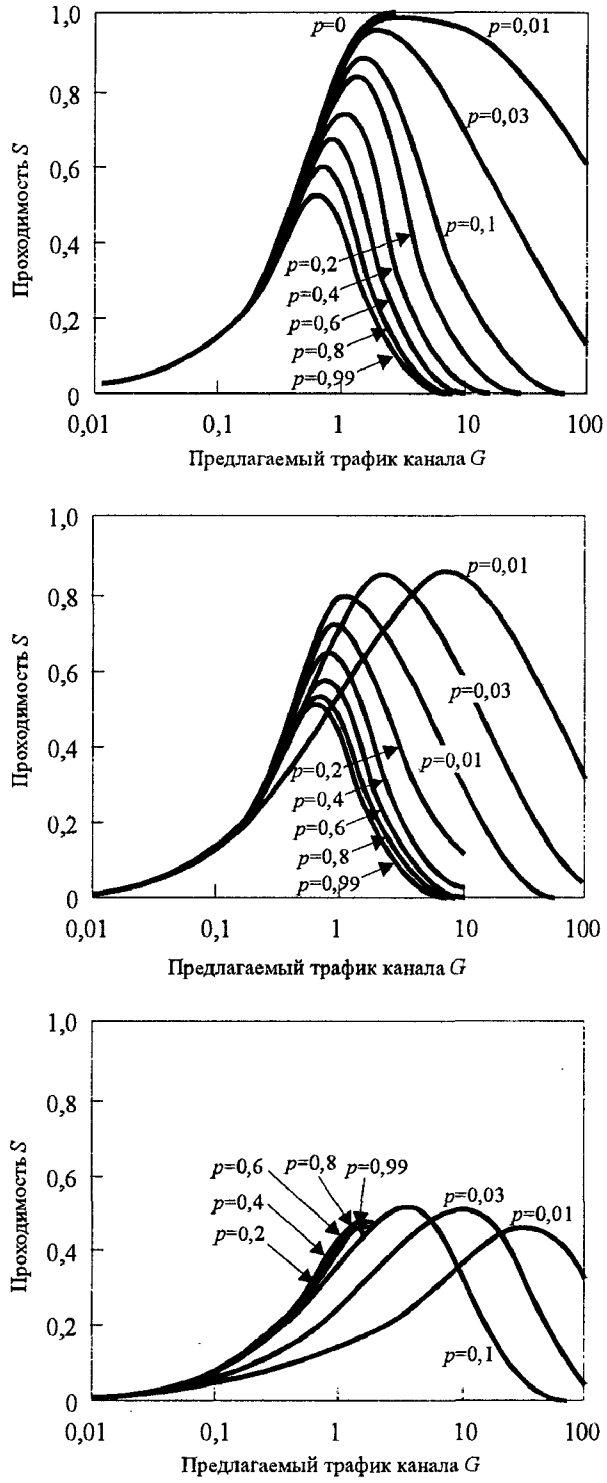


Рис. 15.2.6. Проходимость канала в p-настойчивой системе CSMA:
(*a*) a=0; (*b*) a=0,01; (*c*) a=0,1; [*Kleinrock* и *Tobagi* (1975), © *IEEE*]

Время передачи было также рассчитано Клейнроком и Тобаджи (1975). Рис.15.4.7 иллюстрирует графики задержки (нормированны к T_p) в зависимости от проходимости S для протоколов щелевой ненастойчивой и p-настойчивой CSMA.

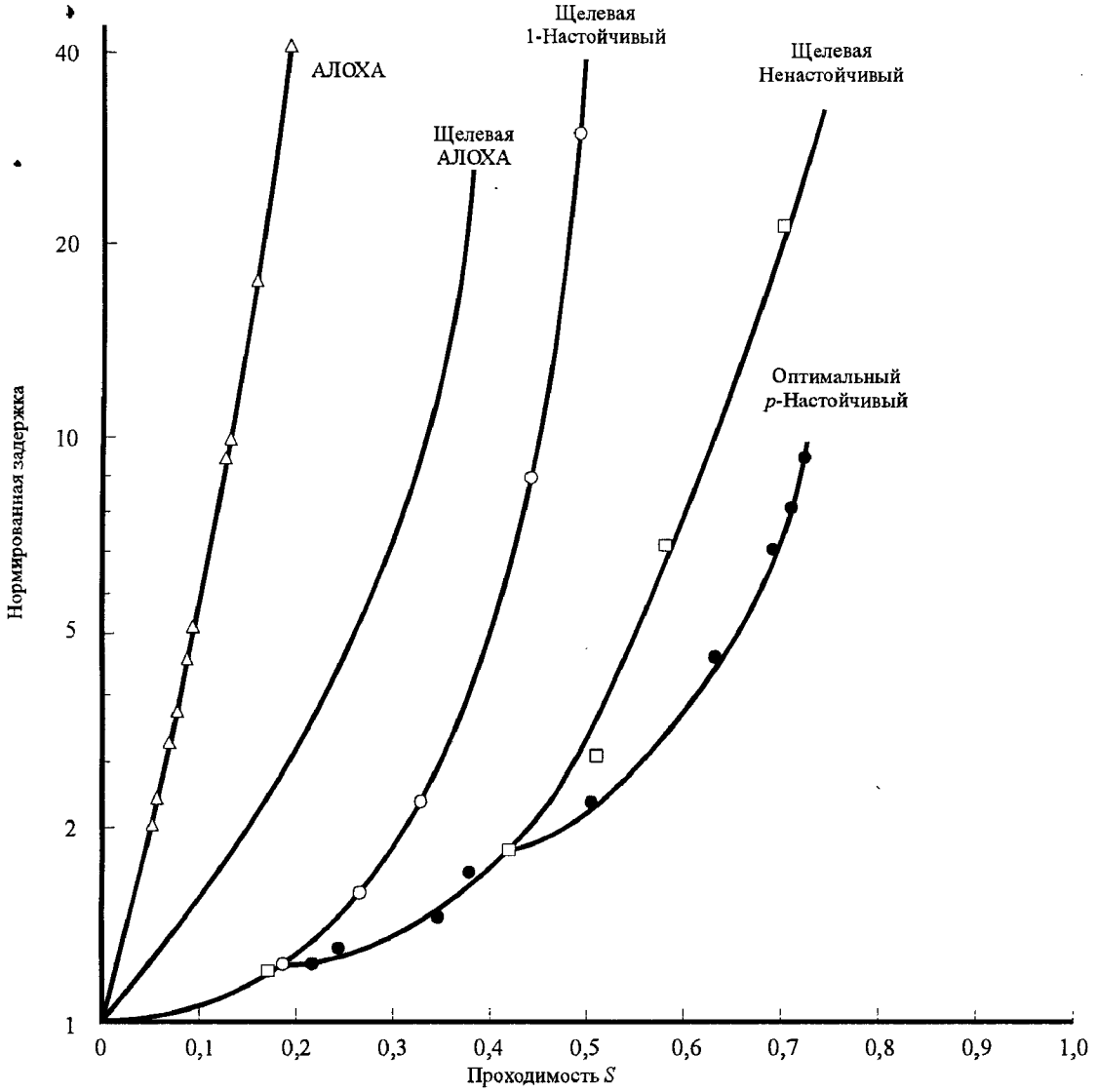


Рис. 15.4.7. Проходимость в обмен на задержку (моделирование при a=0,01) [*Kleinrock* и *Tobagi* (1975), © *IEEE*]

Для сравнения также показана зависимость задержки от характеристик прохождения для протоколов щелевой и бесщелевой Алохи. При этом моделировании только вновь генерированные пакеты считались независимыми и с пуассоновским распределением. Столкновения и неравномерно распределенные случайные ретрансляции обходятся без последующих предположений. Результаты этого моделирования иллюстрируют превосходство качества p-настойчивых и ненастойчивых протоколов относительно протоколов Алохи. Заметим, что значение, обозначенное на графике как "оптимальное p-настойчивое" получено путем нахождения оптимального значения p для каждого значения проходимости. Мы видим, что для малых значений проходимости 1-настойчивой $(p = 1)$ протокол является оптимальным.

15.5. БИБЛИОГРАФИЧЕСКИЕ ЗАМЕЧАНИЯ И ССЫЛКИ

FDMA была основной схемой для множественного доступа[1], которая долгое время использовалась в телефонных систем связи для аналоговой передачи голоса (?). С наступлением цифровой передачи речи, использующей ИКМ, ДИКМ и другие методы кодирования речи, TDMA заменила FDMA, как основная схема множественного доступа в связи. CDMA и методы случайного доступа в целом были разработаны за последние три десятилетия, сначала для беспроводной передачи сигналов и в канальных сетях в свободном пространстве.

Теория информации систем со многими пользователями имеет дело с базовыми информационно-теоретическими ограничениями при кодировании источников для многих пользователей и канального кодирования и модуляции для каналов с множественным доступом. Существует большое количество литературы по этим темам. В контексте нашей трактовки методов множественного доступа читатель может найти особо относящийся к делу материал в статьях Ковера (1972). ЭльГамаля и Ковера (1980), Бергманса и Ковера (1974) и Хая (1984). способность сотовых систем CDMA была рассмотрена в статье Гильхоузена и др. (1991).

Демодуляция и детектирование сигналов в системах со многими пользователями привлекают особое внимание в последние годы. Читатель отсылается к статьям Верду (1986 a-c, 1989), Лупаса и Верду (1990), Хая и др. (1990 a,b), Пур и Верду (1988), Занга и Браду (1993) Звонара и Браду (1995). Ранние работы по синтезу сигналов и демодуляции для систем связи со многими пользователями можно найти в статьях Ван Эттена (1975, 1976), Хорвуда и Гаглиарди (1975), и Кайе и Джордже (1970).

Система ALOHA, которая была одна из самых ранних систем со случайным доступом трактуется в статьях Абрамсона (1970, 1977) и Робертса (1975). Эти статьи содержат анализ проходимости для щелевых и не щелевых систем. Исследования стабильности касательно протоколов ALOHA можно найти в статьях Карлеиала и Хеллмана (1975), Жеза и др. (1988) и Месси (1988). Устойчивые протоколы, основанные на трёх алгоритмах для каналов со случайным доступом были впервые даны Капетанакисом (1977). Протоколы множественного доступа с обнаружением несущей, которые мы описали, принадлежат Клейнроку и Тобаджи (1975). Наконец, мы упомянем IEEE Press book, под ред. Абрамсона (1993), который содержит набор статей, имеющих дело с системами связи со множественным доступом.

ЗАДАЧИ

15.1. При определении сигналов CDMA и моделей типа, описанных в разделе 15.3.1, мы предположили, что принимаемые сигналы вещественны. Для $K>1$ это предположение предполагает фазовый синхронизм на всех передатчиках, что не очень реально в практических системах. Чтобы представить случай когда фазы несущих не синхронизированы мы должны просто изменить адресные сигналы для K пользователей, определенные (15.1.1), на комплексные в виде

$$g_k(t)=e^{\theta_k}\sum_{n=0}^{N-1}a_k(n)p(t-nT_c),\ 1\le k\le K,$$

где θ_k представляет постоянный фазовый сдвиг на k-м передатчике, как это видится общим приёмником.

a По данной комплексной форме адресных сигналов, определите форму оптимального МП приёмника, который вычисляет корреляционные метрики, аналогично (15.3.15).

b повторите нахождение оптимального МП детектора для асинхронной передачи, являющейся аналогом (15.3.19).

---

[1] Основополагающая работа [81] по теории линейного разделения сигналов многих пользователей выполнена Д.В. Агеевым в 1935 г. (прп).

15.2. Рассмотрите TDMA систему, когда каждый пользователь ограничен переданной мощностью P, независимо от числа пользователей. Определите пропускную способность на пользователя C'_K и общую пропускную способность KC'_K. Изобразите C'_K и KC'_K, как функцию от $\mathcal{E}_b/N_0$ и прокомментируйте результат при $K \to \infty$.

15.3. Рассмотрим систему FDMA при $K = 2$ пользователей в канале с АБГШ, где пользователю 1 предназначается полоса $W_1 = aW$, а пользователю 2 предназначается полоса $W_2 = (1-a)W$, где $0 \le a \le 1$. Пусть P_1 и P_2 – средние мощности двух пользователей.

a Определите пропускные способности C'_1 и C'_2 двух пользователей и их сумму $C' = C'_1 + C'_2$, как функцию от α. На двумерном графике зависимости скорости R_2 от R_1 изобразите график точек (C'_2, C'_1) при изменении α в пределах $0 \le a \le 1$.

b Напомним, что скорости двух пользователей должны удовлетворять условиям

$$R_1 < W_1 \log_2\left(1 + \frac{P_1}{W_1 N_0}\right)$$

$$R_2 < W_2 \log_2\left(1 + \frac{P_2}{W_2 N_0}\right)$$

$$R_1 + R_2 < W \log_2\left(1 + \frac{P_1 + P_2}{W_1 N_0}\right)$$

Определите суммарную пропускную способность C', когда $P_1/\alpha = P_2/(1-\alpha) = P_1 + P_2$ и затем покажите, что максимальная скорость достигается, когда $a/(1-\alpha) = P_1/P_2 = W_1/W_2$.

15.4. Рассмотрите систему TDMA с $K = 2$ пользователями в канале с АБГШ. Предположите, что два передатчика ограничены по пиковой мощности P_1 и P_2 и пусть пользователь 1 передает $100\alpha\%$ возможного времени, а пользователь 2 передает $100(1-\alpha)\%$ времени. Имеющаяся в распоряжении полоса частот равна W.

a Определите пропускные способности C'_1 и C'_2 и $C' = C'_1 + C'_2$ как функции от α.

b Нарисуйте графики точек (C'_2, C'_1) при изменении α в области $0 \le a \le 1$.

15.5. Рассмотрите систему TDMA с $K = 2$ пользователями в канале с АБГШ. Предположите, что два передатчика ограничены по средней мощности P_1 и P_2 и пусть пользователь 1 передает $100\alpha\%$ возможного времени, а пользователь 2 передает $100(1-\alpha)\%$ времени. Имеющаяся в распоряжении полоса частот равна W.

a Определите пропускные способности C'_1 и C'_2 и $C' = C'_1 + C'_2$ как функции от α.

b Нарисуйте графики точек (C'_2, C'_1) при изменении α в области $0 \le a \le 1$.

c В чём схожесть между этим решением и решением для системы FDMA из задачи 15.3.

15.6. Рассмотрите синхронный канал с множественным доступом при наличии двух пользователей и адресных последовательностей, показанных на рис. P15.6.

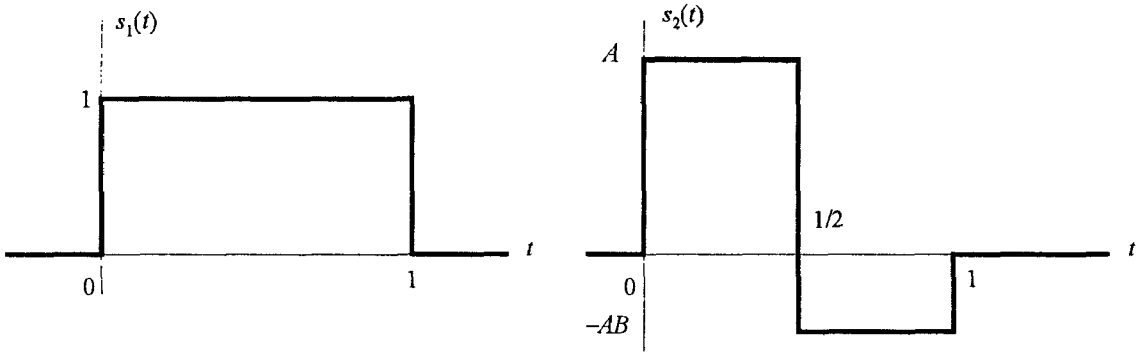


Рис. P15.6

Параметр $A \ge 0$ описывает относительную мощность двух пользователей, а $0 \le B \le 1$ описывает степень корреляции между сигналами. Пусть

$$r(t)=\sum_{k=1}^{2}\sum_{i=-\infty}^{\infty}b_k(i)S_k(t-i)+n(t)$$

означает принимаемый сигнал в момент t, где $n(t)$ – белый гауссовский шум со спектральной плотностью σ^2, а $b_k(i)\in\{-1,+1\}$. В следующих задачах вы должны сравнить структуру общепринятого детектора многих пользователей со структурой оптимального приёмника для различных величин A, $0\le B\le 1$ и σ^2.

a покажите, что при данном наблюдении $\{r(t),-\infty<t\le 1\}$ достаточная статистика для данных $b_1(0)$ и $b_2(0)$ является наблюдение на интервале $t\in[0,1]$.

b Общепринятый (субоптимальный) детектор многих пользователей выбирает данные $b_k(0)$ согласно следующему правилу

$$b_k(0)=\operatorname{sgn}(y_k),$$

где

$$y_k=\int_0^1 r(t)s_k(t)dt.$$

Определите выражение для вероятности ошибки на бит для пользователя 1, использую обозначения

$$w_k=\int_0^1 s_k^2(t)dt,$$

$$\rho_{12}=\int_0^1 s_1(t)s_2(t)dt.$$

c Какова форма этого выражения для $A\to 0$, $B<1$ и произвольном σ^2?

d Какова форма этого выражения для произвольно большого A, $B<1$ и произвольных σ^2? О чем это говорит для общепринятого детектировании?

e Какова форма этого выражения для $B=1$ и произвольных σ^2 и A? Почему она отличается от результата (d)?

f Определите форму этого выражения для произвольно больших σ^2, произвольных A и $B<1$.

g Определите форму этого выражения для $\sigma^2\to 0$, произвольного A и $B<1$.

15.7. Обратитесь к задаче 15.6. При максимально правдоподобном приёме последовательности в этом канале выбираются данные $b_1(0)$ и $b_2(0)$, переданные на интервале $[0,1]$ согласно правилу

$$\left(\hat{b}_1(0),\hat{b}_2(0)\right)=\arg\max_{b_1,b_2}\Lambda\left[\{r(t),0<t<1\}|b_1,b_2\right],$$

где $\Lambda\left[\{r(t),0<t<1\}|b_1,b_2\right]$ – функция правдоподобия b_1 и b_2 при данном наблюдении $\{r(t),0<t<1\}$. Полезно записать эту максимизацию так

$$\left(\hat{b}_1(0),\hat{b}_2(0)\right)=\arg\max_{b_1}\arg\max_{b_2}\Lambda\left[\{r(t),0<t<1\}|b_1,b_2\right],$$

где величина b_2^*, удовлетворяющая внутренней максимизации, может зависеть от b_1. Заметим, что потребность в «последовательном детектировании» мы избежали.

a Выразите эту максимизацию в наипростейшем виде, используя те же обозначения, что в задаче 15.6 (b). Сведите эту максимизацию к простейшей форме, используя равенство

$$\arg\max_x Ke^{f_1(x)}=\arg\max_x f_1(x),$$

если, $K>0$ не зависит от x.

b какова простейшая структура НК приёмника, если относительная мощность пользователей $A\to 0$? Как она по сравнению с общепринятом детектировании?

c Какова простейшая структура НК приёмника для $B=1$ и произвольных A, σ^2? Как он по сравнению с общепринятом детектированием? Почему?

d Какова простейшая структура НК приёмника для произвольно больших σ^2 и произвольных A и B? Как она по сравнению с общепринятым детектировании? Определите вероятность ошибки для пользователя 1 в этом случае. [Подсказка: используйте факт, что $\operatorname{sgn}(y_2)=\operatorname{sgn}(y_2\pm\rho_{12})$ с высокой вероятностью в рассматриваемом случае].

e Определить вероятность ошибки НК приёмиика для пользователя 1 при $\sigma^2 \to 0$ и произвольно большом A и $B<1$? Как она по сравнению с общепринятым детектированием?

f Какова структура НК приёмника для произвольно большом A и $B<1$ и произвольных σ^2? Как она по сравнению с общепринятым детектированием? Что это говорит об общепринятом детектировании в этом случае? [Подсказка: используйте факт, что $E|y_2|$ грубо в A раз больше чем $E|y_1|$].

15.8. Рассмотрите асинхронную систему связи, показанную на рисунке Р15.8.

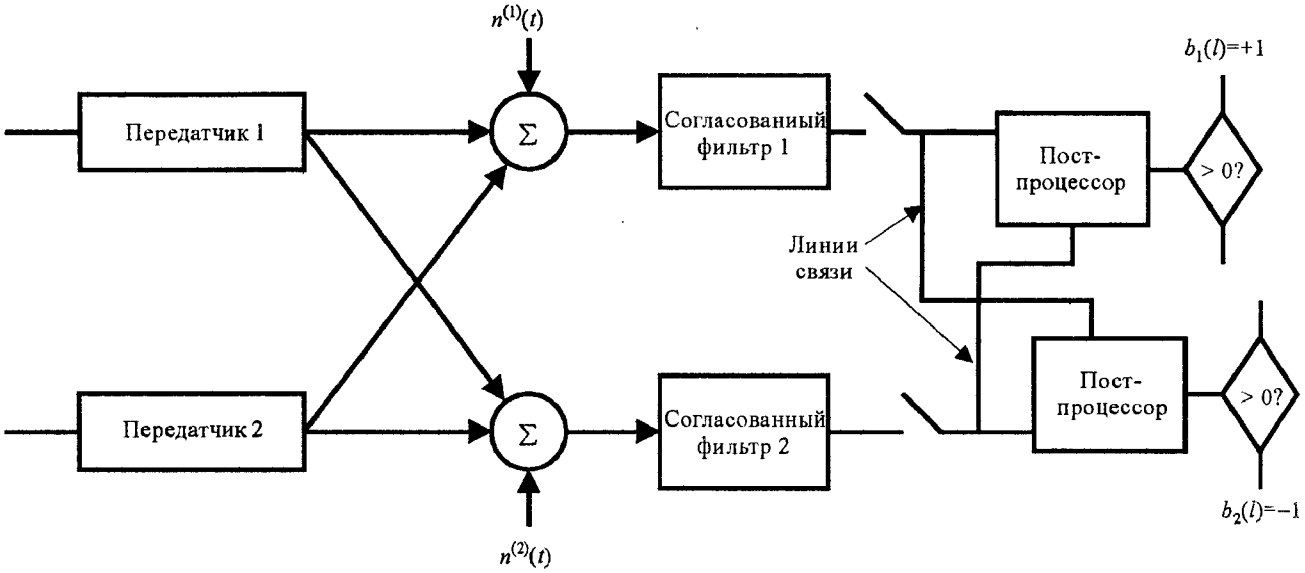


Рис. Р.15.8

Два приёмника разнесены в пространстве, и белые шумовые процессы $n^{(1)}(t)$ и $n^{(2)}(t)$ можно рассматривать как независимые. Шумовые процессы распределены одинаково со спектральной плотностью мощности σ^2, и имеют нулевые средние. Поскольку приёмники разнесены в пространстве, относительные задержки между пользователями не одинаковые – обозначим относительную задержку пользователя k на i-м приёмнике $\tau_k^{(i)}$. Все остальные параметры сигналов совпадают для приёмников, а принимаемый сигнал на i-м приёмнике

$$r^{(i)}(t)=\sum_{k=1}^{2}\sum_{l=-\infty}^{\infty} b_k(l)S_k\left(t-lT-\tau_k^{(i)}\right)+n^{(i)}(t),$$

где s_k определен на интервале $[0,T]$. Вы можете предположить, что i-й приёмник имеет полную информацию об энергиях сигналов и относительных задержках $\tau_1^{(i)}$ и $\tau_2^{(i)}$. Хотя i-й приёмник возможно интересуется только данными от i-го передатчика, заметьте, что имеется свободная линия связи между моделью одного приёмника и постпроцессорными цепями другого. За каждым постпроцессором осуществляется решение пороговым детектированием. В этой задаче, вы можете рассмотреть право выбора за постпроцессором и за линией связи для того, чтобы улучшить качество.

a Какова вероятность ошибки на бит для пользователей 1 и 2 на паре приёмников, которые не используют линии связи и не образуют постпроцессоров. Используйте следующие обозначения:

$$y_k(l)=\int s_k\left(t-lT-\tau_k^{(k)}\right)r^{(k)}(t)dt$$

$$\rho_{12}^{(i)}=\int s_1\left(t-\tau_1^{(i)}\right)s_2\left(t-\tau_2^{(i)}\right)dt$$

$$\rho_{21}^{(i)}=\int s_1\left(t-\tau_1^{(i)}\right)s_2\left(t+T-\tau_2^{(i)}\right)dt$$

$$w_k=\int s_k^2\left(t-\tau_k^{(1)}\right)dt=\int s_k^2\left(t-\tau_k^{(2)}\right)dt.$$

b Рассмотрите постпроцессор для приёмника 1, который принимает $y_2(l-1)$ и $y_2(l)$ с линии связи и реализует следующую послепроцессорную обработку над $y_1(l)$:

$$z_1(l) = y_1(l) - \rho_{21}^{(1)} \operatorname{sgn}[y_2(l-1)] - \rho_{12}^{(1)} \operatorname{sgn}[y_2(l)].$$

Определите точное выражение для вероятности ошибки на бит для пользователя 1.

c Определите асимптотическую эффективность приёмника многих пользователей в (b) и сравните с (a). Будет ли этот приёмник всегда значительно лучше чем прнёмник в (a).

15.9. Базовые сигналы, показанные на рисунке P15.6, синтезированы для двух пользователей, которые делят один и тот же асинхронный, узкополосной канал. Предположите, что $B=1$ и $A=4$. Мы хотели бы сравнить качество различных приёмников по критерию вероятности ошибки $\mathscr{P}_1(0)$. Поскольку это выражение слишком сложное в некоторых случаях, мы хотим также поинтересоваться сравнением аснмптотической эффективности систем многих пользователей η_1 для каждого приёмника. Предположим, что $\tau_1 = 0$, $0 < \tau_2 < T$ фиксировано и известно на приёме и предположим, что мы имеем неограниченный диапазон передачн $2M+1 \to \infty$.

a Для общепринятого многопользовательского приёмника:

(i) Найдите точное выражение для вероятности ошибки для пользователя 1. Выразите этот результат через w_i, ρ_{12}, ρ_{21}, σ^2. [Подсказка: может помочь условная вероятность при фиксированном $b_2(-1)$ и $b_2(0)$.

(ii) Изобразите асимптотическую эффектнвность многих пользователей η_1, как функцию τ_2. Укажите и объясните максимальные н минимальные значения η_1 на этом графике.

b Для НК приёмника:

(i) Изобразите η_1 как функцию τ_2. Объясните максимальные и минимальные значения, сравните с a(ii).

(ii) Какие последовательности ошибок более вероятны для каждого значения τ_2?

c Для ограничивающего детектора с декорреляцией:

(i) Найдите точное выражение для вероятностн ошибки для пользователя 1 с теми же параметрами, как в **a** (i) [Подсказка: не забудьте нормировать ρ_{12} и ρ_{21}].

(ii) Изобразите η_1 как функцию от τ_2. Объясните минимальное значение η_1 в этом случае н сравните с **a** (ii).

15.10. Посимвольный детектор, который минимизирует вероятность ошибки на символ, отличается от минимально правдоподобного детектора последовательности. Последний более сложно описать чем детектор, который выбирает каждый $b_k(0)$ согласно правилу

$$\hat{b}_k(0) = \arg\max_{b_k(0)} \Lambda\left[\{r(t), 0 < r < 1\} b_k(0)\right].$$

a Покажите, что это правило решения мннимизирует $\Lambda\left[b_k(0) \neq \hat{b}_k(0)\right]$ по всем правилам решения при наблюденни $\{r(t), 0 < t < 1\}$. Согласно этому критерию превосходство за НК приёмнником.

b Покажите, что простейшая структура приёмника, обеспечивающего минимизацию вероятности ошибки для пользователя l, определяется так

$$\hat{b}_1(0) = \arg\max_{b_1} \left[\exp\left(\frac{b_1 y_1}{\sigma^2}\right) \operatorname{ch}\left(\frac{y_2 - b_1 \rho_{12}}{\sigma^2}\right)\right].$$

c Найдите простейшую форму для приёмника, обеспечивающего минимум вероятности ошибки, для $B=1$ и произвольных A и σ^2. Как она по сравнению с вышеуказанными приёмниками?

d Найдите предельную форму приёмника, минимизирующего вероятность ошибки, прн произвольно больших σ^2 и произвольных A и B. Сравните с вышеуказанными приёмниками.

e Найдите предельную форму приёмника, минимизирующего вероятность ошибки, для $A >> 1$ и произвольных σ^2, B. Сравните с вышеприведенными приёмниками.

f Найдите предельную форму прнёмника, минимизирующего вероятность ошибки, для $A >> 1$, $\sigma^2 \to 0$ и произвольного B. Сравните с вышеприведенными приёмниками.

15.11. В чистой системе Алоха канальная битовая скорость передачи равна 2400 бит/с. Предположим, что каждый терминал передает в среднем 100 бит сообщений.

a Определите максимальное число терминалов, которые можно использовать в канале.

b повторите (a), если используется щелевая ALOHA.

15.12. Определите максимальный входной трафик для чистой ALOHA и протокол для шелевой ALOHA.

15.13. Для пуассоновского процесса вероятность появления k событий на интервале T равна

$$P(k) = \frac{e^{-\lambda T}(\lambda T)^k}{k!}, \; k = 0,1,2,\ldots .$$

a Определите среднее число появлений событий на интервале T.

b Определите дисперсию σ^2 числа появлений событий на интервале T.

c Какова вероятность того, что на интервале T произойдет хотя бы одно появление события.

d Какова вероятность того, что произойдет точно одно появление события.

15.14. Обратитесь к задаче 15.13. Средняя скорость появления пакетов равна $\lambda = 10$ пакетов/с. Определите:

a Среднее время между появлениями пакетов.

b Вероятность того что другой пакет появится в пределах 1 с, 100 мс.

15.15. Рассмотрите чистую систему ALOHA, которая работает с проходимостью $G = 0,1$, а пакеты генерируются по закону Пуассона с интенсивностью (скоростью) появления λ. Определите:

a Величину λ.

b Среднее число попыток передачи для отправки пакета.

15.16. Рассмотрите систему CSMA/CD в которой скорость передачи по трассе равна 20 Мбит/с. Трасса имеет длину 2 км и задержку при распространении 5 мкс/км. Пакеты имеют по 1000 бит. Определите:

a Задержку из конца в конец τ_d.

b Длину пакета T_p.

c Отношение $a = \tau_d / T_p$.

d Максимальное использование трассы и максимальную скорость передачи символа.

ПРИЛОЖЕНИЕ А

АЛГОРИТМ ЛЕВИНСОНА-ДУРБИНА

Алгоритм Левинсона-Дурбина – рекуррентный метод первого порядка для определения решения системы линейных уравнеиий

$$\mathbf{\Phi}_p \mathbf{a}_p = \boldsymbol{\phi}_p, \tag{A.1}$$

где $\mathbf{\Phi}_p - p \times p$ матрица Тёплица, $\mathbf{a}_p$ – вектор коэффициентов предсказания, выраженный так

$$\mathbf{a}_p^{\mathrm{T}} = \left[a_{p1}\ a_{p2}\ \ldots\ a_{pp}\right],$$

а $\boldsymbol{\phi}_p - p$-мерный вектор с элементами

$$\boldsymbol{\phi}_p^{'} = \left[\phi(1)\ \phi(2) \ldots \phi(p)\right].$$

Для предсказания первого порядка (p=1) имеем решение

$$\begin{aligned} \phi(0) a_{11} &= \phi(1) \\ a_{11} &= \phi(1)/\phi(0). \end{aligned} \tag{A.2}$$

Остаточный средний квадрат ошибки (СКО) для предсказателя первого порядка

$$\mathcal{E}_1 = \phi(0) - a_{11}\phi(1) = \phi(0) - a_{11}^2\phi(0) = \phi(0)\left(1 - a_{11}^2\right). \tag{A.3}$$

В общем, мы можем выразить решение для коэффициентов предсказателя m-го порядка через коэффициенты $(m-1)$-го. Так мы выразим $\mathbf{a}_m$ как сумму двух векторов, именно

$$\mathbf{a}_m = \begin{bmatrix} a_{m1} \\ a_{m2} \\ \vdots \\ a_{mm} \end{bmatrix} = \begin{bmatrix} \mathbf{a}_{m-1} \\ \cdots\cdots \\ 0 \end{bmatrix} + \begin{bmatrix} \mathbf{d}_{m-1} \\ \cdots\cdots \\ k_m \end{bmatrix}, \tag{A.4}$$

где вектор $\mathbf{d}_{m-1}$ и скаляр k_m надо определить. Таким образом, $\mathbf{\Phi}_m$ можно выразить так

$$\mathbf{\Phi}_m = \begin{bmatrix} \mathbf{\Phi}_{m-1} & \vdots & \boldsymbol{\phi}_{m-1}^r \\ \cdots\cdots & \cdots & \cdots\cdots \\ \boldsymbol{\phi}_{m-1}^r & \vdots & \phi(0) \end{bmatrix}, \tag{A.5}$$

где $\boldsymbol{\phi}_{m-1}^r$ как раз вектор $\boldsymbol{\phi}_{m-1}$ в обратном порядке. Теперь

$$\begin{bmatrix} \mathbf{\Phi}_{m-1} & \vdots & \boldsymbol{\phi}_{m-1}^r \\ \cdots\cdots & \cdots & \cdots\cdots \\ \boldsymbol{\phi}_{m-1}^r & \vdots & \phi(0) \end{bmatrix} \left(\begin{bmatrix} \mathbf{a}_{m-1} \\ \cdots\cdots \\ 0 \end{bmatrix} + \begin{bmatrix} \mathbf{d}_{m-1} \\ \cdots\cdots \\ k_m \end{bmatrix} \right) = \begin{bmatrix} \boldsymbol{\phi}_{m-1} \\ \cdots\cdots \\ \phi(m) \end{bmatrix}. \tag{A.6}$$

Из (A.6) мы получаем два уравнения. Первое – это матричное уравнение

$$\mathbf{\Phi}_{m-1}\mathbf{a}_{m-1} + \mathbf{\Phi}_{m-1}\mathbf{d}_{m-1} + k_m\boldsymbol{\phi}_{m-1}^r = \boldsymbol{\phi}_{m-1}. \tag{A.7}$$

Но $\mathbf{\Phi}_{m-1}\mathbf{a}_{m-1} = \boldsymbol{\phi}_{m-1}$. Следовательно, (A.7) упрощается:

$$\mathbf{\Phi}_{m-1}\mathbf{d}_{m-1} + k_m\boldsymbol{\phi}_{m-1}^r = 0. \tag{A.8}$$

Это уравнение имеет решение

$$\mathbf{d}_{m-1} = -k_m\mathbf{\Phi}_{m-1}^{-1}\boldsymbol{\phi}_{m-1}^r. \tag{A.9}$$

Но $\boldsymbol{\phi}^{r}_{m-1}$ равно $\boldsymbol{\phi}_{m-1}$ в обратном порядке. Следовательно, решение (А.9) равно $\mathbf{a}_{m-1}$ в обратном порядке, умноженном на $-k_m$. Это значит

$$\mathbf{d}_{m-1} = -k_m \begin{bmatrix} a_{m-1\,m-1} \\ a_{m-1\,m-2} \\ \vdots \\ a_{m-1\quad 1} \end{bmatrix}. \tag{A.10}$$

Второе уравнение, получаемое из (А.6), – скалярное уравнение

$$\boldsymbol{\phi}^{r}_{m-1}\mathbf{a}_{m-1} + \boldsymbol{\phi}^{r}_{m-1}\mathbf{d}_{m-1} + \phi(0)k_m = \phi(m). \tag{A.11}$$

Мы исключаем $\mathbf{d}_{m-1}$ из (А.11), используя (А.10). Окончательное уравнение дает нам k_m, то есть

$$k_m = \frac{\phi(m) - \boldsymbol{\phi}^{r}_{m-1}\mathbf{a}_{m-1}}{\phi(0) - \boldsymbol{\phi}^{r}_{m-1}\boldsymbol{\Phi}^{-1}_{m-1}\boldsymbol{\phi}_{m-1}} = \frac{\phi(m) - \boldsymbol{\phi}^{r}_{m-1}\mathbf{a}_{m-1}}{\phi(0) - \mathbf{a}^{r}_{m-1}\boldsymbol{\phi}_{m-1}} = \frac{\phi(m) - \boldsymbol{\phi}^{r}_{m-1}\mathbf{a}_{m-1}}{\mathscr{E}_{m-1}}, \tag{A.12}$$

где $\mathscr{E}_{m-1}$ – остаточный СКО, определяемый так

$$\mathscr{E}_{m-1} = \phi(0) - \mathbf{a}^{r}_{m-1}\boldsymbol{\phi}_{m-1}. \tag{A.13}$$

Подстановкой (А.10) для $\mathbf{d}_{m-1}$ в (А.4) мы получаем рекуррентное соотношение первого порядка

$$a_{mk} = a_{m-1\,k} - k_m a_{m-1\,m-k}, \qquad k = 1, 2, \ldots, m-1, \qquad m = 1, 2, \ldots, p \tag{A.14}$$

и

$$a_{mm} = k_m.$$

Минимум СКО можно также вычислить рекуррентно. Мы имеем

$$\mathscr{E}_m = \phi(0) - \sum_{k=1}^{m} a_{mk}\phi(k). \tag{A.15}$$

Используя (А.14) в (А.15) мы получим

$$\mathscr{E}_m = \phi(0) - \sum_{k=1}^{m-1} a_{m-1\,k}\phi(k) - a_{mm}\left[\phi(m) - \sum_{k=1}^{m-1} a_{m-1\,km-k}\phi(k)\right]. \tag{A.16}$$

Но слагаемое в квадратных скобках в (А.16) – это и есть числитель для k_m в (А.12). Следовательно,

$$\mathscr{E}_m = \mathscr{E}_{m-1} - a^2_{mm}\mathscr{E}_{m-1} = \mathscr{E}_{m-1}\left(1 - a^2_{mm}\right). \tag{A.17}$$

ПРИЛОЖЕНИЕ **B**

ВЕРОЯТНОСТЬ ОШИБКИ ДЛЯ МНОГОКАНАЛЬНЫХ ДВОИЧНЫХ СИГНАЛОВ

В многоканальных системах связи, которые используют двоичные сигналы для передачи информации по каналу с АБГШ, величины для решения у детектора можно выразить как частный случай общей квадратичной формы

$$D = \sum_{k=1}^{L}\left(A|X_k|^2 + B|Y_k|^2 + CX_kY_k^* + C^*X_k^*Y_k\right) \tag{B.1}$$

через комплексные гауссовские случайные переменные, A, B и C являются константами; X_k и Y_k – пара коррелированных гауссовских случайных величин. Для рассматриваемых каналов L пар $\{X_k, Y_k\}$ – статистически взаимно независимы и одинаково распределены.

Вероятность ошибки определяется вероятностью того, что $D<0$. Эта вероятность вычисляется ниже.

Вычисления начинаются с характеристической функции, обозначаемое $\psi_D(j\upsilon)$, общей квадратичной формы. Вероятность того, что $D<0$, обозначаемое здесь как вероятность ошибки P_b, равна

$$P_b = P(D<0) = \int_{-\infty}^{0} p(D)\,dD, \tag{B.2}$$

где $p(D)$ – ФПВ для D связанна с $\psi_D(j\upsilon)$ преобразованием Фурье, то есть

$$p(D) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \psi_D(j\upsilon)\mathrm{e}^{-j\upsilon D}\,d\upsilon.$$

Следовательно,

$$P_b = \int_{-\infty}^{0} dD \frac{1}{2\pi}\int_{-\infty}^{\infty} \psi_D(j\upsilon)\mathrm{e}^{-j\upsilon D}\,d\upsilon. \tag{B.3}$$

Переменим порядок интегрирования и сначала выполним интегрирование по D. В результате получаем

$$P_b = -\frac{1}{2\pi j}\int_{-\infty+j\varepsilon}^{\infty+j\varepsilon} \frac{\psi_D(j\upsilon)}{\upsilon}\,d\upsilon, \tag{B.4}$$

где малое положительное число ε введено для того, что сдвинуть путь интегрирования от точки сингулярности $\upsilon=0$. Оно должно быть положительным для того, чтобы было возможным менять порядок интегрирования.

Поскольку D является суммой статистически независимых случайных переменных характеристическая функция D определяется произведением L характеристических функций, причем каждая функция соответствует индивидуальным случайным переменным d_k, где

$$d_k = A|X_k|^2 + B|Y_k|^2 + CX_kY_k^* + C^*X_k^*Y_k.$$

Характеристическая функция d_k равна

$$\phi_{d_k}(j\upsilon) = \frac{\upsilon_1\upsilon_2}{(\upsilon+j\upsilon_1)(\upsilon-j\upsilon_2)}\exp\left[\frac{\upsilon_1\upsilon_2\left(-\upsilon^2\alpha_{1k}+j\upsilon\alpha_{2k}\right)}{(\upsilon+j\upsilon_1)(\upsilon-j\upsilon_2)}\right], \tag{B.5}$$

где параметры υ_1, υ_2, α_{1k} и α_{2k} зависят от средних $\overline{X}_k$ и $\overline{Y}_k$ и вторых (центральных) моментов μ_{xx}, μ_{yy}, и μ_{xy} комплексных гауссовских величин X_k и Y_k через следующее определенне $\left(|C|^2 - AB > 0\right)$:

$$\upsilon_1 = \sqrt{w^2 + \frac{1}{4\left(\mu_{xx}\mu_{yy} - |\mu_{xy}|^2\right)\left(|C|^2 - AB\right)}} - w; \quad \upsilon_2 = \sqrt{w^2 + \frac{1}{4\left(\mu_{xx}\mu_{yy} - |\mu_{yy}|^2\right)\left(|C|^2 - AB\right)}} + w$$

$$w = \frac{A\mu_{xx} + B\mu_{yy} + C\mu_{xy}^* + C^*\mu_{xy}}{4\left(\mu_{xx}\mu_{yy} - |\mu_{xy}|^2\right)\left(|C|^2 - AB\right)}; \quad \alpha_{1k} = 2\left(|C|^2 - AB\right)\left(|\overline{X}_k|^2\mu_{yy} + |\overline{Y}_k|^2\mu_{xx} - \overline{X}_k^*\overline{Y}_k\mu_{xy} - \overline{X}_k\overline{Y}_k^*\mu_{xy}^*\right) \tag{B.6}$$

$$\alpha_{2k} = A|\overline{X}_k|^2 + B|\overline{Y}_k|^2 + C\overline{X}_k^*\overline{Y}_k + C^*\overline{X}_k\overline{Y}_k^*; \quad \mu_{xy} = \tfrac{1}{2}E\left[\left(X_k - \overline{X}_k\right)\left(Y_k - \overline{Y}_k\right)^*\right].$$

Теперь, как результат независимости случайных величин d_k, характеристическая функция D равна

$$\psi_D(j\upsilon) = \prod_{k=1}^{L}\psi_{d_k}(j\upsilon); \qquad \psi_D(j\upsilon) = \frac{(\upsilon_1\upsilon_2)^L}{(\upsilon+j\upsilon_1)^L(\upsilon-j\upsilon_2)^L}\exp\left[\frac{\upsilon_1\upsilon_2\left(j\upsilon\alpha_2 - \upsilon^2\alpha_1\right)}{(\upsilon+j\upsilon_1)(\upsilon-j\upsilon_2)}\right], \tag{B.7}$$

где

$$\alpha_1 = \sum_{k=1}^{L}\alpha_{1k}, \quad \alpha_2 = \sum_{k=1}^{L}\alpha_{2k}. \tag{B.8}$$

Результат (B.7) подставляется для $\psi_D(j\upsilon)$ в (B.4) и мы получаем

$$P_b = -\frac{(\upsilon_1\upsilon_2)^L}{2\pi j}\int_{-\infty+j\varepsilon}^{\infty+j\varepsilon}\frac{d\upsilon}{\upsilon(\upsilon+j\upsilon_1)^L(\upsilon-j\upsilon_2)^L}\exp\left[\frac{\upsilon_1\upsilon_2\left(j\upsilon\alpha_2 - \upsilon^2\alpha_1\right)}{(\upsilon+j\upsilon_1)(\upsilon-j\upsilon_2)}\right]. \tag{B.9}$$

Этот интеграл вычисляется следующим образом.

Первый шаг сводится к выражению экспоненциальной функции в виде

$$\exp\left(-A_1+\frac{jA_2}{\upsilon+j\upsilon_1}-\frac{jA_3}{\upsilon-j\upsilon_2}\right),$$

причём можно легко проверить, что константы A_1, A_2 и A_3 определяются так:

$$A_1=\alpha_1\upsilon_1\upsilon_2;\quad A_2=\frac{\upsilon_1^2\upsilon_2}{\upsilon_1+\upsilon_2}(\alpha_1\upsilon_1+\alpha_2);\quad A_3=\frac{\upsilon_1\upsilon_2^2}{\upsilon_1+\upsilon_2}(\alpha_1\upsilon_2-\alpha_2). \tag{B.10}$$

На втором шаге выполняется конформное преобразование из υ-плоскости в p-плоскость посредством преобразования переменной

$$p=-\frac{\upsilon_1\upsilon-j\upsilon_2}{\upsilon_2\upsilon+j\upsilon_1}. \tag{B.11}$$

В p-плоскости интеграл (B.9) приводится к выражению

$$P_b=\frac{\exp\left[\upsilon_1\upsilon_2\left(-2\alpha\upsilon_1\upsilon_2+\alpha_2\upsilon_1-\alpha_2\upsilon_2\right)/(\upsilon_1+\upsilon_2)^2\right]}{(1+\upsilon_2/\upsilon_1)^{2L-1}}\frac{1}{2\pi j}\int_\Gamma f(p)dp, \tag{B.12}$$

где

$$f(p)=\frac{[1+(\upsilon_2/\upsilon_1)p]^{2L-1}}{p^L(1-p)}\exp\left[\frac{A_2(\upsilon_2/\upsilon_1)}{\upsilon_1+\upsilon_2}p+\frac{A_3(\upsilon_2/\upsilon_1)}{\upsilon_1+\upsilon_2}\frac{1}{p}\right], \tag{B.13}$$

а Г– круговой контур радиуса меньше единицы включает начало координат.

Третий шаг сводится к вычислению интеграла

$$\frac{1}{2\pi j}\int_\Gamma f(p)dp=\frac{1}{2\pi j}\int_\Gamma\frac{(1+(\upsilon_2/\upsilon_1)p)^{2L-1}}{p^L(1-p)}\exp\left[\frac{A_2(\upsilon_2/\upsilon_1)}{\upsilon_1+\upsilon_2}p+\frac{A_3(\upsilon_2/\upsilon_1)}{\upsilon_1+\upsilon_2}\frac{1}{p}\right]dp. \tag{B.14}$$

Для того, что облегчить последующие выкладки, вводятся константы $a\ge 0$ и $b\ge 0$

$$\tfrac{1}{2}a^2=\frac{A_3(\upsilon_2/\upsilon_1)}{\upsilon_1+\upsilon_2},\qquad \tfrac{1}{2}b^2=\frac{A_2(\upsilon_2/\upsilon_1)}{\upsilon_1+\upsilon_2}. \tag{B.15}$$

Выразим также функцию $(1+(\upsilon_2/\upsilon_1)p)^{2L-1}$ как биномиальный ряд. В результате получаем

$$\frac{1}{2\pi j}\int_\Gamma f(p)dp=\sum_{k=0}^{2L-1}\binom{2L-1}{k}\left(\frac{\upsilon_2}{\upsilon_1}\right)^k\frac{1}{2\pi j}\int_\Gamma\frac{p^k}{p^L(1-p)}\exp\left(\frac{\frac{1}{2}a^2}{p}+\tfrac{1}{2}b^2p\right)dp \tag{B.16}$$

Контурный интеграл в (B.16) является одним из представлений функции Бесселя. Его можно разрешить, используя соотношение

$$I_n(ab)=\begin{cases}\dfrac{1}{2\pi j}\left(\dfrac{a}{b}\right)^n\displaystyle\int_\Gamma\frac{1}{p^{n+1}}\exp\left(\frac{\frac{1}{2}a^2}{p}+\tfrac{1}{2}b^2p\right)dp\\[2ex]\dfrac{1}{2\pi j}\left(\dfrac{b}{a}\right)^n\displaystyle\int_\Gamma p^{n-1}\exp\left(\frac{\frac{1}{2}a^2}{p}+\tfrac{1}{2}b^2p\right)dp,\end{cases}$$

где $I_n(x)$– модифицированная функции Бесселя n-го порядка первого рода. Представление Q-функции Маркума в ряд функций Бесселя имеет вид

$$Q_1(a,b)=\exp\left[-\tfrac{1}{2}\left(a^2+b^2\right)\right]+\sum_{n=0}^{\infty}\left(\frac{a}{b}\right)^nI_n(ab).$$

Рассмотрим сначала случай $0\le k\le L-2$ в (B.16). В этом случае результирующий контурный интеграл можно записать в форме[1]

---

[1] Этот контурный интеграл связаи с обобщенной функцией Маркума, определяемой так:

$$Q_m(a,b)=\int_b^\infty x(x/a)^{m-1}\exp\left[-\tfrac{1}{2}\left(x^2+a^2\right)\right]I_{m-1}(ax)dx,\quad m\ge 1,\ \text{или}$$

$$Q_m(a,b)\exp\left[\tfrac{1}{2}\left(x^2+a^2\right)\right]=\frac{1}{2\pi j}\int_\Gamma\frac{1}{p^m(1-p)}\exp\left(\frac{\frac{1}{2}a^2}{p}+\tfrac{1}{2}b^2p\right)dp.$$

$$\frac{1}{2\pi j}\int_{\Gamma}\frac{1}{p^{L-k}(1-p)}\exp\left(\frac{\frac{1}{2}a^2}{p}+\frac{1}{2}b^2p\right)dp=Q_1(a,b)\exp\left[\tfrac{1}{2}\left(a^2+b^2\right)\right]+\sum_{n=1}^{L-1-k}\left(\frac{a}{b}\right)^n I_n(ab). \qquad \text{(B.17)}$$

Далее, рассмотрим слагаемое $k=L-1$. Результирующий контурный интеграл может быть выражен через Q-функцию так:

$$\frac{1}{2\pi j}\int_{\Gamma}\frac{1}{p(1-p)}\exp\left(\frac{\frac{1}{2}a^2}{p}+\frac{1}{2}b^2p\right)dp=Q_1(a,b)\exp\left[\tfrac{1}{2}\left(a^2+b^2\right)\right]. \qquad \text{(B.18)}$$

Наконец, рассмотрим случай $L\le k\le 2L-1$. Имеем

$$\begin{aligned}\frac{1}{2\pi j}\int_{\Gamma}\frac{p^{k-L}}{1-p}\exp\left(\frac{\frac{1}{2}a^2}{p}+\frac{1}{2}b^2p\right)dp&=\sum_{n=0}^{\infty}\frac{1}{2\pi j}\int_{\Gamma}p^{k-L+n}\exp\left(\frac{\frac{1}{2}a^2}{p}+\frac{1}{2}b^2p\right)dp=\\ =\sum_{n=k+1-L}^{\infty}\left(\frac{a}{b}\right)^n I_n(ab)&=Q_1(a,b)\exp\left[\tfrac{1}{2}\left(a^2+b^2\right)\right]-\sum_{n=0}^{k-L}I_n(ab).\end{aligned} \qquad \text{(B.19)}$$

Собирая слагаемые, указанные в первой части (B.16), и используя результаты, даваемые (B.17)–(B.19), получаем после некоторых преобразований следующее выражение для контурного интеграла

$$\begin{aligned}\frac{1}{2\pi j}\int_{\Gamma}f(p)dp=\left(1+\frac{\upsilon_2}{\upsilon_1}\right)^{2L-1}&\left[\exp\left[\tfrac{1}{2}\left(a^2+b^2\right)\right]Q_1(a,b)-I_0(ab)\right]+I_0(ab)\sum_{k=1}^{L-1}\binom{2L-1}{k}\left(\frac{v_2}{v_1}\right)^k+\\ &+\sum_{n=1}^{L-1}I_0(ab)\sum_{k=1}^{L-1-n}\binom{2L-1}{k}\left[\left(\frac{b}{a}\right)^n\left(\frac{\upsilon_2}{\upsilon_1}\right)^k-\left(\frac{a}{b}\right)^n\left(\frac{\upsilon_2}{\upsilon_1}\right)^{2L-1-k}\right].\end{aligned} \qquad \text{(B.20)}$$

Уравнение (B.20) в соединении с (B.12) даёт результат для вероятности ошибки. Следует дальнейшее упрощение, если использовать следующее тождество, которое можно легко доказать

$$\exp\left[\frac{\upsilon_1\upsilon_2}{(\upsilon_1+\upsilon_2)^2}\left(-2\alpha_1\upsilon_1\upsilon_2-\alpha_2\upsilon_1-\alpha_2\upsilon_2\right)\right]=\exp\left[-\tfrac{1}{2}\left(a^2+b^2\right)\right].$$

Таким образом, следует, что

$$\begin{aligned}P_b&=Q_1(a,b)-I_0(ab)\exp\left[-\tfrac{1}{2}\left(a^2+b^2\right)\right]+\\ &+\frac{I_0(ab)\exp\left[-\frac{1}{2}\left(a^2+b^2\right)\right]}{(1+\upsilon_2/\upsilon_1)^{2L-1}}\sum_{k=0}^{L-1}\binom{2L-1}{k}\left(\frac{\upsilon_2}{\upsilon_1}\right)^k+\frac{\exp\left[-\frac{1}{2}\left(a^2+b^2\right)\right]}{(1+\upsilon_2/\upsilon_1)^{2L-1}}\times\\ &\times\sum_{n=1}^{L-1}I_0(ab)\sum_{k=0}^{L-1-n}\binom{2L-1}{k}\times\left[\left(\frac{b}{a}\right)^n\left(\frac{v_2}{v_1}\right)^k-\left(\frac{a}{b}\right)^n\left(\frac{v_2}{v_1}\right)^{2L-1-k}\right]\qquad (L>1),\end{aligned} \qquad \text{(B.21)}$$

$$P_b=Q_1(a,b)-\frac{\upsilon_2/\upsilon_1}{1+\upsilon_2/\upsilon_1}I_0(ab)\exp\left[-\tfrac{1}{2}\left(a^2+b^2\right)\right]\qquad (L=1).$$

Это и есть искомое выражение для вероятности ошибки. Теперь несложно связать параметры a и b с моментами пары $\{X_k, Y_k\}$. Подставив выражения для A_2 и A_3 из (B.10) в (B.15), получаем

$$a=\left[\frac{2\upsilon_1^2\upsilon_2(\alpha_1\upsilon_2-\alpha_2)}{(\upsilon_1+\upsilon_2)^2}\right]^{1/2},\qquad b=\left[\frac{2\upsilon_1\upsilon_2^2(\alpha_1\upsilon_1+\alpha_2)}{(\upsilon_1+\upsilon_2)^2}\right]^{1/2}. \qquad \text{(B.22)}$$

Поскольку υ_1, υ_2, α_1 и α_2 определяются (B.6) и (B.8) непосредственно через моменты пар X_k и Y_k наша задача завершена.

ПРИЛОЖЕНИЕ C

ВЕРОЯТНОСТИ ОШИБКИ ДЛЯ АДАПТИВНОГО ПРИЁМА *M*-ФАЗНЫХ СИГНАЛОВ

В этом приложенни мы определим вероятности ошибки для двух- и четырехфазовых сигналов при передачи по неизменному во времени каналу с аддиивным белым гауссовским шумом при L-кратном разнесении и для M-фазовых сигналов по каналу с релеевскими замираниями и адаптивным белым гауссовским шумом с L-кратным разнесением. Оба канала искажают передаваемые сигналы путем введения аддитивного белого гауссовского шума и случайного мультипликативного ослабления и фазового сдвига в сигнале. Обработка сигнала в приёмнике состоит из нахождения взаимной корреляции сигнала в смеси с шумом, принимаемого в каждой ветви разнесения, с опорным сигналом, получаемым или от предыдущего принятого информационного сигнала, или от пилот-сигнала, и суммирования выходов всех L каналов разнесения для формирования величины для решения.

C.1. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ДЛЯ СИСТЕМЫ СВЯЗИ С *M*-ФАЗОВЫМ СИГНАЛОМ

В общем случае M-фазовой модуляции передаваемый сигнал имеет вид[1]:

$$s_n(t) = \mathrm{Re}\left[s_{ln}(t)\mathrm{e}^{j2\pi f_c t}\right],$$

где

$$s_{ln}(t) = g(t)\exp\left[j\frac{2\pi}{M}(n-1)\right], \quad n = 1,2,...,M, \quad 0 \le t \le T, \tag{C.1}$$

а T – длительность сигнального иитервала.

Рассмотрим случай, когда на протяжении сигнального интервала передаётся один из M сигналов по L каналам. Предположим, что в каждом из каналов передаваемый сигнал искажается введением мультипликативного ослабления и фазового сдвига, представленных комплексным множителем g_k и аддитивным шумом $z_k(t)$. Так, если передаваемый сигнал $s_{ln}(t)$, то принимаемый сигнал в k-м канале

$$r_{lk}(t) = g_k s_{ln}(t) + z_k(t), \quad 0 \le t \le T, \quad k = 1,2,...,L. \tag{C.2}$$

Шум $\{z_k(t)\}$ считается реализацией стационарного белого гауссовского случайного процесса с нулевым средним и автокорреляционной функцией $\phi_z(\tau) = N_0\delta(\tau)$, где N_0 – спектральная плотность мощности шума..

В демодуляторе $r_{lk}(t)$ пропускается через фильтр, импульсная характеристика которого согласована с сигналом $g(t)$. Выход этого фильтра в момент стробирования $t = T$ обозначаем так:

$$X_k = 2\mathscr{E}g_k \exp\left[j\frac{2\pi}{M}(n-1)\right] + N_k, \tag{C.3}$$

где $\mathscr{E}$ – энергия переданного сигнала, а N_k – отсчёт шума на выходе в k-го фильтра. Для того, чтобы демодулятор мог решить, какая из M фаз передана по каналу на сигнальном интсрвале $0 \le t \le T$, следует попытаться убрать фазовый сдвиг, введённый в каждом канале. На практике это осуществляется путем умножения выхода фильтра X_k на комплексно сопряженную величину оценки $\hat{g}_k$ канального ослабления и фазового сдвига. Результатом является взвешенный и сдвинутый по фазе выходной отсчёт фильтра в k-м канале, который затем суммируется со взвешенным н сдвинутым по фазе выходными отсчетами остальных $L-1$ канальных фильтров.

---

[1] Повсюду используется комплексное представление вещественных сигналов. Комплексное сопряжение отмечено звездочкой.

Считается, что оценка $\hat{g}_k$ ослабления и фазового сдвига в k-м канале определяется или от передачи пилот-сигнала или путем снятия модуляции в информационном сигнале, принятом на предыдущем сигнальном интервале. Как пример формирователя, предположим, что пилот-сигнал, обозначенный $s_{pk}(t)$, $0 \le t \le T$, передаётся по k-му каналу с целью измерения ослабления и фазового сдвига в канале. Принимаемый сигнал равен

$$g_k s_{pk}(t) + z_{pk}(t), \qquad 0 \le t \le T,$$

где $z_{pk}(t)$ – отсчётная функция стационарного белого гауссовского случайного процесса с нулевым средним и автокорреляционной функцией $\phi_p(\tau) = N_0 \delta(\tau)$. Этот сигнал плюс шум пропускается через фильтр, согласованный с $s_{pk}(t)$. Отсчёт выхода фильтра в момент $t = T$ содержит случайную переменную $X_{pk} = 2\mathscr{E}_p g_k + N_{pk}$, где $\mathscr{E}_p$ – энергия пилот сигнала, которая считается одинаковой для всех каналов, а N_{pk} – отсчёт аддитивного шума. Оценка g_k получается путем соответствующей нормировки X_{pk}, то есть $\hat{g}_k = g_k + N_{pk}/2\mathscr{E}_p$.

С другой стороны, оценку g_k можно получить из информационного сигнала следующим образом. Если знать информационные компоненты, содержащиеся в выходе согласованного фильтра, то оценку g_k можно получить соответствующей нормировкой этого выхода. Для примера, информационная компонента выхода фильтра в (С.3) равна $2\mathscr{E}g_k \exp[j(2\pi/M)(n-1)]$, и, следовательно, оценка

$$\hat{g}_k = \frac{X_k}{2\mathscr{E}} \exp\left[-j\frac{2\pi}{M}(n-1)\right] = g_k + \frac{N_k'}{2\mathscr{E}},$$

где $N_k' = N_k \exp[-j(2\pi/M)(n-1)]$, а ФПВ N_k' аналогична ФПВ N_k. Оценка, полученная от информационного сигнала таким путём, иногда называется *ясновидящей* оценкой. Хотя физически реализуемый приемник не может обладать таким «ясновидением», он может аппроксимировать эту оценку путем использования временной задержки сигнального интервала и получить путём обратной связи оценку переданной фазы на предшествующем временном интервале.

Получена ли оценка g_k от пилот-сигнала или от информационного сигнала, оценку можно улучшить путем расширения временного интервала в ее формировании, чтобы включить несколько предыдущих сигнальных интервалов способом, описанном Прайсом (1962a, b). Результатом расширения интервала измерения является увеличение отношения сигнал-шум в оценке g_k. В общем случае, если интервалы оценивания не ограничены, нормированная *оценка по пилот-сигналу*

$$\hat{g}_k = g_k + \sum_{l=1}^{\infty} c_l N_{pkl} \Big/ 2\mathscr{E}_p \sum_{l=1}^{\infty} c_l, \qquad \text{(С.4)}$$

где c_l – взвешивающее коэффициенты для подоценок g_k, полученных на l-м предшествующем сигнальном интервале, а N_{pk} – отсчёт аддитивного гауссовского шума на выходе фильтра, согласованного с $s_{pk}(t)$ на l-м предшествующем сигнальном интервале. Аналогично «ясновидящая» оценка, полученная из информационного сигнала путем снятия модуляции, при неограниченном интервале обработки, равна

$$\hat{g}_k = g_k + \sum_{l=1}^{\infty} c_l N_{kl} \Big/ 2\mathscr{E} \sum_{l=1}^{\infty} c_l. \qquad \text{(С.5)}$$

Как указывалось, демодулятор образует произведение между $\hat{g}_k^*$ и X_k и суммирует его с аналогичными произведениями остальных $L-1$ каналов. В результате получаются случайные величины

$$z = \sum_{k=1}^{L} X_k \hat{g}_k^* = \sum_{k=1}^{L} X_k Y_k^* = z_r + j z_i, \qquad \text{(С.6)}$$

где, по определению, $Y_k = \hat{g}_k$, $z_r = \mathrm{Re}(z)$, $z_i = \mathrm{Im}(z)$. Фаза z является величиной для решения. Она равна

$$\theta = \mathrm{arctg}\left(\frac{z_i}{z_r}\right) = \mathrm{arctg}\left[\mathrm{Im}\left(\sum_{k=1}^{L} X_k Y_k^*\right) \Big/ \mathrm{Re}\left(\sum_{k=1}^{L} X_k Y_k^*\right)\right]. \qquad \text{(С.7)}$$

С.2. ХАРАКТЕРИСТИЧЕСКАЯ ФУНКЦИЯ И ФУНКЦИЯ ПЛОТНОСТИ ВЕРОЯТНОСТИ ФАЗЫ

Будем исходить из предположения, что фаза переданного сигнала равна нулю, то есть $n=1$. При необходимости ФПВ для θ при условии передачи фазы другого переданного сигнала можно получить преобразованием $p(\theta)$ на угол $2\pi(n-1)/M$. Мы также предполагаем, что комплексные слагаемые $\{g_k\}$, характеризующие L каналов, статистически взаимно независимые и одинаково распределенные гауссовские случайные величины с нулевыми средними. Такая характеристика приемлема для каналов с медленными релеевскими замираниями. Как следствие, случайные величины (X_k, Y_k) гауссовские, коррелированные, комплексные, с нулевыми средними и статистически независимые, но одинаково распределённые с любой другой парой (X_i, Y_i).

Метод, который используется при расчёте плотности вероятности $p(\theta)$, в общем случае разнесённого приёма, следующий. Во-первых, определяется характеристическая функция совместной функции плотности вероятности z_r и z_i, где z_r и z_i – две компоненты, которые определяют величину для решения θ. Во-вторых, выполняются двойное преобразование Фурье характеристической функции, что даст $p(z_r, z_i)$. Затем преобразование

$$r=\sqrt{z_r^2+z_i^2}, \qquad \theta=\operatorname{arctg}\left(\frac{z_i}{z_r}\right) \tag{С.8}$$

даст совместную ФПВ огибающей r и фазы θ. В заключение интегрирование совместной ФПВ по случайной величине r даст ФПВ для θ.

Совместная характеристическая функция случайных величин z_r и z_i можно выразить в виде

$$\psi(j\upsilon_1, j\upsilon_2)=\left[\frac{4}{\left(\upsilon_1-j\dfrac{2|\mu|\cos\varepsilon}{\sqrt{m_{xx}m_{yy}\left(1-|\mu|^2\right)}}\right)\left(\upsilon_2-j\dfrac{2|\mu|\sin\varepsilon}{\sqrt{m_{xx}m_{yy}\left(1-|\mu|^2\right)}}\right)+\dfrac{4}{m_{xx}m_{yy}\left(1-|\mu|^2\right)}}\right]^L, \tag{С.9}$$

где, по определению,

$$\left.\begin{aligned} m_{xx}&=E\left(|X_k|^2\right)\\ m_{yy}&=E\left(|Y_k|^2\right)\\ m_{xy}&=E\left(|X_kY_k^*|^2\right)\end{aligned}\right\}\text{ одинаковы для всех } k$$

$$\mu=\frac{m_{xy}}{\sqrt{m_{xx}m_{yy}}}=|\mu|e^{j\varepsilon}. \tag{С.10}$$

Результат преобразования Фурье $\psi(j\upsilon_1, j\upsilon_2)$ по величинам υ_1 и υ_2 даёт

$$p(z_r, z_i)=\frac{\left(1-|\mu|^2\right)^L}{(L-1)!\pi 2^L}\left(\sqrt{z_r^2+z_i^2}\right)^{L-1}\exp\left[|\mu|\left(z_r\cos\varepsilon+z_i\sin\varepsilon\right)\right]K_{L-1}\left(\sqrt{z_r^2+z_i^2}\right), \tag{С.11}$$

где $K_n(x)$ – модифицированная функция Ханкеля порядка n. Затем преобразование случайных величин, указанных в (С.8), даёт совместную ФПВ огибающей r и фазы θ в виде

$$p(r,\theta)=\frac{\left(1-|\mu|^2\right)^L}{(L-1)!\pi 2^L}r^L\exp\left[|\mu|r\cos(\theta-\varepsilon)\right]K_{L-1}(r). \tag{С.12}$$

Теперь интегрирование по величине r даёт собственно ФПВ фазы θ. Мы вычислим интеграл для получения $p(\theta)$ в виде

$$p(\theta)=\frac{(-1)^{L-1}\left(1-|\mu|^2\right)^L}{2\pi(L-1)!}\left\{\frac{\partial^{L-1}}{\partial b^{L-1}}\left[\frac{1}{b-|\mu|^2\cos^2(\theta-\varepsilon)}+\right.\right.$$
$$\left.\left.+\frac{|\mu|\cos(\theta-\varepsilon)}{\left[b-|\mu|^2\cos^2(\theta-\varepsilon)\right]^{3/2}}\arccos\left(-\frac{|\mu|\cos(\theta-\varepsilon)}{b^{1/2}}\right)\right]\right\}\Bigg|_{b=1}. \qquad \text{(C.13)}$$

В этом уравнении обозначение

$\left.\frac{\partial^L}{\partial b^L}f(b,\mu)\right|_{b=1}$ означает L-ю частную производную функции $f(b,\mu)$, при $b=1$.

C.3. ВЕРОЯТНОСТИ ОШИБКИ ДЛЯ КАНАЛОВ С МЕДЛЕННЫМИ РЕЛЕЕВСКИМИ ЗАМИРАНИЯМИ

В этом разделе определяется вероятность характерных ошибок и вероятность ошибочного приёма двоичных символов для M-фазных сигналов. Вероятности вычисляются через функцию плотности вероятности и функций распределения вероятностей θ.

Функция распределения вероятности фазы. Чтобы определить вероятность ошибки, мы должны вычислить определенный интеграл.

$$P(\theta_1\le\theta\le\theta_2)=\int_{\theta_1}^{\theta_2}p(\theta)d\theta,$$

где θ_1 и θ_2 – пределы интегрирования, а $p(\theta)$ определяется (C.13). Все последующие вычисления выполняются для вещественных коэффициентов взаимной корреляции μ. Вещественность μ означает, что сигнал имеет симметричный спектр. Такая ситуация обычно встречается. Поскольку комплексная величина μ обуславливает сдвиг ε в ФПВ θ, то есть ε просто слагаемое наклона, результаты, даваемые для вещественной μ можно тривиальным путем изменить, чтобы охватить более общий случай комплексного μ.

При интегрирования $p(\theta)$ рассматривается только область $0\le\theta\le\pi$, поскольку $p(\theta)$ является чётной функцией. Далее, непрерывность интегрируемой функции и её производных и тот факт, что пределы θ_1 и θ_2 не зависят от b позволяют менять местами интегрирование и дифференцирование. Если это выполнить, результирующая интеграл можно вычислить совсем легко, и его можно выразить так:

$$\int_{\theta_1}^{\theta_2}p(\theta)d\theta=\frac{(-1)^{L-1}\left(1-\mu^2\right)^L}{2\pi(L-1)!}\times\frac{\partial^{L-1}}{\partial b^{L-1}}\left\{\frac{1}{b-\mu^2}\left[\frac{\mu\sqrt{1-\left(b/\mu^2-1\right)x^2}}{b^{1/2}}\operatorname{arcctg}x-\right.\right.$$
$$\left.\left.-\operatorname{arcctg}\left(\frac{xb^{1/2}\mu}{\sqrt{1-\left(b/\mu^2-1\right)x^2}}\right)\right]\right\}\Bigg|_{x_1}^{x_2}\Bigg|_{b=1}, \qquad \text{(C.14)}$$

где, по определению,

$$x_i=\frac{-\mu\cos\theta_i}{\sqrt{b-\mu^2\cos\theta_i}},\quad i=1,2. \qquad \text{(C.15)}$$

Вероятность ошибочного приёма символа. Вероятность ошибки символа для M-фазной системы сигналов равна

$$P_M=2\int_{\pi/M}^{\pi}p(\theta)d\theta.$$

Если (C.14) проинтегрировать в этих пределах, то результат равен

$$P_M = \frac{(-1)^{L-1}(1-\mu^2)^L}{\pi(L-1)!}\frac{\partial^{L-1}}{\partial b^{L-1}}\left\{\frac{1}{b-\mu^2}\left[\frac{\pi}{M}(M-1)-\right.\right.$$
$$\left.\left.-\frac{\mu\sin(\pi/M)}{\sqrt{b-\mu^2\cos^2(\pi/M)}}\operatorname{arcctg}\left(\frac{-\mu\cos(\pi/M)}{\sqrt{b-\mu^2\cos^2(\pi/M)}}\right)\right]\right\}\Bigg|_{b=1}. \qquad \text{(C.16)}$$

Вероятность ошибки двоичных символов. Сначала рассмотрим двухфазовых сигналов. В этом случае вероятность ошибки двоичных символов получается интегрированием ФПВ $p(\theta)$ в области $\frac{1}{2}\pi < 0 < 3\pi$. Поскольку $p(\theta)$ – чётная функция, а сигналы априорно равновероятны, эту вероятность можно выразить так:

$$P_2 = 2\int_{\pi/2}^{\pi} p(\theta)d\theta .$$

Легко показать, что $\theta_1 = \pi/2$ предполагает $x_1 = 0$ и $\theta_2 = \pi$ предполагает $x_2 = \mu/\sqrt{b-\mu^2}$. Таким образом,

$$P_2 = \frac{(-1)^{L-1}(1-\mu^2)^L}{2(L-1)!}\frac{\partial^{L-1}}{\partial b^{L-1}}\left[\frac{1}{b-\mu^2}-\frac{\mu}{b^{1/2}(b-\mu^2)}\right]\Bigg|_{b=1}. \qquad \text{(C.17)}$$

После выполнения дифференцирования, указанного в (C.17), и вычисления результата при $b=1$, вероятность ошибки для двоичного символа получается в виде

$$P_2 = \frac{1}{2}\left[1-\mu\sum_{k=1}^{L-1}\binom{2k}{k}\left(\frac{1-\mu^2}{4}\right)^k\right]. \qquad \text{(C.18)}$$

Далее мы рассмотрим случай четырех фазовых сигналов с использованием кода Грея для отображения пары двоичных символов в определенную фазу. Снова предположив, что передаваемый сигнал $s_{11}(t)$, становится ясно, что одиночная ошибка совершается, когда принимаемая фаза $\frac{1}{4}\pi < 0 < \frac{1}{4}\pi$, а двойная ошибка совершается, когда принимаемая фаза $\frac{3}{4}\pi < 0 < \pi$. Это значит, что вероятность ошибки на бит для четырехфазовой системы символов

$$P_{4b} = \int_{\pi/4}^{3\pi/4} p(0)d0 + 2\int_{3\pi/4}^{\pi} p(0)d0 . \qquad \text{(C.19)}$$

Легко получить с учётом (C.14) и (C.19), что

$$P_{4b} = \frac{(-1)^{L-1}(1-\mu^2)^L}{2(L-1)!}\frac{\partial^{L-1}}{\partial b^{L-1}}\left[\frac{1}{b-\mu^2}-\frac{\mu}{(b-\mu^2)(2b-\mu^2)^{1/2}}\right]\Bigg|_{b=1}.$$

Таким образом, окончательно для вероятности ошибки на бит для четырёхфазовых сигналов

$$P_{4b} = \frac{1}{2}\left[1-\frac{\mu}{\sqrt{2-\mu^2}}\sum_{k=0}^{L-1}\binom{2k}{k}\left(\frac{1+\mu^2}{4-2\mu^2}\right)^k\right]. \qquad \text{(C.20)}$$

Заметим, что если ввести замену $\rho = \mu/\sqrt{2-\mu^2}$, выражение для P_{4b} можно выразить через ρ так:

$$P_{4b} = \frac{1}{2}\left[1-\rho\sum_{k=0}^{L-1}\binom{2k}{k}\left(\frac{1-\rho^2}{4}\right)^k\right]. \qquad \text{(C.21)}$$

Другими словами, P_{4b} имеет тот же вид что P_2, определяемое (C.18). Далее заметим, что ρ подобно μ можно интерпретировать как коэффициент взаимной корреляции, так как при $0 \le \mu \le 1$ область определения $0 \le \rho \le 1$. Этот простой факт будет использован в разделе С.4.

Вышеизложенная процедура получения вероятности ошибки на бит для M-фазовых сигналов с кодом Грея можно использовать для получения результатов при $M = 8$, 16 и так далее, как показано Прокисом (1968).

Вычисление коэффициентов взаимной корреляции. Выражение для вероятности ошибок данных выше зависит от единственного параметра, именно, коэффициентов взаимной корреляции μ. «ясновидящая» оценка дана (C.5) а выход согласованного фильтра, когда передается сигнал $s_{11}(t)$, равен $N_k = 2\mathcal{E}g_k + N_k$. Следовательно, коэффициент взаимной корреляции равен

$$\mu = \frac{\sqrt{\nu}}{\sqrt{\left(\bar{\gamma}_c^{-1}+1\right)\left(\bar{\gamma}_c^{-1}+\nu\right)}}, \tag{C.22}$$

где, по определению,

$$\nu = \left|\sum_{i=1}^{\infty} c_i\right|^2 \Big/ \sum_{i=1}^{\infty} |c_i|^2; \qquad \bar{\gamma}_c = \frac{\mathscr{E}}{N_0} E\left(|g_k|^2\right), \quad k = 1,2,\ldots,L. \tag{C.23}$$

Параметр ν представляет эффективное число сигнальных интервалов, на которых формируется оценка, а $\bar{\gamma}_c$ – это среднее ОСШ на канале.

Для случая дифференциальной ФМ взвешивающие коэффициенты раны $c_1 = 1$, $c_i = 0$ для $i \neq 1$. Следовательно, $\nu = 1$ и $\mu = \bar{\gamma}_c/(1+\bar{\gamma}_c)$.

Когда $\nu = \infty$, совершенная оценка равна

$$\lim_{\nu\to\infty} \mu = \sqrt{\frac{\bar{\gamma}_c}{\bar{\gamma}_c+1}}.$$

Таблица С.1. Каиал с релеевскими замираниями

Тип оценки	**Коэффициент взаимной корреляции** μ
«ясновидящая» оценка	$\dfrac{\sqrt{\nu}}{\sqrt{\left(\bar{\gamma}_c^{-1}+1\right)\left(\bar{\gamma}_c^{-1}+\nu\right)}}$
Оценка по пилот-сигналу	$\dfrac{\sqrt{r\nu}}{(r+1)\sqrt{\left(\dfrac{1}{\bar{\gamma}_l}+\dfrac{r}{r+1}\right)\left(\dfrac{1}{\bar{\gamma}_l}+\dfrac{\nu}{r+1}\right)}}$
Дифференциальная ФМ	$\dfrac{\bar{\gamma}_c}{\bar{\gamma}_c+1}$
Совершенная оценка	$\sqrt{\dfrac{\bar{\gamma}_c}{\bar{\gamma}_c+1}}$

Наконец, для случая оценки на основе пилот-сигнала, даваемой (С.4), коэффициент взаимной корреляции равен

$$\mu = \left[\left(1+\frac{r+1}{r\bar{\gamma}_l}\right)\left(1+\frac{r+1}{\nu\bar{\gamma}_l}\right)\right]^{-1/2}, \tag{C.24}$$

где, по определению,

$$\bar{\gamma}_l = \frac{\mathscr{E}_l}{N_0} E\left(|g_k|^2\right), \quad \mathscr{E}_l = \mathscr{E} + \mathscr{E}_p, \quad r = \mathscr{E}/\mathscr{E}_p.$$

Величина μ, определённая выше, приведена в таблице С.1.

С.4. ВЕРОЯТНОСТЬ ОШИБКИ ДЛЯ НЕИЗМЕННЫХ ВО ВРЕМЕНИ КАНАЛОВ И КАНАЛОВ С РАЙСОВСКИМИ ЗАМИРАНИЯМИ

В разделе С.2 комплексные коэффициенты ослабления канала $\{g_k\}$ характеризовались как гауссовские случайные величины с нулевым средним, что соответствует каналам с релеевскими замираниями. В этом разделе коэффициенты ослабления канала $\{g_k\}$ предполагаются гауссовскими случайными величинами с не нулевыми средними. Оценки коэффициентов ослабления канала формируются демодулятором, и они используются так, как описано в разделе С.1. Более того, величины для решения 0 опять-таки определяются (С.2). Однако в рассматриваемом случае гауссовские случайные величины X_k и Y_k, которые определяют выходы согласных фильтров и оценок, соответственно, для k-го канала, имеют ненулевые средние, обозначаемые $\bar{X}_k$ и $\bar{Y}_k$. Далее, вторые моменты равны

$$\left.\begin{aligned} m_{xx} &= E\left(\left|X_k - \bar{X}_k\right|^2\right) \\ m_{yy} &= E\left(\left|Y_k - \bar{Y}_k\right|^2\right) \\ m_{xy} &= E\left(\left|\left(X_k - \bar{X}_k\right)\left(Y_k^* - \bar{Y}_k^*\right)\right|^2\right) \end{aligned}\right\} \text{одинаковы для всех каналов,}$$

а нормированная ковариация равна

$$\mu = \frac{m_{xy}}{\sqrt{m_{xx} m_{yy}}}.$$

Для этой модели канала ниже даны вероятности ошибки только для двух- и четырехфазовых сигналов. Мы интересуемся частным случаем, когда флуктуирующие компоненты каждого из канальных ослаблений $\{g_k\}$ отсутствуют, так что каналы неизменны во времени. Если дополнительно к неизменности во времени параметров канала шумы оценки и выхода согласованного фильтра не коррелированны, то $\mu = 0$.

В общем случае вероятность ошибки передачи двухфазовых сигналов по статистически независимым L каналам, характеризуемым так, как описано выше, можно получить из результатов *приложения* В. В наиболее общем виде выражение для вероятности ошибки двоичной системы символов

$$\begin{aligned} P_2 &= Q_1(a,b) - I_0(a)\exp\left[-\tfrac{1}{2}\left(a^2+b^2\right)\right] + \\ &+ \frac{I_0(ab)\exp\left[-\frac{1}{2}\left(a^2+b^2\right)\right]}{\left[2/(1-\mu)\right]^{2L-1}} \sum_{k=0}^{L-1} \binom{2L-1}{k} \left(\frac{1+\mu}{1-\mu}\right)^k + \frac{\exp\left[-\frac{1}{2}\left(a^2+b^2\right)\right]}{\left[2/(1-\mu)\right]^{2L-1}} \times \\ &\times \sum_{n=1}^{L-1} I_n(ab) \sum_{k=0}^{L-1-n} \binom{2L-1}{k} \left[\left(\frac{b}{a}\right)^n \left(\frac{1+\mu}{1-\mu}\right)^k - \left(\frac{a}{b}\right)^n \left(\frac{1+\mu}{1-\mu}\right)^{2L-1-k}\right] \quad (L \geq 2), \end{aligned} \tag{C.25}$$

$$P_2 = Q_1(a,b) - \tfrac{1}{2}(1+\mu) I_0(ab) \exp\left[-\tfrac{1}{2}\left(a^2+b^2\right)\right] \qquad (L=1),$$

где, по определению,

$$a = \left\{\frac{1}{2}\sum_{k=1}^{L}\left|\frac{\bar{X}_k}{\sqrt{m_{xx}}} - \frac{\bar{Y}_k}{\sqrt{m_{yy}}}\right|^2\right\}^{1/2}, \quad b = \left\{\frac{1}{2}\sum_{k=1}^{L}\left|\frac{\bar{X}_k}{\sqrt{m_{xx}}} + \frac{\bar{Y}_k}{\sqrt{m_{yy}}}\right|^2\right\}^{1/2} \tag{C.26}$$

$$Q_1(a,b) = \int_b^\infty x \exp\left[-\tfrac{1}{2}\left(a^2+x^2\right)\right] I_0(ax)\,dx.$$

$I_n(x)$ – модифицированная функция Бесселя первого ряда порядка n.

Определим константы a и b, когда канал неизменен во времени, $\mu = 0$, а оценка ослабления канала и фазы даны в разделе С.1. Напомним, что когда передается $s_1(t)$, выход согласованного фильтра $X_k = 2\mathscr{E} g_k + N_k$. «ясновидящая» оценка дана (С.5). Следовательно, для этой оценки моменты равны $\bar{X}_k = 2\mathscr{E} g_k$, $\bar{Y}_k = g_k$, $m_{xx} = 4\mathscr{E} N_0$ и $m_{yy} = N_0/\mathscr{E}v$, где $\mathscr{E}$ – энергия сигнал, N_0 – значение спектральной плотности шума, а v определено (С.23). Подстановка этих моментов в (С.26) дает следующее выражение для a и b:

$$a = \sqrt{\tfrac{1}{2}\gamma_b}\left|\sqrt{v} - 1\right|, \quad b = \sqrt{\tfrac{1}{2}\gamma_b}\left|\sqrt{v} + 1\right|, \quad \gamma_b = \frac{\mathscr{E}}{N_0}\sum_{k=1}^{L}\left|g_k\right|^2. \tag{C.27}$$

Этот результат первоначально получен Прайсом (1962).

Вероятность ошибки для диффсренциальной ФМ можно получить, положив $v = 1$ в (С.27).

Далее рассмотрим оценку по пилот сигналу. В этом случае оценка дается (С.4), а выход согласованного фильтра снова $X_k = 2\mathscr{E} g_k + N_k$. Если вычислить моменты и их подставить в (С.26) получаются следующие выражения для a и b:

$$a = \sqrt{\frac{\gamma_l}{2}}\left|\sqrt{\frac{v}{r+1}} - \sqrt{\frac{r}{r+1}}\right|, \quad b = \sqrt{\frac{\gamma_l}{2}}\left|\sqrt{\frac{v}{r+1}} + \sqrt{\frac{r}{r+1}}\right| \tag{C.28}$$

где

$$\gamma_l = \frac{\mathscr{E}_l}{N_0}\sum_{k=1}^{L}\left|g_k\right|^2, \quad \mathscr{E}_l = \mathscr{E} + \mathscr{E}_p, \quad r = \mathscr{E}/\mathscr{E}_p.$$

Наконец, рассмотрим вероятность ошибки при передаче четырехфазовой системы сигналов по неизменному во времени каналу, при условии $\mu = 0$. Один из подходов, который можно использовать для определения вероятности ошибки сводится к определению ФПВ θ и затем к ее интегрированию по соответствующей области значений θ. К сожалению, такой подход математически трудно осуществить. Вместо этого можно использовать простой, хотя и обходной метод, включающий преобразование Лапласа. Интеграл (14.4.14), который связывает вероятность ошибки $P_2(\gamma_b)$ в канале с АБГШ с вероятностью ошибки P_{4b} в канале с релеевскими замираниями, является преобразованием Лапласа. Поскольку вероятность ошибки на бит P_2 и P_{4b} для канала с релеевским замираниями, определяемые (С.18) и (С.21) соответственно имеет туже форму, но отличается только коэффициентом корреляции, то следует, что вероятность ошибки на бит для неизмененного во времени канала также имеет туже форму. Это значит, что (С.25) с $\mu = 0$ является также выражением для вероятности ошибки на бит для четырех фазовой системы сигналов с модифицированными параметрами a и b, отражающими разницу коэффициентов корреляции. Дальнейшие исследования можно найти в статье Прокиса (1968). Выражение для a и в даны в таблице С.2.

Таблица С.2. Канал, неизменный во времени

Тип оценки	a	b
	Двоичная ФМ	
«ясновидящая» оценка	$\sqrt{\frac{1}{2}\gamma_b}\left\|\sqrt{\nu}-1\right\|$	$\sqrt{\frac{1}{2}\gamma_b}\left(\sqrt{\nu}+1\right)$
Дифференциальная ФМ	0	$\sqrt{2\gamma_b}$
Оценка по пилот-сигналу	$\sqrt{\frac{\gamma_I}{2}}\left\|\sqrt{\frac{\nu}{r+1}}-\sqrt{\frac{r}{r+1}}\right\|$	$\sqrt{\frac{\gamma_I}{2}}\left(\sqrt{\frac{\nu}{r+1}}+\sqrt{\frac{r}{r+1}}\right)$
	4-позиционная ФМ	
«Ясновидящая» оценка	$\sqrt{\frac{1}{2}\gamma_b}\left\|\sqrt{\nu+1+\sqrt{\nu^2+1}}-\sqrt{\nu+1-\sqrt{\nu^2+1}}\right\|$	$\sqrt{\frac{1}{2}\gamma_b}\left(\sqrt{\nu+1+\sqrt{\nu^2+1}}+\sqrt{\nu+1-\sqrt{\nu^2+1}}\right)$
Дифференциальная ФМ	$\sqrt{\frac{1}{2}\gamma_b}\left\|\sqrt{2+\sqrt{2}}-\sqrt{2-\sqrt{2}}\right\|$	$\sqrt{\frac{1}{2}\gamma_b}\left\|\sqrt{2+\sqrt{2}}+\sqrt{2-\sqrt{2}}\right\|$
Оценка по пилот-сигналу	$\sqrt{\frac{\gamma_I}{4(r+1)}}\left(\sqrt{\nu+r+\sqrt{\nu^2+r^2}}-\sqrt{\nu+r-\sqrt{\nu^2+r^2}}\right)$	$\sqrt{\frac{\gamma_I}{4(r+1)}}\left(\sqrt{\nu+r+\sqrt{\nu^2+r^2}}+\sqrt{\nu+r-\sqrt{\nu^2+r^2}}\right)$

ПРИЛОЖЕНИЕ D

ФАКТОРИЗАЦИЯ МАТРИЦЫ (ПРИ РЕШЕНИИ СИСТЕМЫ ЛИНЕЙНЫХ УРАВНЕНИЙ)

Рассмотрим решение системы линейных уравнений

$$\mathbf{R}_N\mathbf{C}_N = \mathbf{U}_N, \tag{D.1}$$

где $\mathbf{R}_N - N \times N$ положительно определённая симметричная матрица, $\mathbf{C}_N - N$-мерный вектор коэффициентов, которых надо определить, а $\mathbf{U}_N -$ произвольный N-мерный вектор. Уравнение (D.1) можно эффективно разрешить путём факторизации $\mathbf{R}_N$ в виде произведения

$$\mathbf{R}_N = \mathbf{S}_N \mathbf{D}_N \mathbf{S}_N^{\mathrm{T}}, \tag{D.2}$$

где $\mathbf{S}_N$ – нижняя треугольная матрица с элементами $\{s_{ik}\}$, а $\mathbf{D}_N$ – диагональная матрица с диагональными элементами $\{d_k\}$. Диагональные элементы $\mathbf{S}_N$ образуют ряд единиц, то есть $s_{ii} = 1$. Тогда имеем

$$\begin{aligned} r_{ij} &= \sum_{k=1}^{i} s_{jk} d_k s_{ik}, \ 1 \le j \le i-1, \ i \ge 2, \\ r_{11} &= d_1, \end{aligned} \tag{D.3}$$

где $\{r_{ij}\}$ – элементы $\mathbf{R}_N$. Следовательно, элементы $\{s_{ik}\}$ и $\{d_k\}$ определяются (D.3) согласно правилам

$$\begin{aligned} d_1 &= r_{11} \\ s_{ij} d_j &= r_{ij} - \sum_{k=1}^{i-1} s_{ik} d_k s_{jk}, \quad 1 \le j \le i-1, \quad 2 \le i \le N \\ d_i &= r_{ii} - \sum_{k=1}^{i-1} s_{ik}^2 d_k, \quad 2 \le i \le N. \end{aligned} \tag{D.4}$$

(D.4) определяет $\mathbf{S}_N$ и $\mathbf{D}_N$ через элементы $\mathbf{R}_N$.

Решение (D.1) осуществляется в два этапа. При подстановке (D.2) в (D.1) имеем

$$\mathbf{S}_N \mathbf{D}_N \mathbf{S}_N^{\mathrm{T}} \mathbf{C}_N = \mathbf{U}_N.$$

Пусть

$$\mathbf{Y}_N = \mathbf{D}_N \mathbf{S}_N^{\mathrm{T}} \mathbf{C}_N. \tag{D.5}$$

Тогда

$$\mathbf{S}_N \mathbf{Y}_N = \mathbf{U}_N. \tag{D.6}$$

Сначала решим (D.6) для $\mathbf{Y}_N$. С учётом треугольной формы $\mathbf{S}_N$ имеем

$$\begin{aligned} y_1 &= u_1, \\ y_i &= u_i - \sum_{j=1}^{i-1} s_{ij} y_j, \quad 2 \le i \le N. \end{aligned} \tag{D.7}$$

Получив $\mathbf{Y}_N$, на втором этапе определяем $\mathbf{C}_N$. Имеем

$$\begin{aligned} \mathbf{D}_N \mathbf{S}_N^{\mathrm{T}} \mathbf{C}_N &= \mathbf{Y}_N, \\ \mathbf{S}_N^{\mathrm{T}} \mathbf{C}_N &= \mathbf{D}_N^{-1} \mathbf{Y}_N. \end{aligned}$$

Начинаем с

$$c_N = y_N / d_N, \tag{D.8}$$

а оставшиеся коэффициенты $\mathbf{C}_N$ получаются рекуррентно:

$$c_i = \frac{y_i}{d_i} - \sum_{j=i-i}^{N} s_{ij} c_j, \quad 1 \le i \le N-1. \tag{D.9}$$

Число умножений и делений, требуемых для формирования факторизации $\mathbf{R}_N$, пропорционально N^N. Число умножений и делений, требуемых для вычисления $\mathbf{C}_N$, когда $\mathbf{S}_N$ определено, пропорционально N^2. Если $\mathbf{R}_N$ – матрица Тёплица, алгоритм Левинсона-Дурбина эффективно используется для нахождения решения (D.1), так как число умножений и делении пропорционально N^2. С другой стороны, непосредственно методом наименьших квадратов (НК) $\mathbf{S}_N$ и $\mathbf{D}_N$ не вычисляются, как в (D.3), но они поддаются рекуррентному обновлению. Обновление выполнятся N^2 операциями (умножений и делений). Затем решение для вектора $\mathbf{C}_N$ производится шагами (D.5)–(D.9). Как следствие, вычислительные затраты для рекуррентной НК процедуры пропорционально N^2.

ССЫЛКИ И БИБЛИОГРАФИЯ

Abend K. and Fritchman B.D. (1970) "Statistical Detection for Communication Channels with Intersymbol Interference", Proc. IEEE. P. 779–785. May.

Abramson N. (1963) Information Theory and Coding. McGraw-Hill. New York.

Abramson N. (1970) "The ALOHA System – Another Alternative for Computer Communications", 1970. Fall Joint Comput. Conf. AFIDS Conf. Proc. vol. 37. P. 281–285. AFIPS Press. Montvale. N.J.

Abramson N. (1977) "The Throughput of Packet Broadcasting Channels", IEEE Trans. Commun., vol. COM–25. P. 117–128. January.

Abramson N. (1993) Multiple Access Communications. IEEE Press, New York.

Adler R.L., Coppersmith D. and Hassner M. (1983). "Algorithms for Sliding Block Codes". IEEE Trans. Inform. Theory. vol. IT–29. P. 5–22, January.

Al-Hussaini E. and Al-Bassiouni (1985). "Performance of MRC Diversity Systems for the Detection of Signals with Nakagami Fading", IEEE Trans. Commun. vol. COM–33. P. 1315–1319. December.

Altekar S.A. and Beaulieu N.C. (1993) "Upper Bounds on the Error Probability of Decision Feedback Equalization", IEEE Trans. Inform. Theory. vol. IT–39. P. 145–156. January.

Anderberg M.R. (1973). Cluster Analysis for Applications. Academic, New York.

Anderson J.B., Aulin T. and Sundberg C.W. (1986) Digital Phase Modulation. Plenum, New York.

Anderson R.R. and Salz J. (1965). "Spectra of Digital FM", Bell Syst. Tech. J. vol. 44. P. 1165–1189. July – August.

Ash R.B. (1965) Information Theory, Interscience, New York.

Aulin T. (1980) "Viterbi Detection of Continuous Phase Modulated Signals", Nat. Telecommun. Conf. Record. P. 14.2.1–14.2.7, Houston Texas, November.

Aulin T., Rydbeck N. and Sundberg C.W. (1981) "Continuous Phase Modulation – Part II: Partial Response Signaling", IEEE Trans. Commun. vol. COM–29. P. 210–225, March.

Aulin T. and Sundberg C.W. (1981) "Continuous Phase Modulation – Part I: Full Response Signaling", IEEE Trans. Commun. vol. COM–29. P. 196–209, March.

Aulin T. and Sundberg C.W. (1982a) "On the Minimum Euclidean Distance for a Class of Signal Space Codes", IEEE Trans. Inform Theory. vol. IT–28. P. 43–55, January.

Aulin T. and Sundberg C.W. (1982b) "Minimum Euclidean Distance and Power Spectrum for a Class of Smoothed Phase Modulation Codes with Constant Envelope" IEEE Trans. Commun. vol. COM–30. P. 1721–1729, July.

Aulin T. and Sundberg C.W. (1984) "CPM – An Efficient Constant Amplitude Modulation Scheme" Int. J. Satellite Commun. vol. 2. P. 161–186.

Austin M.E. (1967). "Decision-Feedback Equalization for Digital Communication Over Dispersive Channels". MIT Lincoln Laboratory, Lexington. Mass., Tech. Report No 437, August.

Barrow B. (1963) "Diversity Combining of Fading Signals with Unequal Mean Strengths", IEEE Trans. Commun. Syst. vol. CS–11. P. 73–78. March.

Beaulieu N.C. (1990) "An Infinite Series for the Computation of the Complementary Probability Distribution Function of a Sum of Independent Random Variables and Its Application to the Sum of Rayleigh Random Variables". IEEE Trans. Commun. vol. COM–38. P. 1463–1474. September.

Beaulieu N.C. and Abu-Dayya A.A. (1991). "Analysis of Equal Gain Diversity on Nakagami Fading Channels", IEEE Trans. Commun. vol. COM–39. P. 225–234. February.

Bekir N.E. Scholtz R.A. and Welch L.R. (1978). "Partial Period Correlation Properties of PN Sequences". 1978 Nat. Telecommun. Conf. Record. P. 35.1.1–35.1.4. Birmingham, Alabama, November.

Belfiore C.A. and Park J.N., Jr (1979) "Decision Feedback Equalization" Proc. IEEE vol. 67. P. 1143–1156. August.

Bellini J. (1986) "Bussgang Techniques for Blind Equalization", Proc. GLOBECOM'86. P. 46.1.1–46.1.7. Houston, Texas, December.

Bello P.A. and Nelin B.D. (1962a). "Predetection Diversity Combining with Selectivity Fading Channels". IRE Trans. Commun. Syst. vol. CS–10. P. 32–42. March.

Bello P.A. and Nelin B.D. (1962b). "The Influence of Fading Spectrum on the Binary Error Probabilities of Incoherent and Differentially Coherent Matched Filter Receivers". IRE Trans. Commun. Syst., vol. CS–10. P. 160–168. June.

Bello P.A. and Nelin B.D. (1963). "The Effect of Frequency Selective Fading on the Binary Error Probabilities of Incoherent and Differentially Coherent Matched Filter Receivers", IEEE Trans. Commun. Syst., vol. CS–11. P. 170–186. June.

Bennett W.R. and Davey J.R (1965) Data Transmission. McGraw-Hill, New York.

Bennett W.R. and Rice S.O. (1963) "Spectral Density and Autocorrelation Functions Associated with Binary Frequency-Shift Keying", Bell Syst. Tech. J., vol. 42. P. 2355–2385, September.

Benveniste A. and Goursat M. (1984). "Blind Equalizers". IEEE Trans. Commun., vol. COM–32. P. 871–883. August.

Berger T. (1971). Rate Distortion Theory. Prentice-Hall. Englewood Cliffs, N.J.

Berger T. and Tufts D.W. (1967). "Optimum Pulse Amplitude Modulation. Part I: Transmitter-Receiver Design and Bounds from Information Theory" IEEE Trans. Inform. Theory, vol. IT–13. P. 196–208.

Bergmans P.P. and Cover T.M. (1974). "Cooperative Broadcasting", IEEE Trans. Inform. Theory. vol. IT–20. P. 317–324. May.

Berlekamp E.R. (1968). Algebraic Coding Theory. McGraw-Hill, New York.

Berlekamp E.R. (1973). "Goppa Codes", IEEE Trans. Inform. Theory, vol. IT–19. P. 590–592.

Berlekamp E.R. (1974). Key Papers in the Development of Coding Theory. IEEE Press, New York.

Bierman G.J. (1977). Factorization Methods for Discrete Sequential Estimation. Academic, New York.

Biglieri E., Divsalar D., McLane P.J. and Simon M.K. (1991). Introduction to Trellis-Coded Modulation with Application. Macmillan, New York.

Bingham J.A.C. (1990). "Multicarrier Modulation for Data Transmission: An Idea Whose Time Has Come", IEEE Commun. Mag., vol. 28. P. 5–14, May.

Blahut R.E. (1983). Theory and Practice of Error Control Codes, Addison-Wesley, Reading, Mass.

Blahut R.E. (1987). Principles and Practice of Information Theory, Addison-Wesley, Reading. Mass.

Bose R.C. and Ray-Chaudhuri D.K. (1960a). "On a Class of Error Correcting Binary Group Codes", Inform. Control, vol. 3. P. 68–79, March.

Bose R.C. and Ray-Chaudhuri D.K. (1960b). "Further Results in Error Correcting Binary Group Codes", Inform. Control, vol. 3. P. 279–290, September.

Brennan D.G. (1959). "Linear Diversity Combining Techniques", Proc. IRE, vol. 47. P. 1075–1102.

Busgang J.J. (1952). "Crosscorrelation Functions of Amplitude-Distorted Gaussian Signals", MIT RLE Tech. Report 216.

Bucher E.A. (1980). "Coding Options for Efficient Communications on Non-Stationary Channels", Rec. IEEE Int. Conf. Commun. P. 4.1.1–4.1.7.

Burton H.O. (1969). "A Class of Asymptotically Optimal Burst Correcting Block Codes", Proc. ICCC, Boulder, Col., June.

Buzo A., Gray A.H., Jr., Gray R.M. and Markel J.D. (1980). "Speech Coding Based Upon Vector Quantization", IEEE Trans. Acoust., Speech. Signal Processing, vol. ASSP-28, P. 562–574, October.

Cahn C.R. (1960). "Combined Digital Phase and Amplitude Modulation Communication Systems", IRE Trans. Commun. Syst., vol. CS–8. P. 150–155, September.

Calderbank A.R. and Sloane N.J.A. (1987). "New Trellis Codes Based on Lattices and Cosets", IEEE Trans. Inform. Theory vol. IT–33. P. 177–195, March.

Campanella S.J. and Robinson G.S. (1971). "A Comparison of Orthogonal Transformations for Digital Speech Processing", IEEE Trans. Commun., vol. COM–19. P. 1045–1049, December.

Campopiano C.N. and Glazer B.G. (1962). "A Coherent Digital Amplitude and Phase Modulation Scheme", IRE Trans. Commun. Syst., vol. CS–10. P. 90–95, June.

Capetanakis J.I. (1979). "Tree Algorithms for Packet Broadcast Channels", IEEE Trans. Inform. Theory. vol. IT–25. P. 505–515. September.

Caraiscos C. and Liu B. (1984). "A Round off Error Analysis of the LMS Adaptive Algorithm", IEEE Trans. Acoust., Speech, Signal Processing. vol. ASSP–32. P. 34–41. January.

Carayannis G., Manolakis D.G. and Kalouptsidis N. (1983). "A Fast Sequential Algorithm for Least-Squares Filtering and Prediction", IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP–31. P. 1394–1402, December.

Carayannis G., Manolakis D.G. and Kalouptsidis N. (1986). "A Unified View of Parametric Processing Algorithms for Prewindowed Signals", Signal Processing, vol. 10. P. 335–368. June.

Carleial A.B. and Hellman M.E. (1975). "Bistable Behavior of ALOHA-Type Systems", IEEE Trans. Commun., vol. COM–23. P. 401–410. April 1975.

Carlson A.B. (1975). Communication Systems, McGraw-Hill, New York.

Chang D.Y., Gersho A., Ramamurthi B. and Shohan Y. (1984). "Fast Search Algorithms for Vector Quantization and Pattern Matching", Proc. IEEE Int. Conf. Acoust. Speech, Signal Processing, paper 9.11, San Diego, Calif., March.

Chang R.W. (1966). "Synthesis of Band-Limited Orthogonal Signals for Multichannel Data Transmission", Bell Syst. Tech. J. vol. 45. P. 1775–1796, December.

Charash U. (1979). "Reception Trough Nakagami Fading Multipath Channels with Random Delays", IEEE Trans. Commun. vol. COM–27. P. 657–670, April.

Chase D. (1972). "A Class of Algorithms for Decoding Block Codes with Channel Measurement Information" IEEE Trans. Inform. Theory, vol. IT–18. P. 170–182, January.

Chase D. (1976). "Digital Signal Design Concepts for a Time-Varying Ricean Channel" IEEE Trans. Commun., vol. COM–24. P. 164–172, February.

Chien R.T. (1964). "Cyclic Decoding Procedures for BCH Codes", IEEE Trans. Inform. Theory, vol. IT–10. P. 357–363, October.

Chow J.S. Tu J.C. and Cioffi J.M. (1991). "A Discrete Multitone Transceiver System for HDSL Applications", IEEE J. Selected Areas Commun., vol. SAC–9. P. 895–908, August.

Chyi G.T., Proakis J.G. and Keller C.M. (1988). "Diversity Selection/Combining Schemes with Excess Noise-Only Diversity Reception over a Rayleigh-Fading Multipath Channel", Proc. Conf. Inform. Sci. Syst., Princeton University, Princeton, N.J. March.

Cioffi J.M. and Kailath T. (1984). "Fast Recursive-Least Squares Transversal Filters for Adaptive Filtering", IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP–32. P. 304–337, April.

Cook C.E., Ellersick F.W., Milstein L.B. and Schilling D.L. (1983). Spread Spectrum Communication, IEEE Press, New York.

Costas J.P. (1956). "Synchronous Communications". Proc. IRE, vol. 44. P. 1713–1718, December.

Cover T.M. (1972). "Broadcast Channels", IEEE Trans. Inform. Theory, vol. IT–18. P. 2–14, January.

Cramer H. (1946). Mathematical Methods of Statistics, Princeton University Press, Princeton, N.J.

Daut D.G. Modestino J.W. and Wismer L.D. (1982). "New Short Constraint Length Convolutional Code Construction for Selected Rational Rates", IEEE Trans. Inform. Theory, vol. IT–28. P. 793–799, September.

Davenport W.B., Jr. (1970). Probability and Random Processes, McGraw-Hill, New York.

Davenport W.B., Jr. and Root W.L. (1958). Random Signals and Noise. McGraw-Hill, New York.

Davisson L.D. (1973). "Universal Noiseless Coding", IEEE Trans. Inform. Theory, vol. IT–19. P. 783–795.

Davisson L.D., McEliece R.J., Pursley M.B. and Wallace M.S. (1981). "Efficient Universal Noiseless Source Codes", IEEE Trans. Inform. Theory, vol. IT–27. P. 269–279.

DeBuda R. (1972). "Coherent Demodulation of Frequency Shift Keying with Low Deviation Ratio", IEEE Trans. Commun. vol. COM–20. P. 429–435. June.

Deller J.P., Proakis J.G. and Hansen H.L. (1993). Discrete-Time Processing of Speech Signals, MacMillan, New York.

Ding Z. (1990). Application Aspects of Blind Adaptive Equalizers in QAM Data Communications, Ph. D. Thesis, Department of Electrical Engineering, Cornell University.

Ding Z., Kennedy R.A. Anderson B.D. O. and Johnson C.R. (1989). "Existence and Avordance of Ill-Convergence of Godard Blind Equalizers in Data Communication Systems", Proc. 23rd Conf. On Inform. Sci. Systems. Baltimore, Md.

Divsalar D., Simon M.K. and Yuen J.H. (1987). "Trellis Coding with Asymmetric Modulation", IEEE Trans. Commun., vol. COM–35. P. 130–141, February.

Divsalar D. and Yuen J.H. (1984). "Asymmetric MPSK for Trellis Codes", Proc. GLOBECOM'84. P. 20.6.1–20.6.8, Atlanta, Georgia, November.

Dixon R.C. (1976). Spread Spectrum Techniques. IEEE Press, New York.

Doelz M.L. Heald E.T. and Martin D.L. (1957). "Binary Data Transmission Techniques for Linear Systems", Proc. IRE, vol. 45. P. 656–661. May.

Drouilhet P.R., Jr. and Bernstein S.L. (1969). "TATS – A Bandspread Modulation-Demodulation System for Multiple Access Tactical Satellite Communication", 1969 IEEE Electronics and Aerospace Systems (EASCON) Conv. Record, Washington, D. C. P. 126–132, October 27–29.

Duffy E.P. and Tratcher T.W. (1971). "Analog Transmission Performance on the Switched Telecommunications Network", Bell Syst. Tech. J., vol. 50. P. 1311–1347. April.

Durbin J. (1959). "Efficient Estimation of Parameters in Moving-Average Models", Biometrika, vol. 46, parts 1 and 2. P. 306–316.

Duttweiler D.L., Mazo J.E. and Messerschmitt D.G. (1974). "Error Propagation in Decision-Feedback Equalizers", IEEE Trans. Inform. Theory, vol. IT–20. P. 490–497, July.

Eleftheriou E. and Falconer D.D. (1987). Adaptive Equalization Techniques for HF Channels", IEEE J. Selected Areas Commun., vol. SAC–5. P. 238–247, February.

El Gamal A. and Cover T.M. (1980). "Multiple User Information Theory, " Proc. IEEE, vol. 68. P. 1466–1483, December.

Elias P. (1954). "Error Free Coding", IRE Trans. Inform. Theory, vol. IT–4. P. 29–37, September.

Elias P. (1955). "Coding for Noisy Channels", IRE Convention Record, vol. 3, part 4. P. 37–46.

Esposito R. (1967). "Error Probabilities for the Nakagami Channel", IRE Trans. Inform. Theory, vol. IT–13. P. 145–148. January.

Eyuboglu V.M. (1988). "Detection of Coded Modulation Signals on Linear, Severely Distorted Channels Using Dicision-Feedback Noise Prediction with Interleaving", IEEE Trans. Commun., vol. COM–36. P. 401–409, April.

Falconer D.D. (1976). "Jointly Adaptive Equalization and Carrier Recovery in Two-Dimensional Digital Communication Systems", Bell Syst. Tech. J., vol. 55. P. 317–334, March.

Falconer D.D. and Ljung L. (1978). "Application of Fast Kalman Estimation to Adaptive Equalization", IEEE Trans. Commun. vol. COM–26. P. 1439–1446, October.

Falconer D.D. and Salz J. (1977). "Optimal Reception of Digital Data Over the Gaussian Channel with Unknown Delay and Phase Jitter", IEEE Trans. Inform. Theory, vol. IT–23. P. 117–126, January.

Fano R.M. (1961). Transmission of Information. MIT Press. Cambridge, Mass.

Fano R.M. (1963). "A Heuristic Discussion of Probabilistic Coding". IEEE Trans. Inform. Theory vol. IT–19. P. 64–74, April.

Feinstein A. (1958). Foundations of Information Theory. McGraw-Hill, New-York.

Fire P. (1959). "A Class of Multiple-Error-Correcting Binary Codes for Non-Independent Errors", Sylvania Report No. RSL-E-32. Sylvania Electronic Defense Laboratory. Mountain View, Calif., March.

Flanagan J.L. et al. (1979). "Speech Coding", IEEE Trans. Commun. vol. COM–27. P. 710–736. April.

Forney G.D. Jr. (1965). "One Decoding BCH Codes", IEEE Trans. Inform. Theory vol. IT–11. P. 549–557, October.

Forney G.D. Jr. (1966a). Concatenated Codes. MIT Press, Cambridge, Mass.

Forney G.D. Jr. (1966b). "Generalized Minimum Distance Decoding", IEEE Trans. Inform. Theory vol. IT–12. P. 125–131. April.

Forney G.D. Jr. (1968). "Exponential Error Bounds for Erasure, List and Decision-Feedback Schemes", IEEE Trans. Inform. Theory vol. IT–14. P. 206–220, March.

Forney G.D. Jr. (1970a). "Coding and Its Application in Space Communications", IEEE Spectrum., vol. 7. P. 47–58, June.

Forney G.D. Jr. (1970b)", Convolutional Codes I: Algebraic Structure", IEEE Trans. Inform. Theory vol. IT–16. P. 720–738. November.

Forney G.D. Jr. (1971). "Burst Correcting Codes for the Classic Bursty Channel". IEEE Trans. Commun. Tech. vol. COM–19. P. 772–781, October.

Forney G.D. Jr. (1972). "Maximum-Likelihood Sequence Estimation of Digital Sequences in the Presence of Intersymbol Interference", IEEE Trans. Inform. Theory vol. IT–18. P. 363–378, May.

Forney G.D. Jr. (1974). "Convolutional Codes III: Sequential Decoding", Inform Control, vol. 25. P. 267–297, July.

Forney G.D. Jr. (1988). "Coset Codes I: Introduction and Geometrical Classification", IEEE Trans. Inform. Theory. vol. IT-34. P. 671–680, September.

Forney G.D. Jr., Gallager R.G., Lang G.R., Longstaff F.M. and Qureshi S.U.H. (1984). "Efficient Modulation for Band-Limited Channels", IEEE J. Selected Areas Commun., vol. SAC–2. P. 632–647, September.

Foschini G.J. (1985). "Equalizing Without Altering or Detecting Data". Bell Syst. Tech J., vol. 64. P. 1885–1911, October.

Foschini G.J., Gitlin R.D. and Weinstein S.B. (1974). "Optimization of Two-Dimensional Signal Constellations in the Presence of Gaussian Noise", IEEE Trans. Commun. vol. COM–22. P. 28–38, January.

Franaszek P.A. (1968). "Sequence-State Coding for Digital Transmission", Bell Syst. Tech. J., vol. 27. P. 143.

Franaszek P.A. (1969). "On Synchronous Variable Length Coding for Discrete Noiseless Channels", Inform. Control. vol. 15. P. 155–164.

Franaszek P.A. (1970). "Sequence-State Methods for Run-Length-Limited Codes", IBM J. Res. Dev. P. 376–383, July.

Franks L.E. (1969). Signal Theory. Prentice-Hall, Englewood Cliff, N.J.

Franks L.E. (1983). "Carrier and Bit Synchronization in Data Communication – A Tutorial Review", IEEE Trans. Commun. vol. COM–28. P. 1107–1121. August.

Franks L.E. (1981). "Synchronization Subsystems: Analysis and Design", in Digital Communications, Satellite/Earth Station Engineering, K. Feher (ed.), Prentice-Hall, Englewood Cliffs, N.J.

Fredricsson S. (1975). "Pseudo-Randomness Properties of Binary Shift Register Sequences", IEEE Trans. Inform. Theory vol. IT–21. P. 115–120, January.

Freiman C.E. and Wyner A.D. (1964). "Optimum Block Codes for Noiseless Input Restricted Channels". Inform. Control. vol. 7. P. 398–415.

Gaarder N.T. (1971). "Signal Design for Fast-Fading Gaussian Channels". IEEE Trans. Inform. Theory, vol. IT–17. P. 247–256, May.

Gabor A. (1967). "Adaptive Coding for Self Clocking Recording", IEEE Trans. Electronic Comp. vol. EC–16. P. 866.

Gallager R.G. (1965). "Simple Derivation of the Coding Theorem and Some Applications", IEEE Trans. Inform. Theory vol. IT–11. P. 3–18, January.

Gallager R.G. (1968). Information Theory and Reliable Communication. Wiley, New York.

Gardner F.M. (191979). Phaselock Techniques. Wiley, New York.

Gardner F.M. (1984). "Learning Characteristic of Stochastic Gradient Descent Algorithms: A General Study, Analysis and Critique", Signal Processing, vol. 6. P. 113–133, April.

George D.A., Bowen R.R. and Storey J.R. (1971). "An Adaptive Decision-Feedback Equalizer", IEEE Trans. Commun. Tech. vol. COM–19. P. 281–293, June.

Gercho A. (1982). "On the Structure of Vector Quantizers", IEEE Trans. Inform. Theory vol. IT–28. P. 157–166. March.

Gercho A. and Gray R.M. (1992). Vector Quantization and Signal Compression. Kluwer Academic Publishers, Boston.

Gercho A. and Lawerence V.B. (1984). "Multidimensional Signal Constellations for Voiceband Data Transmission", IEEE J. Selected Areas Commun. vol. SAC–2. P. 687–702. September.

Gerst I. and Diamond J. (1961). "The Elimination of Intersymbol Interference by Input Pulse Shaping". Proc. IRE, vol. 53. July.

Ghez S., Verdu S. and Schwartz S.C. (1988). "Stability Properties of Slotted Aloha with Multipacket Reception Capability", IEEE Trans. Autom. Control. vol. 33. P. 640–649, July.

Ghosh M. and Weber C.L. (1991). "Maximum Likelihood Blind Equalization", Proc. 1991 SPIE Conf., San Diego. Calif. July.

Gilbert E.N. (1952). "A Comparison of Signaling Alphabets", Bell Syst. Tech. J., vol. 31. P. 504–522. May.

Gilhousen K.S., Jacobs I.M., Podovani R., Viterbi A.J., Weaver L.A. and Wheatley G.E. III (1991). "On the Capicity of a Cellular CDMA System", IEEE Trans. Vehicular Tech., vol. 40. P. 303–312. May.

Gitlin R.D., Meadors H.C. and Weinstein S.B. (1982). "The Tap Leakage Algorithm: An Algorithm for the Stable Operation of a Digitally Implemented Fractionally Spaced. Adaptive Equalizer", Bell Syst. Tech. J., vol. 61. P. 1817–1839. October.

Gitlin R.D. and Weinstein S.B. (1979). "On the Required Tap-Weight Precision for Digitally Implemented Mean-Squared Equalizers", Bell Syst. Tech. J., vol. 58. P. 301–321, February.

Gitlin R.D. and Weinstein S.B. (1981). "Fractionally-Spaced Equalization: An Improved Digital Transversal Equalizer", Bell Syst. Tech. J., vol. 60. P. 275–296. February.

Glave F.E. (1972). "An Upper Bound on the Probability of Error due to Intersymbol Interference for Correlated Digital Signals", IEEE Trans. Inform. Theory vol. IT–18. P. 365–362, May.

Goblick T.J. Jr. and Holsinger J.L. (1967). "Analog Source Digitization: A Comparison of Theory and Practice", IEEE Trans. Inform. Theory, vol. IT–13. P. 323–326. April.

Godard D.N. (1974). "Channel Equalization Using a Kalman Filter for Fast Data Transmission", IBM J. Res. Dev. vol. 18. P. 267–273. May.

Godard D.N. (1980). "Self-Recovering Equalization and Carrier Tracking in Two-Dimensional Data Communications Systems", IEEE Trans. Commun. vol. COM–28. P. 1867–1875. November.

Golay M.J.E. (1949). "Note on Digital Coding". Proc. IRE, vol. 37. P. 657. June.

Gold R. (1967). "Optimal Binary Sequences for Spread Spectrum Multiplexing", IEEE Trans. Inform. Theory vol. IT–13. P. 619–621. October.

Gold R. (1968). "Maximal Recursive Sequences with 3-Valued Recursive Cross Correlation Functions", IEEE Trans. Inform. Theory vol. IT–14. P. 154–156. January.

Golomb S.W. (1967). Shift Register Sequences. Holden – Day, San Francisco, Calif.

Goppa D. (1970). "New Class of Linear Correcting Codes". Probl. Peredach. Inform. vol. 6. P. 24–30.

Goppa D. (1971). "Rational Presentation of Codes and (L, g)-Codes", Probl. Peredach. Inform. vol. 7. P. 41–49.

Gray R.M. (1975). "Sliding Block Source Coding" IEEE Trans. Inform. Theory vol. IT–21. P. 357–368. July.

Gray R.M. (1990). Source Coding Theory. Kluwer Academic Publishers, Boston.

Greefkes J.A. (1970). "A Digitally Companded Delta Modulation Modem for Speech Transmission". Proc. IEEE Int. Conf. on Commun. P. 7.33–7.48, June.

Green P.E. Jr. (1962). "Radar Astronomy Measurement Techniques", MIT Lincoln Laboratory. Lexington, Mass., Tech. Report No. 282, December.

Gronemeyer S.A. and McBride A.L. (1976). "MSK and Offset QPSK Modulation", IEEE Trans. Commun. vol. COM–24. P. 809–820. August.

Gupta S.C. (1975). "Phase-Locked Loops", Proc. IEEE, vol. 63. P. 291–306. February.

Hahn P.M. (1962). "Theoretical Diversity Improvement in Multiple Frequency Shift Keying", IRE Trans. Commun. Syst. vol. CS–10. P. 177–184. June.

Hamming R.W. (1950). "Error Detecting and Error Correcting Codes", Bell Syst. Tech. J., vol. 29. P. 147–160. April.

Hamming R.W. (1986). Coding and Information Theory. Prentice-Hall. Englewood Cliffs. N.J.

Hancock J.C. and Lucky R.W. (1960). "Performance of Combined Amplitude and Phase Modulated Communication Systems", IRE Trans. Commun. Syst. vol. CS–8. P. 232–237, December.

Hartley R.V. (1928). "Transmission of Information", Bell Syst. Tech. J., vol. 7. P. 535.

Hatzinakos D. and Nikias C.L. (1991). "Blind Equalization Using a Tricepstrum-Based Algorithm", IEEE Trans. Commun. vol. COM–39. P. 669–682, May.

Hecht M. and Guida A. (1969). "Delay Modulation", Proc. IEEE, vol. 57. P. 1314–1316, July.

Heller J.A. (1968). "Short Constraint Length Convolutional Codes", Jet Propulsion Laboratory, California Institute of Technology, Pasadena, Calif., Space Program Summary 37–54, vol. 3. P. 171–174, December.

Heller J.A. (1975). "Feedback Decoding of Convolutional Codes", in Advances in Communication System, vol. 4, A. J. Viterbi (ed.), Academic, New York.

Helstrom C.W. (1955). "The Resolution of Signals in White Gaussian Noise", Proc. IRE, vol. 43. P. 1111–1118, September.

Helstrom C.W. (1968). Statistical Theory of Signal Detection, Pergamon, London..

Helstrom C.W. (1991). Probability and Stochastic Processes for Engineers. Macmillan. New York.

Hildebrand F.B. (1960). Methods of Applied Mathematics. Prentice-Hall. Englewood Cliffs, N.J.

Hirosaki B. (1981). "An Orthogonality Multiplexed QAM System Using the Discrete Fourier Transform", IEEE Trans. Commun. vol. COM–29. P. 982–989, July.

Hirosaki B., Hasegawa S. and Sabato A. (1986). "Advanced Group-Band Modem Using Orthogonally Multiplexed QAM Techniques", IEEE Trans. Commun. vol. COM–34. P. 587–592. June.

Ho E.Y. and Yeh Y.S. (1970). "A New Approach for Evaluating the Error Probability in the Presence of Intersymbol Interference and Additive Gaussian Noise", Bell Syst. Tech. J., vol. 49. P. 2249–2265, November.

Hocquenghem A. (1959). "Codes Correcteurs d'Erreurs", Chiffres, vol. 2. P. 147–156.

Holmes J.K. (1982). Coherent Spread Spectrum Systems, Wiley-Interscience. New York.

Horwood D. and Gagliardi R. (1975). "Signal Design for Digital Multiple Access Communications", IEEE Trans. Commun. vol. COM–23. P. 378–383. March.

Hsu F.M. (1982). "Square-Root Kalman Filtering for High-Speed Data Received over Fading Dispersive HF Channels", IEEE Trans. Inform. Theory vol. IT–28. P. 753–763, September.

Huffman D.A. (1952). "A Method for the Construction of Minimum Redundancy Codes", Proc. IRE, vol. 40. P. 1098–1101, September.

Hui J.Y.N. (1984). "Throughput Analysis for Code Division Multiple Accessing of the Spread Spectrum Channel", IEEE J. Selected Areas Commun., vol. SAC–2. P. 482–486, July.

Immink K.A.S. (1990). "Runlength-Limited Sequences", Proc. IEEE vol. 78. P. 1745–1759, November.

Itakura F. (1975). "Minimum Prediction Residual Principle Applied to Speech Recognition", IEEE Trans. Acoust. Speech Signal Processing, vol. ASSP–23. P. 67–72. February.

Itacura F. and Saito S. (1968). "Analysis Synthesis Telephony Based on the Maximum-Likelihood Methods", Proc. 6th. Int. Congr. Acoust., Tokyo, Japan. P. C17–C20.

Jacobs I.M. (1974). "Practical Applications of Coding", IEEE Trans. Inform. Theory vol. IT–20. P. 305–310, May.

Jacoby G.V. (1977). "A New Look-Ahead Code for Increased Data Density", IEEE Trans. Magnetics, vol. MAG–13. P. 1202–1204.

Jayant N.S. (1970). "Adaptive Delta Modulation with a One-Bit Memory", Bell Syst. Tech. J. P. 321–342, March.

Jayant N.S. (1974). "Digital Coding of Speech Waveforms: PCM, DPCM and DM Quantizers", Proc. IEEE vol. 62. P. 611–632. May.

Jayant N.S. (1976). Waveform Quantization and Coding. IEEE Press, New York.

Jayant N.S. and Noll P. (1984). Digital Coding of Waveforms. Prentice-Hall. Englewood Cliffs, N.J.

Jelinek F. (1968). Probabilistic Information Theory. McGraw-Hill New York.

Jelinek F. (1969). "Fast Sequential Decoding Algorithm Using a Stack", IBM J. Res. Dev. vol. 13. P. 675–685. November.

Johnson C.R. (1991). "Admissibility in Blind Adaptive Channels Equalization", IEEE Control Syst. Mag P. 3–15, January.

Jones S.K., Cavin R.K. and Reed W.M. (1982). "Analysis of Error-Gradient Adaptive Linear Equalizers for a Class of Stationary-Dependent Processes", IEEE Trans. Inform. Theory, vol. IT–28. P. 318–329, March.

Jordan K.L. Jr. (1966). "The Performance of Sequential Decoding in Conjunction with Efficient Modulation", IEEE Trans. Commun. Syst. vol. CS–14. P. 283–287. June.

Justesen J. (1972). "A Class of Constructive Asymptotically Good Algebraic Codes", IEEE Trans. Inform. Theory, vol. IT–18. P. 652–656, September.

Kailath T. (1960). "Correlation Detection of Signals Perturbed by a Random Channel", IRE Trans. Inform. Theory, vol. IT–6. P. 361–366, June.

Kailath T. (1961). "Channel Characterization: Time-Variant Dispersive Channels, In Lecture on Communication System Theory, Chap 6, E. Baghdady (ed.). McGraw-Hill New York.

Kalet I. (1989). "The Multitone Channel", IEEE Trans. Commun. vol. COM–37. P. 119–124. February.

Karabed R. and Siegel P.H. (1991). "Matched-Spectral Null Codes for Partial-Response Channels", IEEE Trans. Inform. Theory vol. IT–37. P. 818–855, May.

Kasami T. (1966). "Weight Distribution Formula for Some Class of Cyclic Codes", Coordinated Science Laboratory, University of Illinois, Urbana, Ill., Tech. Report No. R-285, April.

Kaye A.R. and George D.A. (1970). "Transmission of Multiplexed PAM Signals over Multiple Channel and Diversity Systems", IEEE Trans. Commun. vol. COM–18. P. 520–525, October.

Kelly E.J., Reed I.S. and Root W.L. (1960). "The Detection of Radar Echoes in Noise, Pt. I", J. SIAM, vol. 8. P. 309–341. September.

Kleinrock L. and Tobagi F.A. (1975). "Packet Switching in Radio Channels: Part I – Carrier Sence Multiple-Access Modes and Their Throughput-Delay Characteristics", IEEE Trans. Commun. vol. COM–23. P. 1400–1416, December.

Kobayashi H. (1971). "Simultaneous Adaptive Estimation and Decision Algorithm for Carrier Modulated Data Transmission Systems", IEEE Trans. Commun. Tech., vol. COM–19. P. 268–280. June.

Kotelnikov V.A. (1947). "The Theory of Optimum Noise Immunity", Ph. D. Dissertation, Molotov Energy Institute, Moskow. [Translated by R. A. Silverman, McGraw-Hill, New York.]

Kretzmer E.R. (1966). "Generalization of a Technique for Binary Data Communication". IEEE Trans. Commun. Tech. vol. COM–14. P. 67–68, February.

Larsen K.J. (1973). "Short Convolutional Codes with Maximal Free Distance for Rates 1/2, 1/3 and 1/4", IEEE Trans. Inform. Theory, vol. IT–19. P. 371–372. May.

Lender A. (1963). "The Duobinary Techiques for High Speed Data Transmission", AIEE Trans. Commun. Electronicks, vol. 82. P. 214–218.

Leon-Garsia A. (1994). Probability and Random Processes for Electrical Engineering. Addison-Wesley, Reading, Mass.

Levinson N. (1947). "The Wiener RMS (Root Mean Square) Error Criterion in Filter Design and Predection", J. Math. and Phys., vol. 25. P. 261–278.

Lin S. and Costello D.J. Jr. (1983). Error Control Coding: Fundamentals and Applications. Prentice-Hall. Englewood Cliffs, N.J.

Linde Y., Buzo A. and Gray R.M. (1980). "An Algorithm for Vector Quantizer Design", IEEE Trans. Commun. vol. COM–28, P. 84–95, January.

Lindell G. (1985). "On Coded Continuous Phase Modulation", Ph.D. Dissertation, Telecommunication Theory, University of Lund, Lund, Sweden. May.

Lindholm J.H. (1968). "An Analysis of the Pseudo-Randomness Properties of Subsequences of Long *m*-Sequences", IEEE Trans. Inform. Theory vol. IT–14. P. 569–576, July.

Lindsey W.C. (1964). "Error Probabilities for Ricean Fading Multichannel Reception of Binary and *N*-Ary Signals", IEEE Trans. Inform. Theory, vol. IT–10. P. 339–350, October.

Lindsey W.C. (1972). Synchronization Systems in Communication. Prentice-Hall. Englewood Cliffs, N.J.

Lindsey W.C. and Chie C.M. (1981). "A Survey of Digital Phase-Locked Loops", Proc. IEEE, vol. 69. P. 410–432.

Lindsey W. C. and Simon M. K. (1973). Telecommunication Systems Engineering. Prentice-Hall. Englewood Cliffs, N.J.

Ling F. (1988). "Convergence Characteristics of LMS and LS Adaptive Algorithms for Signals with Rank-Deficient Correlation Matrices", Proc. Int. Conf. Acoust., Speech, Signal Processing, New York 25.D.4.7, April.

Ling F., Manolakis D.G. and Proakis J.G. (1986a). "Finite Word-Length Effects in Recursive Least Squares Algorithms with Application to Adaptive Equalization". Annales des Telecommunications, vol. 41. P.1–9. May–June.

Ling F., Manolakis D.G. and Proakis J.G. (1986b). "Numerically Robust Least-Squares Lattice-Ladder Algorithms with Direct Updating of the Reflection Coefficient", IEEE Trans. Acoust. Speech, Signal Processing, vol. ASSP–34. P. 837–845, August.

Ling F. and Proakis J.G. (1982). "Generalized Least Squares Lattice and Its Applications to DFE", Proc. 1982. IEEE Int. Conf. on Acoustics, Speech, Signal Processing, Paris, France, May.

Ling F. and Proakis J.G. (1984a). "Numerical Accuracy and Stability: Two Problems of Adaptive Estimation Algorithms Caused by Round-Off Error", Proc. Int. Conf. Speech Signal Processing. P. 30.3.1–30.3.4, San Diego, Calif., March.

Ling F. and Proakis J.G. (1984b). "Nonstationary Learning Characteristics of Least Squares Adaptive Estimation Algorithms", Proc. Int. Conf. Acoust., Speech. Signal Processing. P. 3.7.1–3.7.4, San Diego, Calif., March.

Ling F. and Proakis J.G. (1984c). "A Generalized Multichannel Least-Squares Lattice Algorithm with Sequential Processing Stage", IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP–32. P. 381–389. April.

Ling F. and Proakis J.G. (1985). "Adaptive Lattice Decision-Feedback Equalizers – Their Performance and Application to Time-Variant Multipath Channels", IEEE Trans. Commun. vol. COM–33. P. 348–356, April.

Ling F. and Proakis J.G. (1986). "A recursive Modified Gram-Schmidt Algorithm", IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP–34. P. 829–836, August.

Ling F. and Qureshi S.U.H. (1986). "Lattice Predictive Decision-Feedback Equalizer for Digital Communication Over Fading Multipath Channels", Proc. GLOBECOM '86, Houston, Texas, December.

Ljung S. and Ljung L. (1985). "Error Propagation Properties of Recursive Least-Squares Adaptation Algorithms". Automatica, vol. 21. P. 159–167.

Lloyd S.P. (1982). "Least Squares Quantization in PCM", IEEE Trans. Inform. Theory vol. IT–28. P. 129–137. March.

Loeve M. (1955). Probability Theory. Van Nostrand. Princeton, N.J.

Long G., Ling F. and Proakis J.G. (1987). "Adaptive Transversal Filters with Delayed Coefficient Adaptation", Proc. Int. Conf. Acoust., Speech, Signal Processing, Dallas, Texas, March.

Long G., Ling F. and Proakis J.G. (1988a). "Fractionally-Spaced Equalizers Based on Singular-Value Decomposition", Proc. Int. Conf. Acoust., Speech, Signal Processing, New York 25.D.4.10, April.

Long G., Ling F. and Proakis J.G. (1988b)., "Applications of Fractionally-Spaced Decision-Feedback Equalizers to HF

Fading Channels", Proc. MILCM, San Diego, Calif., October.

Long G., Ling F. and Proakis J.G. (1989). "The LMS Algorithm with Delayed Coefficient Adaptation", IEEE Trans. Acoust. Speech, Signal Processing, vol. ASSP–37, October.

Lucky R.W. (1965). "Automatic Equalization for Digital Communications", Bell Syst. Tech. J., vol. 44. P. 547–588, April.

Lucky R.W. (1966). "Techniques for Adaptive Equalization for Digital Communications", Bell Syst. Tech. J., vol. 45. P. 255–286.

Lucky R.W. and Hancock J.C. (1962). "On the Optimum Performance of N-ary Systems Having Two Degrees of Freedom", IRE Trans. Commun. Syst. vol. CS–10. P. 185–192, June.

Lucky R. W., Salz J. and Weldon E.J., Jr. (1968). Principles of Data Communication. McGraw-Hill, New York.

Lugannani R. (1969). "Intersymbol Interference and Probability of Error in Digital Systems", IEEE Trans. Inform. Theory vol. IT–15. P. 682–688, November.

Lundgren C.W. and Rummler W.D. (1979). "Digital Radio Outage Due to Selective Fading–Observation vs. Prediction from Laboratory Simulation", Bell Syst. Tech. J., vol. 58. P. 1074–1100, May–June.

Lupas R. and Verdu S. (1989). "Linear Multiuser Detector for Synchronous Code-Division Multiple Access Channels", IEEE Trans. Inform. Theory vol. IT–35. P. 123–136, January.

Lupas R. and Verdu S. (1990). "Near-Far Resistance of Multiuser Detectors in Asynchronous Channels", IEEE Trans. Commun. vol. COM–38. P. 496–508, April.

MakKenchnie L.R. (1973). "Maximum Likelihood Receivers for Channels Having Memory", Ph.D. Dissertation, Department of Electrical Engineering, University of Notre Dame, Notre Dame, Ind., January.

MacWilliams F.J. and Sloane J.J. (1977). The Theory of Error Correcting Codes, North Holland. New York.

Magee F.P. and Proakis J.G. (1973). "Adaptive Maximum-Likelihood Sequence Estimation for Digital Signaling in the Presence of Intersymbol Interference", IEEE Trans. Inform. Theory, vol. IT–19. P. 120–124, January.

Makhoul J. "A Class of All-Zero Lattice Digital Filters: Properties and Applications", IEEE Trans. Acoust. Speech, Signal Processing, vol. ASSP–26. P. 304–314, August.

Makhoul J., Roucos S. and Gish H. (1985). "Vector Quantization in Speech Coding", Proc. IEEE vol. 73. P. 1551–1587, November.

Martin D.R. and McAdam P.L. (1980). "Convolutional Code Performance with Optimal Jamming", Conf. Rec. Int. Conf. Commun. P.4.3.1–4.3.7, May.

Massey J.L. (1963). Threshold Decoding. MIT Press, Cambridge, Mass.

Massey J.L. (1965). "Step-by-Step Decoding of the BCH Codes", IEEE Trans. Inform. Theory vol. IT–11. P. 580–585. October.

Massey J.L. (1988). "Some New Approaches to Random Access Communications", Performance'87. P. 551–569. [Reprinted 1993 in Multiple Access Communications, N. Abramson (ed.), IEEE Press, New York.]

Massey J.L. and Sain M. (1968). "Inverces of Linear Sequential Circuits", IEEE Trans. Comput., vol. C–17. P. 330–337, April.

Matis K.R. and Modestino J.W. (1982). "Reduced-State Soft-Decision Trellis Decoding of Linear Block Codes", IEEE Trans. Inform. Theory vol. IT–28. P. 61–68, January.

Max J. (1960). "Quantizing for Minimum Distortion", IRE Trans. Inform. Theory vol. IT–6. P. 7–12, March.

Mazo J.E. (1975). "Faster-Than-Nyquist Signaling", Bell Syst. Tech. J., vol. 54. P. 1451–1462, October.

Mazo J.E. (1979). "On the Independence Theory of Equalizer Convergence", Bell Syst. Tech. J., vol. 58. P. 963–993. May.

McMahon M.A. (1984). The Making of a Profession – A Century of Electrical Engineering in America. IEEE Press. New York.

Mengali U. (1977). "Joint Phase and Timing Acquisition in Data Transmission", IEEE Trans. Commun. vol. COM–25. P. 1174–1185, October.

Meyers M.H. and Franks L.E. (1980). "Joint Carrier Phase and Symbol Timing for PAM Systems", IEEE Trans. Commun. vol. COM–28. P. 1121–1129. August.

Meyr H. and Ascheid G. (1990). Synchronization in Digital Communications. Wiley Interscience, New York.

Miller K.S. (1964). Multidimensional Gaussian Distributions. Wiley, New York.

Millman S. (ed.) (1984). A History of Engineering and Science in the Bell System–Communication Sciences (1925–1980), AT&T Bell Laboratories.

Miyagaki Y., Morinaga N., and Namekawa T. (1978). "Error Probability Characteristics for CPSK Signal Through *m*-Distributed Fading Channel", IEEE Trans. Commun. vol. COM–26. P. 88–100, January.

Monsen P. (1971). "Feedback Equalization for Fading Dispersive Channels", IEEE Trans. Inform. Theory, vol. IT–17. P. 56–64, January.

Morf M. (1977). "Ladder Forms in Estimation and System Identification", Proc. 11th Annual Asilomar Conf. on Circuits, Systems and Computers, Monterey, Calif., Nov. 7–9.

Morf M., Dickinson B., Kailath T. and Vieira A. (1977). "Efficient Solution of Covariance Equations for Linear Prediction", IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP–25. P. 429–433, October.

Morf M. and Lee D. (1979). "Recursive Lest Squares Ladder Forms for Fast Parameter Tracking", Proc. 1978 IEEE Conf. on Decision and Control, San Diego, Calif. P. 1362–1367, January 12.

Morf M., Lee D., Nickolls J. and Vieira A. (1977). "A Classification of Algorithms for ARMA Models and Ladder Realizations", Proc. 1997 IEEE Int. Conf. On Acoustics, Speech, Signal Processing, Hartford, Conn. P. 13–19, May.

Mort M., Vieira A. and Lee D. (1977). "Ladder Forms for Identification and Speech Processing", Proc. 1997 IEEE Conf. on Decision and Control, New Orleans, La. P. 1074–1078, December.

Mueller K.H. and Muller M.S. (1976). "Timing Recovery in Digital Synchronous Data Receivers", IEEE Trans. Commun. vol. COM–24. P. 516–531. May.

Muller D.E. (1954). "Application of Boolean Algebra to Switching Circuit Design and to Error Detection", IRE

Trans. Electronic Comput., vol. EC–3. P. 6–12, September.

Milligan M.G. (1988). "Multi-Amplitude Continuous Phase Modulation with Convolutional Coding", Ph. D. Dissertation, Department of Electrical and Computer Engineering, North-eastern University, June.

Nakagami M. (1960). "The *m*-Distribution – A General Formula of Intensity Distribution of Rapid Fading", in Statictical Methods of Radio Wave Propagation, W. C. Hoffman (ed.). P. 3–36, Pergamon Press, New York.

Natali F.D. and Walbesser W.J. (1969). "Phase-Locked Loop Detection of Binary PSK Signals Utilizing Decision Feedback", IEEE Trans. Aerospace Electronic Syst. vol. AES–5. P. 83–90, January.

Neyman J. and Pearson E.S. (1933). "On the Problem of the Most Efficient Tests of Statistical Hypotheses", Phil. Trans. Roy. Soc. London Series A, vol. 231. P. 289–337.

North D.O. (1943). "An Analysis of the Factors Which Determine Signal/Noise Discrimination in Pulse-Carrier Systems", RCA Tech. Report No. 6 PTR-6C.

Nyquist H. (1924). "Certain Factors Affecting Telegraph Speed", Bell Syst. Tech. J., vol. 3. P. 324.

Nyquist H. (1928). "Certain Topics in Telegraph Transmission Theory", AIEE Trans., vol. 47. P. 617–644.

Odenwalder J.P. (1970). "Optimal Decoding of Convolutional Codes", Ph. D. Dissertation, Department of Systems Sciences, School of Engineering and Applied Sciences, University of California, Los Angeles.

Odenwalder J.P. (1976). "Dual-*k* Convolutional Codes for Noncoherently Demodulated Channels", Proc. Int. Telemetering Conf., vol. 12. P. 165–174, September.

Olsen J.D. (1977). "Nonlinear Binary Sequences with Asymptotically Optimum Periodic Cross Correlation", Ph. D. Dissertation, University of Southern California, December.

Omura J. (1971). "Optimal Receiver Design for Convolutional Codes and Channels with Memory Via Control Theoretical Concepts", Inform. Sci., vol. 3. P. 243–266.

Omura J.K. and Levitt B.K. (1982). "Code Error Probability Evaluation for Antijam Communication Systems ", IEEE Trans. Commun. vol. COM–30. P. 896–903, May.

Osborne W.P. and Luntz M.B. (1974). "Coherent and Noncoherent Detection of CPFSK", IEEE Trans. Commun. vol. COM–22. P. 1023–1036, August.

Paaske E. (1974). "Short Binary Convolutional Codes with Maximal Free Distance for Rates 2/3 and 3/4", IEEE Trans. Inform. Theory vol. IT–20. P. 683–689, September.

Pacz M.D. and Glisson T. H. (1972). "Minimum Mean Squared Error Quantization in Speech PCM and DPCM Systems", IEEE Trans. Commun., vol. COM–20. P. 225–230, April.

Pahlavan K. (1985). "Wireless Communications for Office Information Networks", IEEE Commun. Mag., vol. 23, P. 18–27, June.

Papoulis A. (1984). Probability, Random Variables and Stochastic Processes. McGraw-Hill New York.

Paul D.B. (1983). "An 800 bps Adaptive Vector Quantization Vocoder Using a Preceptual Distance Measure", Proc. IEEE Int. Conf. Acoust. Speech, Signal Processing, Boston, Mass. P. 73–76, April.

Pearson K. (1965). Tables of the Incomplete Γ-Function, Cambridge University Press, London.

Peebles P.Z. (1987). Probability, Random Variables, and Random Signal Principles. McGraw-Hill New York.

Peterson W.W. (1960). "Encoding and Error-Correction Procedures for Bose-Chaudhuri Codes", IRE Trans. Inform. Theory vol. IT–6. P. 459–470, September.

Peterson W.W. and Weldon E.J., Jr. (1972). Error-Correcting Codes, 2nd ed., MIT Press, Cambridge, Mass.

Picci G. and Prati G. (1987). "Blind Equalization and Carrier Recovery Using a Stop-and-Go Decision Directed Algorithm", IEEE Trans. Commun. vol. COM–35. P. 877–887, September.

Picinbono B. (1978). "Adaptive Signal Processing for Detection and Communication", in Communication Systems and Random Process Theory, J. K. Skwirzynski (ed.), Sijthoff & Nordhoff, Alphen aan den Rijn, The Netherlands.

Pickholtz R.L., Shilling D.L. and Milstein L.B. (1982). "Theory of Spread Spectrum Communications–A Tutorial", IEEE Trans. Commun. vol. COM–30. P. 855–884, May.

Pieper J.F., Proakis J.G., Reed R.R. and Wolf J.K. (1978). "Design of Efficient Coding and Modulation for a Rayleigh Fading Channel", IEEE Trans. Inform. Theory vol. IT–24. P. 457–468, July.

Pierce J.N. (1958). "Theoretical Diversity Improvement in Frequency-Shift Keying", Proc. IRE vol. 46. P. 903–910, May.

Pierce J.N. and Stein S. (1960). "Multiple Diversity with Non-Independent Fading", Proc. IRE vol. 48. P. 89–104, January.

Plotkin M. (1960). "Binary Codes with Specified Minimum Distance", IRE Trans. Inform. Theory, vol. IT–6. P. 445–450, September.

Poor H.V. and Verdu S. (1988). "Single-User Detectors for Multiuser Channels", IEEE Trans. Commun. vol. 36. P. 50–60, January.

Price R. (1954). "The Detection of Signals Perturbed by Scatter and Noise", IRE Trans. Inform. Theory vol. PGIT–4. P. 163–170, September.

Price R. (1956). "Optimum Detection of Random Signals in Noise, with Application to Scatter-Multipath Communication", IRE Trans. Inform. Theory vol. IT–2. P. 125–135, December.

Price R. (1962a). "Error Probabilities for Adaptive Multichannel Reception of Binary Signals", MIT Lincoln Laboratory, Lexington, Mass., Tech. Report No. 258, July.

Price R. (1962b). "Error Probabilities for Adaptive Multichannel Reception of Binary Signals", IRE Trans. Inform. Theory vol. IT–8. P. 305–316, September.

Price R. (1972). "Nonlinearly Feedback-Equalized PAM vs. Capacity", Proc. 1972 IEEE Int. Conf. on Commun. Philadelphia, Penn. P. 22.12–22.17, June.

Price R. and Green P.E., Jr. (1958). "A Communication Technique for Miltipath Channels", Proc. IRE, vol. 46. P. 555–570, March.

Price R. and Green P.E. Jr. (1960). "Signal Processing in Radar Astronomy–Communication via Fluctuating Multipath Media", MIT Lincoln Laboratory, Lexington, Mass., Tech. Report No. 234, October.

Proakis J.G. (1968). "Probabilities of Error for Adaptive Reception of *M*-Phase Signals", IEEE Trans. Commun. Tech., vol. COM–16. P. 71–81, February.

Proakis J.G. (1975). "Advances in Equalization for Intersymbol Interference", In Advanced in Communication Systems, vol. 4, A. J. Viterbi (ed.) Academic, New York.

Proakis J.G., Drouihet P.R., Jr. and Price R. (1964). "Performance of Coherent Detection Systems Using Decision-Directed Channel Measurement", IEEE Trans. Commun. Syst., vol. CS-12. P. 54–63, March.

Proakis J.G. and Ling F. (1984). "Recursive Least Squares Algorithms for Adaptive Equalization of Time-Variant Multipath Channels", Proc. Int. Conf. Commun. Amsterdam, The Netherlands, May.

Proakis J.G. and Manolakis D.G. (1988) Introduction to Digital Processing, Macmillan, New York.

Proakis J.G. and Miller J.H. (1969). "Adaptive Receiver for Digital Signaling through Channels with Intersymbol Interference", IEEE Trans. Inform. Theory vol. IT-15. P. 484–497, July.

Proakis J.G. and Rahman I. (1979). "Performance of Contenated Dual-*k* Codes on a Rayleigh Fading Channel with a Bandwidth Constraint", IEEE Trans. Commun. vol. COM-27. P. 801–806, May.

Pursley M.B. (1979). "On the Mean-Square Partial Correlation of Periodic Sequences", Proc. 1979 Conf. Inform. Science and Systems, Johns Hopkins University, Baltimore, Md. P. 377–379, March.

Qureshi S.U.H. (1976). "Timing Recovery for Equalized Partial Response Systems", IEEE Trans. Commun. vol. COM-24. P. 1326–1331, December.

Qureshi S.U.H. (1985). "Adaptive Equalization" Proc. IEEE, vol. 53. P. 1349–1387, September.

Qureshi S.U.H. and Forney G.D. Jr. (1977). "Performance and Properties of a *T*/2 Equalizer", Natl. Telecom. Conf. Record, P. 11.1.1–11.1.14, Los Angeles, Calif., December.

Rabiner L.R. and Schafer R.W. (1978). Digatal Processing of Speech Signal. Prentice-Hall. Englewood Cliffs, N.J.

Raheli R., Polydoros A. and Tzou C.-K. (1995). "The Principle of Per-Survivor Processing. A General Approach to Approximate and Adaptive MSLE", IEEE Trans. Commun. vol. COM-43 (to appear).

Rahman I. (1981). "Bandwidth Constrained Signal Design for Digital Communication over Rayleigh Fading Channels and Partial Band Interference Channels", Ph. D. Dissertation, Department of Electrical Engineering, Northeastern University, Boston, Mass.

Ramsey J.L. (1970). "Realization of Optimum Interleavers", IEEE Trans. Inform. Theory vol. IT-16. P. 338–345.

Reed I.S. (1954). "A Class of Multiple-Error Correcting Codes and the Decoding Scheme", IRE Trans. Inform. vol. IT-4. P. 38–49, September.

Reed I.S. and Solomon G. (1960). "Polynomial Codes Over Certain Finite Fields", SIAM J., vol. 8. P. 300–304, June.

Rizos A.D., Proakis J.G. and Nguven T.Q. (1994). "Comparison of DFT and Cosine Modulated Filter Banks in Multicarrier Modulation", Proc. Globecom '94. P. 687–691, San Francisco, Calif., November.

Roberts L.G. (1975). "Aloha Packet System with and without Slots and Capture", Comp. Commun. Rev, vol. 5. P. 28–42, April.

Roucos S., Schwartz R. and Makhoul J. (1982). "Segment Quantization for Very-Low-Rate Speech Coding", Proc. Int. Conf. Acoust., Speech Signal Processing, Paris, France, P. 1565–1569, May.

Rowe H.E. and Prabhu V.K. (1975). "Power Spectrum of a Digital Frequency Modulation Signal", Bell Syst. Tech. J., vol. 54. P. 1095–1125, July–August.

Rummler W.D. (1979). "A New Selective Fading Model: Application to Propagation Data", Bell Syst. Tech. J., vol. 58. P. 1037–1071. May–June.

Ryder J.D. and Fink D.G. (1984). Engineers and Electronics. IEEE Press, New York.

Saltzberg B.R. (1967). "Performance of an Efficient Parallel Data Transmission System", IEEE Trans. Commun. vol. COM-15. P. 805–811, December.

Saltzberg B.R. (1968). "Intersymbol Interference Error Bounds with Application to Ideal Bandlimited Signaling" IEEE Trans. Inform. Theory vol. IT-14. P. 563–568, July.

Salz J. (1973). "Optimum Mean-Square Decision Feedback Equalization", Bell Syst. Tech. J., vol. 52. P. 1341–1373, October.

Salz J., Sheehan J.R. and Paris D.J. (1971). "Data Transmission by Cominbed AM and PM", Bell Syst. Tech. J., vol. 50. P. 2399–2419, September.

Sarwate D.V. and Pursley M.B. (1980). "Crosscorrelation Properties of Pseudorandom and Related Sequences", Proc. IEEE, vol. 68. P. 593–619, May.

Sato Y. (1975). "A methods of Self-Recovering Equalization for Multilevel Amplitude-Modulation Systems", IEEE Trans. Commun. vol. COM-23. P. 679–682, June.

Sato Y. et. al. (1986). Blind Suppression of Time Dependency and its Extension to Multi-Dimensional Equalization", Proc. ICC'86. P. 46.4.1–46.4.5.

Satorius E.H. and Alexander S.T. (1979). "Channel Equalization Using Adaptive Lattice Algorithms", IEEE Trans. Commun. vol. COM-27. P. 899–905, June.

Satorius E.H. and Pack J.D. (1981). "Application of Least Squares Lattice Algorithms to Adaptive Equalization". IEEE Trans. Commun. vol. COM-29. P. 136–142. February.

Savage J.E. (1966). "Sequential Decoding – The Computation Problem", Bell Syst. Tech. J., vol. 45. P. 149–176, August.

Scholtz R.A. (1977). "The Spread Spectrum Concept", IEEE Trans. Commun. vol. COM-25. P. 748–755, August.

Scholtz R.A. (1979). "Optimal CDMA Codes", 1979 Nat. Telecommun. Conf. Rec., Washington, D. C. P. 54.2.1–54.2.4, November.

Scholtz R.A. (1982). "The Origins of Spread Spectrum", IEEE Trans. Commun. vol. COM-30. P. 822–854, May.

Schonhoff T.A. (1976). "Symbol Error Probabilities for *M*-ary CPFSK: Coherent and Noncoherent Detection", IEEE Trans. Commun. vol. COM-24. P. 644–652, June.

Seshadri N. (1994). "Joint Data and Channel Estimation Using Fast Blind Trellis Search Techniques", IEEE Trans. Commun. vol. COM-42. P. 1000–1011. March.

Shalvi O. and Weinstein E. (1990). "New Criteria for Blind Equalization of Nonminimum Phase Systems Channels", IEEE Trans. Inform. Theory vol. IT-36. P. 312–321, March.

Shannon C.E. (1948a). "A Matematical Theory of Communication" Bell Syst. Tech. J., vol. 27. P. 379–423, July.

Shannon C.E. (1948b). "A Matematical Theory of Communication" Bell Syst. Tech. J., vol. 27. P. 623–656, October.

Shannon C.E. (1949). "Communication in the Presence of Noise", Proc. IRE, vol. 37. P. 10–21, January.

Shannon C.E. (1959a). "Coding Theorems for a Discrete Sourse with a Fidelity Criterion", IRE Nat. Conv. Rec., pt. 4. P. 142–163, March.

Shannon C.E. (1959b). "Probability of Error for Optimal Codes in a Gaussian Channel", Bell Syst. Tech. J., vol. 38. P. 611–656, May.

Shannon C.E., Gallger R.G. and Berlekamp E.R. (1967). "Lower Bounds to Error Probability for Coding on Discrete Memoryless Channels, I and II", Inform. Control. vol. 10. P. 65–103, January. P. 527–552, May.

Shimbo O. and Celebiler M. (1971). "The Probability of Error due to Intersymbol Interference and Gaussian Noise in Digital Communication Systems", IEEE Trans. Commun. Tech., vol. COM–19. P. 113–119, April.

Simon M.K. and Divsalar D. (1985). "Combined Trellis Coding with Asymmetric MPSK Modulation", JPL Publ. 85–24, Pasadena, Calif., May.

Simon M.K., Omura J.K., Scholtz R.F. and Levitt B.K. (1985). Spread Spectrum Communications Vol. I, II, III. Computer Science Press, Rockville, Md.

Simon M.K. and Smith J.G. (1973). "Hexagonal Multiple Phase-and-Amplitude-Shift Keyed Signal Sets", IEEE Trans. Commun. vol. COM–21. P. 1108–1115, October.

Slepian D. (1956). "A Class of Binary Signaling Alphabets", Bell Syst. Tech. J., vol. 35. P. 203–234, January.

Slepian D. (1974). Key Papers in the Development of Information Theory. IEEE Press, New York.

Slepian D. and Wolf J.K. (1973). "A Coding Theorem for Multiple Access Channels with Correlated Sources", Bell Syst. Tech. J., vol. 52. P. 1037–1076.

Sloane N.J.A. and Wyner A.D. (1993). The Collected Papers of Shannon. IEEE Press, New York.

Slock D.T.M. and Kailath T. (1988). "Numerically Stable Fast Recursive Least-Squares Transversal Filters", Proc. Int. Conf. Acoust. Speech, Signal Processing. P. 1365–1368, New York, April.

Smith J.W. (1965). "The Joint Optimization of Transmitted Signal and Receiving Filter for Data Transmissiom Systems", Bell Syst. Tech. J., vol. 44. P. 1921–1942, December.

Stenbit J.P. (1964). "Table of Generators for BCH Codes", IEEE Trans. Inform. Theory, vol. IT–10. P. 390–391, October.

Stiffler J.J. (1971). Theory of Synchronous Communications. Prentice-Hall. Englewood Cliffs, N.J.

Sundberg C.E. (1986). "Continuous Phase Modulation. IEEE Commun. Mag., vol. 24. P. 25–38, April.

Suzuki H. (1977). "A Statistical Model for Urban Multipath Channels with Random Delay", IEEE Trans. Commun. vol. COM–25, P. 673–680, July.

Tang D.L. and Bahl L.R. (1970). "Block Codes for a Class of Constrained Noiseless Channels", Inform. Control., vol. 17. P. 436–461.

Titsworth R.C. and Welch L.R. (1961). "Power Spectra of Signals Modulated by Random and Pseodorandom Sequences". JPL Tech. Rep. 32–140, October 10.

Thomas C.M., Weidner M.Y. and Durrani S.H. (1974). "Digital Amplitude-Phase-Keying with *M*-ary Alphabets". IEEE Trans. Commun. vol. COM–22. P. 168–180, February.

Tong L., Xu G. and Kailath (1991). "Blind Identification and Equalization Based on Second-Order Statistics", IEEE Trans. Inform. Theory vol. IT–40. P. 340–349. March.

Tufts D.W. (1965). "Nyquist's Problem – The Joint Optimization of Transmitter and Receiver in Pulse Amplitude Modulation", Proc. IEEE, vol. 53. P. 248–259, March.

Turin G.L. (1961). "On Optimal Diversity Reception". IRE Trans. Inform. Theory vol. IT–7. P. 154–166, July.

Turin G.L. (1962). "On Optimal Diversity Reception II" IRE Trans. Commun. Syst., vol. CS–12. P. 22–31, March.

Turin G.L et al. (1972) "Simulation of Urban Vehicle Monitoring Systems", IEEE Trans. Vehicular Tech. P. 9–16, February.

Tzannes M.A., Tzannes M.C., Proakis J.G. and Heller P.N. (1994). "DMT Systems, DWMT Systems and Digital Filter Banks", Proc. Int. Conf. Commun. P. 311–315, New Orleans, Louisiana, May 1–5.

Ungerboeck G. (1972). "Theory on the Speed of Convergence in Adaptive Equalizers for Digital Communication", IBM J. Res. Dev., vol. 16. P. 546–555, November.

Ungerboeck G. (1974). "Adaptive Maximum – Likelihood Receiver for Carrier-Modulated Data-Transmission Systems", IEEE Trans. Commun. vol. COM–22. P. 624–636, May.

Ungerboeck G. (1976). "Fractional Tap-Spacing Equalizer and Consequences for Clock Recovery in Data Modems". IEEE Trans. Commun. vol. COM–24. P. 856–864, August.

Ungerboeck G. (1982). "Channel Coding with Multilevel/Phase Signals", IEEE Trans. Inform. Theory. vol. IT–28. P. 55–67, January.

Ungerboeck G. (1987). "Trellis-Coded Modulation with Redundant Signal Sets, Part I and II", IEEE Commun. Mag., vol. 25. P. 5–21, February.

Ungerbocck G. and Csajka I. (1976). "On Improving Data-Link Performance by Increasing the Channel Alphabet and Introducing Sequence Coding", 1976 Int. Conf. Inform Theory, Ronneby, Sweden, June.

Vaidyanathan P.P. (1993). Multirate Systems and Filter Banks. Prentice-Hall. Englewood Cliffs, N.J.

Van Etten W. (1975). "An Optimum Linear Receiver for Multiple Channel Digital Transmission Systems", IEEE Trans. Commun. vol. COM–23. P. 828–834, August.

Van Etten W. (1976). "Maximum Likelihood Receiver for Multiple Channel Transmission Systems", IEEE Trans. Commun., vol. COM–24. P. 276–283, February.

Van Trees H.L. (1968). Detection, Estimation and Modulation Theory, Part I, Wiley, New York.

Varsharmov R.R. (1957). "Estimate of the Number of Signals in Error Correcting Codes", Doklady Akad. Nauk. S.S.S.R., vol. 117. P. 739–741.

Verdu S. (1986a). "Minimum Probability of Error for Asynchronous Gaussian Multiple-Access Channels", IEEE Trans. Inform. Theory, vol. IT–32. P. 85–96, January.

Verdu S. (1986b). "Multiple-Access Channels with Point-Process Observation: Optimum Demodulation", IEEE

Trans. Inform. Theory, vol. IT–32. P. 642–651, September.

Verdu S. (1986c). "Optimum Multiuser Asymptotic Efficiency", IEEE Trans. Commun., vol. COM–34. P. 890–897, September.

Verdu S. (1989). "Recent Progress in Multiuser Detection", Advances in Communications and Signal Processing, Springer-Verlag, Berlin. [Reprinted in Multiple Access Communications, N. Abramson (ed.), IEEE Press, New York.].

Viterbi A.J. (1966). Principles of Coherent Communication, McGraw-Hill, New York.

Viterbi A.J. (1967). "Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm", IEEE Trans. Inform. Theory, vol. IT–13. P. 260–269, April.

Viterbi A.J. (1969). "Error Bounds for White Gaussian and Other Very Noisy Memoryless Channels with Generalized Decision Regions", IEEE Trans. Inform. Theory, vol. IT–15. P. 279–287, March.

Viterbi A.J. (1971). "Convolutional Codes and Their Performance in Communication Systems", IEEE Trans. Commun. Tech., vol. COM–19. P. 751–772, October.

Viterbi A.J. (1978). "A Processing Satellite Transponder for Multiple Access by Low-Rate Mobile Users", Proc. Fourth Int. Conf. on Digital Satellite Communications, Monreal, Canada. P. 166–174, October.

Viterbi A.J. (1979). "Spread Spectrum Communication – Myths and Realities", IEEE Commun. Mag., vol. 17. P. 11–18, May.

Viterbi A.J. (1985). "When Not to Spread Spectrum – A Sequel", IEEE Commun. Mag., vol. 23. P. 12–17, April.

Viterbi A.J. and Jacobs (1975). "Advances in Coding and Modulation for Noncoherent Channels Affected by Fading, Partial Band and Multiple-Access Interference", in Advances in Communication Systems, vol. 4, A. J. Viterbi (ed.), Academic, New York.

Viterbi A.J. and Omura J. K. (1979). Principles of Digital Communication and Coding. McGraw-Hill, New York.

Wainberg S. and Wolf J.K. (1970). "Subsequences of Pseudo-Random Sequences", IEEE Trans. Commun. Tech., vol. COM–18, P. 606–612, October.

Wainberg S. and Wolf J.K. (1973). "Algebraic Decoding of Block Codes Over a q-ary Input, Q-ary Output Channel, $Q>q$", Inform Control, vol. 22. P. 232–247, April.

Wald A. (1947). Sequential Analysis. Wiley, New York.

Ward R.B. (1965). "Acquisition of Pseudonoise Signals by Sequential Estimation", IEEE Trans. Commun. Tech., vol. COM–13, P. 474–483, December.

Ward R.B. and Yiu K.P. (1977). "Acquisition of Pseudonoise Signals by Recursion-Aided Sequential Estimation", IEEE Trans. Commun., vol. COM–25. P. 784–794, August.

Weber W.J., III, Stanton P.H. and Sumida J.T. (1978). "A Bandwidth Compressive Modulation System Using Multi-Amplitude Minimum-Shift Keying (MAMSK)", IEEE Trans. Commun., vol. COM–26. P. 543–551, May.

Wei L.F. (1984a). "Rotationally Invariant Convolutional Channel Coding with Expanded Signal Space, Part I: 180^0", IEEE Selected Areas Commun., vol. SAC–2. P. 659–671, September.

Wei L.F. (1984b). "Rotationally Invariant Convolutional Channel Coding with Expanded Signal Space, Part II: Nonlinear Codes", IEEE J. Selected Areas Commun., vol. SAC–2. P. 672–686, September.

Wei L.F. (1987). "Trellis-Coded Modulation with Multi-Dimensional Constellations", IEEE Trans. Inform. Theory, vol. IT–33, P. 483–501, July.

Weinstein S.B. and Ebert P.M. (1971). "Data Transmission by Frequency-Division Multiplexing Using the Discrete Fourier Transform", IEEE Trans. Commun., vol. COM–19. P. 628–634, October.

Welch L.R. (1974). "Lower Bounds on the Maximum Cross Correlation of Signals", IEEE Trans. Inform. Theory, vol. IT–20. P. 397–399, May.

Weldon E.J., Jr. (1971). "Decoding Binary Block Codes on Q-ary Output Channels", IEEE Trans. Inform. Theory, vol. IT–17. P. 713–718, November.

Widrow B. (1966). "Adaptive Filters, I: Fundamentals" Stanford Electronic Laboratory, Stanford University, Stanford, Calif., Tech. Report No. 6764–6, December.

Widrow B. and Hoff M.E., Jr. (1960). "Adaptive Switching Circuits", IRE WESCON Conv. Rec., pt. 4. P. 96–104.

Widrow B. et. al. (1975). "Adaptive Noise Cancelling: Principles and Applications", Proc. IEEE vol. 63. P. 1692–1716, December.

Wiener N. (1949). The Extrapolation, Interpolation and Smoothing of Stationary Time Series with Engineering Applications. Wiley, New York [reprinted of original work published as an MIT Radiation Laboratory Report in 1942].

Wintz P.A. (1972). "Transform Picture Coding", Proc. IEEE. vol. 60. P. 880–920, July.

Wolf J.K. (1978). "Efficient Maximum Likelihood Decoding of Linear Block Codes Using a Trellis", IEEE Trans. Inform. Theory, vol. IT–24. P. 76–81, January.

Wozencraft J.M. (1957). "Sequential Decoding for Reliable Communication", IRE Nat. Conv. Rec., vol. 5, pt. 2. P. 11–25.

Wozencraft J.M. and Jacobs I. M. (1965). Principles of Communication Engineering. Wiley, New York.

Wozencraft J.M. and Kennedy R. S. (1966). "Modulation and Demodulation for Probabilistic Decoding", IEEE Trans. Inform. Theory, vol. IT–12. P. 291–297, July.

Wozencraft J.M. and Reiffen B. (1961). Sequential Decoding. MIT Press, Cambridge, Mass.

Wyner A.D. (1965) "Capacity of the Band-Limited Gaussian Channel", Bull. Syst. Tech. J., vol. 45. P. 359–371, March.

Xie Z., Rushforth C.K. and Short R.T. (1990a). "Multiuser Signal Detection Using Sequential Decoding", IEEE Trans. Commun. vol. COM–38. P. 578–583, May.

Xie Z., Short R.T. and Rushforth C.K. (1990b). "A Family of Suboptimal Detectors for Coherent Multiuser Communications", IEEE J. Selected Areas Commun. vol. SAC–8. P. 683–690, May.

Yao K(1972). "On Minimum Average Probability of Error Expression for Binary Pulse-Communication System with Intersymbol Interference", IEEE Trans. Inform. Theory, vol. IT–18. P. 528–531, July.

Yao K. and Tobin R.M. (1976). "Moment Space Upper and Lower Error Bounds for Digital Systems with Intersymbol Interference", IEEE Trans. Inform. Theory, vol. IT–22. P. 65–74, January.

Yue O. (1983). "Spread Spectrum Mobile Radio 1977–1982", IEEE Trans. Vehicular Tech., vol. VT–32. P. 98–105, February.

Zelinski P. and Noll P. (1977). "Adaptive Transform Coding of Speech Signals", IEEE Trans. Acoustics, Speech, Signal Processing, vol. ASSP–25. P. 299–309, August.

Zervas E., Proakis J.G. and Eyuboglu V. (1991). "A Quantized Channel Approach to Blind Equalization", Proc. ICC'91, Chicago, II, June.

Zhang X. and Brady D. (1993). "Soft-Decision Multistage Detection of Asynchronous AWGN Channels", Proc. 31st Allerton Conf. on Commun. Contr., Comp. Allerton. II., October.

Zhou K. and Proakis J.G. (1988). "Coded Reduced-Bandwidth QAM with Decision-Feedback Equalization", Conf. Rec. IEEE Int. Conf. Commun., Philadelphia, Penn. P. 12.6.1–12.6.5, June.

Zhou K., Proakis J.G. and Ling F. (1987). "Decision-Feedback Equalization of Fading Dispersive Channels with Trellis-Coded Modulation", Int. Conf. Commun. Tech., Nanjing, China, November.

Zhou K. Proakis J.G. and Ling F. (1990). "Decision-Feedback Equalization of Time-Dispersive Channels with Coded Modulation", IEEE Trans. Commun., vol. COM–38. P. 18–24, January.

Ziemer R.E. and Peterson R.L. (1985). Digital Communications and Spread Spectrum Systems, Macmillan, New York.

Zigangirov K.S. (1966). "Some Sequential Decoding Procedures", Probl. Peredach. Inform., vol. 2. P. 13–25.

Ziv J. (1985). "Universal Quantization", IEEE Trans. Inform. Theory vol. 31. P. 344–347.

Ziv J. and Lempel A. (1977). "A Universal Algorithm for Sequential Data Compression", IEEE Trans. Inform. Theory, vol. IT–23, P. 337–343.

Ziv J. and Lempel A. (1978). "Compression of Individual Sequences via Variable-Rate Coding", IEEE Trans. Inform. Theory, vol. IT–24. P. 530–536.

Zvonar Z. and Brady D. (1995). "Differentially Coherent Multiuser Detection in Asynchronous CDMA Flat Rayleigh Fading Channels", IEEE Trans. Commun., vol. COM–43, to appear.

СПИСОК ДОПОЛНИТЕЛЬНОЙ ЛИТЕРАТУРЫ

1. Котельников В.А. О пропускной способности «эфира» и проволоки в электросвязи. – Всесоюзный энергетический комитет. Материалы к I Всесоюзному съезду по вопросам реконструирования дела связи и развития слаботочной промышленности. Изд. ред. Упр. связи РККА, 1933.
2. Khintchine A. Korrelations theorie der stationaren stochastischen Prozessen. – Math. Annalen.1934.109. Рус. пер.: Хинчин А.Я Теория корреляции стационарных стохастических процессов. – Успехи мат. наук. 1938, вып. 5.
3. Колмогоров А.Н. Интерполирование и экстраполирование стационарных случайных последовательностей. Известия АН СССР, серия математическая. 1941, т.5, № 5.
4. Крейн М.Г. Об основной аппроксимационной задаче теории экстраполяции и фильтрации стационарных случайных процессов. – ДАН, 1954, т.94, С.13–16.
5. Стратонович Р.Л. Условные процессы Маркова. – Теория вероятности и ее применения. 1960, т.5, вып.2, С.172–195.
6. Стратонович Р.Л. Избранные вопросы теории флуктуации в радиотехнике. М.: Сов. радио. 1961.
7. Стратонович Р.Л. Условные марковские процессы и их применения к теории оптимального управления. М., изд-во МГУ. 1966.
8. Колмогоров А.Н. Теория передачи информации. Сборник. Сессия АН СССР по научным проблемам автоматизации производства, 1956. М.: Изд-во АН СССР, 1957.
9. Хинчин А.Я. Об основных теоремах информации. – Успехи математических наук. 1956, т.11, вып.1.
10. Добрушин Р.Л. Общая формулировка основных теорем Шеннона. – Успехи математических наук. 1959, т.14, вып.6, С.3–104.
11. Котельников В.А. Теория потенциальной помехоустойчивости. М.: ГЭИ, 1956, 152 с.
12. Харкевич А.А. Очерки общей теории связи. М.: Гостехиздат, 1955.
13. Харкевич А.А. Борьба с помехами. М.: Физматгиз, 1964.
14. Харкевич А.А. Избранные труды. Т.3. М.: Наука, 1973.
15. Щукин А.Н. Об одном методе борьбы с импульсными помехами. – Известия АН СССР. Сер. физическая. 1946, т.10, № 1.
16. Сифоров В.И. Исследование одновременного замирания двух радиосигналов на коротких волнах. Труды международной конференции по ВЧ радиовещанию в Мексике, 1949. Доклад от СССР в комитете по планированию. Мехико-Сити, протокол 11А, доклад № 79-Р.
17. Сифоров В.И. О влиянии помех на прием импульсных сигналов. – Радиотехника. 1946. № 1, С.5–19.
18. Мельников В.С. Частотное радиотелеграфирование. М.: Связьиздат, 1951.
19. Мельников В.С. Вопросы теории помехоустойчивости телеграфных систем. – Доклад (материалы диссертации на степень д.т.н.) – М.: ГНИИ, 1962, 52 с.
20. Бунимович В.И. Флуктуационные процессы в радиоприёмных устройствах. М.: Советское радио, 1951.
21. Вайнштейн Л.А., Зубаков В.Д. Выделение сигналов на фоне случайных помех. М.: Советское радио, 1960.
22. Гуткин Л.С. Теория оптимальных методов радиоприема при флуктуационных помехах. М.: Госэнергоиздат, 1961, 488 с., 2-е изд. 1972, 447 с.
23. Финк Л.М. Теория передачи дискретных сообщений. М.: Советское радио, 1963, 576 с., 2-е изд., 1970.
24. Фельдбаум А.А. Основы теории оптимальных автоматических систем. М.: Физматгиз, 1963.
25. Кловский Д.Д. Передача дискретных сообщений по каналам с переменными во времени параметрами. – Реферат диссертации на соискание ученой степени к.т.н., Ленинград, ЛЭИС, 1960.
26. Кловский Д.Д., Николаев Б.И. Инженерная реализация радиотехнических схем (в системах передачи дискретных сообщений в условиях МСИ).– М.: Связь, 1975. Англ. пер.: Klovsky D., Nikolaev B. Sequential translation of digital information in the presence of intersymbol interference. – M.: Mir Publisher, 1978.
27. Левин Б.Р. Теоретические основы статистической радиотехники. М.: Сов. радио, 1-ое изд., кн. первая 1966, 728 с. кн. вторая, 1968, 503 с., кн. третья, 1976, 2-ое издание, соответственно в 1974, 1978, 1984.
28. Репин В.Г., Тартаковский Г.П. Статистический синтез при априорной неопределенности и адаптация информационных систем. М.: Сов. радио, 1977.
29. Левин Б.Р., Шинаков Ю.С. Совместно оптимальные алгоритмы обнаружения сигналов и оценивания их параметров (обзор). – Радиотехника и электроника. 1977, № 11, С.2239–2256.
30. Трифонов А.П., Шинаков Ю.С. Совместное различение сигналов и оценка их параметров на фоне помех. М.: Радио и связь, 1986, 264 с.
31. Сосулин Ю.Г. Теория обнаружения и оценивания стохастических сигналов. – М.: Сов. радио, 1978. – 320 с.
32. Блох Э.Л., Зяблов В.В. Каскадные итерированные коды и применение их для исправления пакетов ошибок. – В кн.: Передача дискретных сообщений по каналам с группирующимися ошибками. М.: Наука. 1972, т.IX, вып.4, С.3–10.
33. Блох Э.Л., Зяблов В.В. Линейные каскадные коды. – М.: Наука, 1982, 228 с.
34. Колесник В.Д., Мирончиков Е.Г. Декодирование циклических кодов. М.: Связь, 1968.
35. Коржик В.И., Осмоловский С.А., Финк Л.М. Универсальное кодирование для произвольных каналов с обратной связью. – Проблемы передачи информации.1974, № 4, С.25–29.
36. Коржик В.И., Финк Л.М. Многоступенчатое стохастическое кодирование. – Проблемы передачи информации. 1978, № 2.
37. Коржик В.И., Финк Л.М. Помехоустойчивое кодирование дискретных сообщений в каналах со случайной структурой. М.: Связь. 1975.
38. Статистическая теория связи и её практические приложения (под ред. Б.Р. Левина), М.: Связь, 1979. 287 с.
39. Шеннон К. Работы по теории информации и кибернетике. / Пер. с англ. под ред. Н.А. Железнова. – М.: ИЛ. 1963. – 829 с.
40. Левин Б.Р. Теория случайных процессов и ее применение в радиотехнике. М.: Сов. радио. 1957.
41. Пугачёв В.С. Теория случайных функций и её применение к задачам автоматического управления. М.: Госиздат, 1957.

42. Тихонов В.И. Статистическая радиотехника. М.: Радио и связь, 2-е изд., 1982, 624 с.
43. Тихонов В.И., Харисов В.Н. Статистический анализ и синтез радиотехнических устройств и систем. Учеб. пособие для вузов. М.: Радио и связь, 1991, 608 с.
44. Петрович Н.Т. Новые способы осуществления фазовой телеграфии. – Радиотехника. 1957, № 10, С.47–54.
45. Петрович Н.Т. Передача дискретной информации в каналах с фазовой манипуляцией. М.: Сов. радио, 1965, 263 с.
46. Окунев Ю.Б. Системы связи с инвариантными характеристиками помехоустойчивости. М.: Связь, 1973, № 8, 80 с.
47. Окунев Ю.Б. Теория фазоразностной модуляции. М.: Связь, 1979, 215 с.
48. Зюко А.Г. Помехоустойчивость и эффективность систем связи. М.: Связьиздат, 1983, 320 с.
49. Гинзбург В.В., Каяцкас А.А. Теория синхронизации демодуляторов. М.: Связь, 1974, 216 с.
50. Шахгильдян В.В., Ляховкин А.А. Системы фазовой автоподстройки частоты. М.: Связь, 1972.
51. Шахгильдян В.В., Лохвицкий М.С. Методы адаптивного приема сигналов. М.: Связь, 1974.
52. Колесник В.Д., Полтырев Г.Ш. Курс теории информации. М.: Наука, 1982, 416 с.
53. Зигангиров К.Ш. Некоторые последовательные процедуры декодирования. – Проблемы передачи информации. 1966, т.II, вып.4, С.13–25.
54. Зигангиров К.Ш. Процедуры последовательного декодирования. М.: Связь, 1974.
55. Кисель В.А. Синтез гармонических корректоров для высокоскоростных систем связи. – М.: Связь, 1979, – 232 с.
56. Тамм Ю.А. Адаптивная коррекция сигнала ПД. – М.: Связь, 1978. – 144 с.
57. Макаров С.Б., Цикин И.А. Передача дискретных сообщений по радиоканалам с ограниченной полосой пропускания. М.: Радио и связь. 1988, 304 с.
58. Гольденберг Л.М., Кловский Д.Д. Метод приема импульсных сигналов, основанный на использовании вычислительных машин. Труды Ленинградского института связи (ЛЭИС), Ленинград, вып.VII (44), 1959.
59. Гладышев Е.Г. Периодически коррелированные случайные последовательности. – Советская математика, № 2, С.385–388, 1961.
60. Ширман Я.Д. Разрешение и сжатие сигналов. М.: Сов. радио, 1974.
61. Вакман Д.Е. Сложные сигналы и принцип неопределённости в радиолокации. – М.: Сов. радио, 1965.
62. Вакман Д.Е. Регулярный метод синтеза ФМ сигналов. М.: Сов. радио, 1967.
63. Варакин Л.Е. Теория сложных сигналов. М.: Сов. радио, 1970.
64. Варакин Л.Е. Теория систем сигналов. М.: Сов. радио, 1978, 304 с.
65. Свердлик М.Б. Оптимальные дискретные сигналы. М.: Сов. радио, 1975.
66. Кловский Д.Д. Передача дискретных сообщений по радиоканалам – Монография. М.: Связь, 1-ое изд., 1969, 2-ое изд. 1982, 304 с.
67. Полищук Ю.М. Пространственно-временная структура случайных электромагнитных полей при распространении в тропосфере. Томск. изд. ТГУ,1975, 92 с.
68. Николаев Б.И. Последовательная передача дискретных сообщений по непрерывным каналам с памятью – М.: Радио и связь. 1988, 262 с.
69. Кловский Д.Д., Сойфер В.А. Обработка пространственно-временных сигналов (в каналах передачи информации). – М.: Связь, 1978, 207 с.
70. Кловский Д.Д., Конторович В.Я., Широков С.М. Модели непрерывных каналов связи на основе стохастических дифференциальных уравнений. – М.: Радио и связь, 1984, 248 с.
71. Кловский Д.Д. Теория передачи сигналов. Учебник для вузов. – М.: Связь, 1973, 376 с.
72. Зюко А.Г., Кловский Д.Д., Коржик В.И., Назаров М.В. Теория электрической связи (учебник для вузов под ред. Кловского Д.Д.), М.: Радио и связь, 1998, 433 с.
73. Финкельштейн Е.З. Приём дискретных сигналов при быстрых и скачкообразных изменениях параметров канала связи. Диссертация на степень к.т.н. – ЛЭИС. 1967, 180 с.
74. Кловский Д.Д. Система оптимального приема в каналах с эхо-сигналами. – Труды ЛЭИС, № 19, 1964.
75. Кловский Д.Д. Потенциальная помехоустойчивость в каналах с эхо-сигналами. – Радиотехника, т.19, № 12, 1965.
76. Кловский Д.Д. Обработка сигналов при совместной демодуляции-декодировании в каналах с межсимвольной интерференцией. Труды Международной Академии Связи, № 4 (12), 1999.
77. Карташевский В.Г. Обработка пространственно-временных сигналов в каналах с памятью. – М.: Радио и связь, 2000, 272 с.
78. Кловский Д.Д., Карташевский В.Г., Белоус С.А. Прием сигналов со свёрточным кодированием в каналах с МСИ. – Проблемы передачи информации, № 2, 1991.
79. Кловский Д.Д., Кирюшин Г.В. Энергетический выигрыш совместной демодуляции-декодирования по сравнению с раздельной демодуляцией и декодированием в многолучевых стохастических радиоканалах. – Электросвязь, № 3, 1998.
80. Кловский Д.Д., Широков С.М. Замена различения сигналов оценивания в условиях межсимвольной интерференции. Электросвязь, 1981, № 8, С.58–61.
81. Агеев Д.В. Основы теории линейной селекции. – Научно-технический сборник ЛЭИС, № 10, 1935.

СОДЕРЖАНИЕ

ПРЕДИСЛОВИЕ АВТОРА К РУССКОМУ ИЗДАНИЮ ... 5
ПРЕДИСЛОВИЕ РЕДАКТОРА РУССКОГО ПЕРЕВОДА ... 7
СПИСОК СОКРАЩЕНИЙ ... 9
1. ВВЕДЕНИЕ ... 11
1.1. ЭЛЕМЕНТЫ СИСТЕМ ЦИФРОВОЙ СВЯЗИ ... 11
1.2. КАНАЛЫ СВЯЗИ И ИХ ХАРАКТЕРИСТИКИ ... 13
Проводные каналы ... 13
Волоконно-оптические каналы ... 14
Беспроводные (радио) каналы ... 15
Подводные акустические каналы ... 18
Системы хранения информации ... 19
1.3. МАТЕМАТИЧЕСКИЕ МОДЕЛИ КАНАЛОВ СВЯЗИ ... 19
Канал с аддитивным шумом ... 19
Линейный фильтровой канал ... 20
Линейный фильтровой канал с переменными параметрами ... 20
1.4. ИСТОРИЧЕСКИЙ ОБЗОР РАЗВИТИЯ ЦИФРОВОЙ СВЯЗИ ... 21
1.5. БИБЛИОГРАФИЧЕСКИЕ ЗАМЕЧАНИЯ И ССЫЛКИ ... 25
2. ВЕРОЯТНОСТЬ И СЛУЧАЙНЫЕ ПРОЦЕССЫ ... 26
2.1. ВЕРОЯТНОСТЬ ... 26
Совместные события и совместные вероятности ... 27
Условные вероятности ... 28
Статистическая независимость ... 29
2.1.1. Случайные величины, распределение вероятностей и плотности вероятностей ... 30
Многомерные случайные величины, совместные распределения вероятностей и совместные плотности вероятностей ... 32
Условные функции распределения вероятности ... 33
Статистически независимые случайные величины ... 34
2.1.2. Функции от случайных величин ... 35
2.1.3. Статистическое усреднение случайных величин ... 38
Характеристические функции ... 39
2.1.4. Некоторые часто используемые распределения ... 41
Биномиальное распределение ... 41
Равномерное распределение ... 42
Гауссовское распределение ... 43
Хи-квадрат распределение ... 44
Релеевское распределение ... 46
Распределение Райса ... 47
m-распределение Накагами ... 48
Многомерное гауссовское распределение ... 49
2.1.5. Верхняя граница для вероятностей «хвостов» ... 51
Неравенство Чебышева ... 51
Граница Чернова ... 52
2.1.6. Суммы случайных величин и центральная предельная теорема ... 55
Центральная предельная теорема ... 57
2.2. СЛУЧАЙНЫЕ ПРОЦЕССЫ ... 59
Стационарные случайные процессы ... 59
2.2.1. Статистические средние ... 60
Средние для гауссовских процессов ... 61
Средние для совместных случайных процессов ... 61
2.2.2. Спектральная плотность мощности ... 62
2.2.3. Отклик линейной стационарной системы на случайный входной сигнал ... 63
2.2.4. Теорема отсчётов для частотно-ограниченных случайных процессов ... 66
2.2.5. Случайные сигналы и системы с дискретным временем ... 67
2.2.6. Процессы с циклической стационарностью ... 69

2.3. БИБЛИОГРАФИЧЕСКИЕ ЗАМЕЧАНИЯ И ССЫЛКИ ... 70
ЗАДАЧИ ... 70
3. КОДИРОВАНИЕ ИСТОЧНИКА ... 74
3.1. МАТЕМАТИЧЕСКИЕ МОДЕЛИ ДЛЯ ИСТОЧНИКОВ ИНФОРМАЦИИ ... 74
3.2. ЛОГАРИФМИЧЕСКАЯ МЕРА ИНФОРМАЦИИ ... 75
3.2.1. Средняя взаимная информация и энтропия ... 78
3.2.2. Измерение информации для непрерывных случайных величин ... 80
3.3. КОДИРОВАНИЕ ДЛЯ ДИСКРЕТНЫХ ИСТОЧНИКОВ ... 82
3.3.1. Кодирование для дискретных источников без памяти ... 82
Кодовые слова фиксированной длины ... 82
Теорема кодирования источника I ... 83
Кодовые слова переменной длины ... 84
Неравенство Крафта ... 85
Теорема кодирования источника II ... 86
Алгоритм кодирования Хаффмена ... 87
3.3.2. Дискретные стационарные источники ... 90
3.3.3. Алгоритм Лемпела–Зива ... 92
3.4. КОДИРОВАНИЕ ДЛЯ АНАЛОГОВЫХ ИСТОЧНИКОВ – ОПТИМАЛЬНОЕ КВАНТОВАНИЕ 94
3.4.1. Функция скорость-искажение $R(D)$ ... 94
Теорема: Функция скорость-искажение для гауссовского источника без памяти ... 95
Теорема: Кодирование источника с заданной мерой искажения ... 95
Теорема: Верхняя граница для $R(D)$ ... 96
3.4.2. Скалярное квантование ... 98
3.4.3. Векторное квантование ... 102
Алгоритм K средних ... 106
3.5. ТЕХНИКА КОДИРОВАНИЯ АНАЛОГОВЫХ ИСТОЧНИКОВ ... 108
3.5.1. Временное сигнальное кодирование ... 109
Импульсно-кодовая модуляция (ИКМ) ... 109
Дифференциальная импульсно-кодовая модуляция (ДИКМ) ... 110
Адаптивные ИКМ и ДИКМ ... 113
Дельта-модуляция (ДМ) ... 115
3.5.2. Спектральное кодирование сигнала ... 118
Кодирование подполосок ... 118
Адаптивное преобразующее кодирование ... 119
3.5.3. Модельное кодирование источника ... 119
Методы кодирования, применяемые для речевых сигналов ... 124
3.6. БИБЛИОГРАФИЧЕСКИЕ ЗАМЕЧАНИЯ И ССЫЛКИ ... 125
ЗАДАЧИ ... 125
4. ХАРАКТЕРИСТИКИ СИГНАЛОВ И СИСТЕМ СВЯЗИ ... 131
4.1. ПРЕДСТАВЛЕНИЕ ПОЛОСОВЫХ СИГНАЛОВ И СИСТЕМ ... 131
4.1.1. Представление полосовых сигналов ... 131
4.1.2. Представление линейных полосовых систем ... 134
4.1.3. Отклик полосовой системы на полосовой сигнал ... 135
4.1.4. Представление полосовых случайных процессов ... 136
Свойства квадратурных компонент ... 138
Представление белого шума ... 138
4.2. ГЕОМЕТРИЧЕСКОЕ ПРЕДСТАВЛЕНИЕ СИГНАЛОВ ... 139
4.2.1. Концепция векторного пространства ... 139
4.2.2. Концепция пространства сигналов ... 141
4.2.3. Ортогональное разложение сигналов ... 141
Процедура Грама-Шмидта ... 143
4.3. ПРЕДСТАВЛЕНИЕ СИГНАЛОВ ЦИФРОВОЙ МОДУЛЯЦИИ ... 147
4.3.1. Методы модуляции без памяти ... 148
Амплитудно-импульсная модуляция ... 148
Сигналы фазовой модуляции ... 150
Квадратурная амплитудная модуляция ... 151

Многомерные сигналы ... 152
Ортогональные многомерные сигналы ... 154
Биортогональные сигналы ... 155
Симплексные сигналы ... 156
Формы сигналов для двоичных кодов ... 156
4.3.2. Линейная модуляция с памятью ... 157
4.3.3. Нелинейные методы модуляции с памятью ... 161
Частотная модуляция с непрерывной фазой (ЧМНФ) ... 161
Модуляция с непрерывной фазой (МНФ) ... 162
Модуляция с минимальным сдвигом ... 166
Пространственные диаграммы для сигналов МНФ ... 168
Многоуровневая МНФ ... 169
4.4. СПЕКТРАЛЬНЫЕ ХАРАКТЕРИСТИКИ СИГНАЛОВ ЦИФРОВОЙ МОДУЛЯЦИИ ... 173
4.4.1. Спектр мощности сигналов линейной модуляции ... 173
4.4.2. Спектр мощности для сигналов ЧМНФ и МНФ ... 177
Спектральная плотность мощности для ЧМНФ ... 180
Спектральные характеристики МНФ ... 183
4.4.3. Спектр мощности для модулированных сигналов с памятью ... 186
4.5. БИБЛИОГРАФИЧЕСКИЕ ЗАМЕЧАНИЯ И ССЫЛКИ ... 189
ЗАДАЧИ ... 190

5. ОПТИМАЛЬНЫЕ ПРИЕМНИКИ ДЛЯ КАНАЛА С АДДИТИВНЫМ БЕЛЫМ ГАУССОВСКИМ ШУМОМ ... 197
5.1. ОПТИМАЛЬНЫЙ ПРИЕМНИК ДЛЯ СИГНАЛОВ, ПОДВЕРЖЕННЫХ ВОЗДЕЙСТВИЮ АДДИТИВНОГО БЕЛОГО ГАУССОВСКОГО ШУМА ... 197
5.1.1. Корреляционный демодулятор ... 198
5.1.2. Согласованный фильтр как демодулятор ... 201
Свойства согласованного фильтра ... 202
Интерпретация согласованного фильтра в частотной области ... 203
5.1.3. Оптимальный детектор ... 206
5.1.4. Последовательный детектор максимального правдоподобия. Алгоритм Витерби ... 210
5.1.5. Посимвольное детектирование для сигналов с памятью ... 214
5.2. ХАРАКТЕРИСТИКИ КАЧЕСТВА ОПТИМАЛЬНОГО ПРИЁМНИКА ДЛЯ МОДУЛЯЦИИ БЕЗ ПАМЯТИ ... 216
5.2.1. Вероятность ошибки при двоичной модуляции ... 216
5.2.2. Вероятность ошибки для M-позиционных ортогональных сигналов ... 219
Объединённая граница для вероятности ошибки ... 221
5.2.3. Вероятность ошибки для M-позиционной биортогональной системы сигналов ... 222
5.2.4. Вероятность ошибки для симплексных сигналов ... 223
5.2.5. Вероятность ошибки для M-позиционной системы с двоичными кодовыми сигналами ... 224
5.2.6. Вероятность ошибки для M-позиционной АМ ... 224
5.2.7. Вероятность ошибки для M-позиционной ФМ ... 226
5.2.8. Дифференциальная ФМ (ДФМ) и её характеристики качества ... 231
5.2.9. Вероятность ошибки для КАМ ... 234
5.2.10. Сравнение цифровых методов модуляции ... 238
5.3. ОПТИМАЛЬНЫЙ ПРИЁМНИК ДЛЯ СИГНАЛОВ МНФ ... 240
5.3.1. Оптимальные демодуляция и детектирование для МНФ ... 240
Расчёт метрик ... 243
5.3.2. Характеристики качества сигиалов МНФ ... 245
МНФ со многими индексами (multi-h) ... 249
5.3.3. Посимвольные детектирование сигналов МНФ ... 251
5.4. ОПТИМАЛЬНЫЙ ПРИЕМНИК ДЛЯ СИГНАЛОВ СО СЛУЧАЙНОЙ ФАЗОЙ В КАНАЛЕ С АБГШ ... 255
5.4.1. Оптимальный приемник двоичных сигналов ... 255
Оптимальный демодулятор ... 256
Оптимальный детектор ... 257
5.4.2. Оптимальный приемник для M-позиционных ортогональных сигналов ... 260

5.4.3. Вероятность ошибки при детектировании огибающей для M-позиционных ортогональных сигналов ... 262
5.4.4. Вероятность ошибки для коррелированных двоичных сигналов при детектировании огибающей ... 264
5.5. РЕГЕНЕРАТИВНЫЕ ПОВТОРИТЕЛИ И АНАЛИЗ РЕСУРСОВ ЛИНИЙ СВЯЗИ ... 266
5.5.1. Регенеративные повторители ... 266
5.5.2. Анализ ресурсов линий связи ... 268
5.6. БИБЛИОГРАФИЧЕСКИЕ ЗАМЕЧАНИЯ И ССЫЛКИ ... 271
ЗАДАЧИ ... 272

6. СИНХРОНИЗАЦИЯ НЕСУЩЕЙ И ТАКТОВАЯ СИНХРОНИЗАЦИЯ ... 282
6.1. ОЦЕНИВАНИЕ ПАРАМЕТРОВ СИГНАЛА ... 282
6.1.1. Функция правдоподобия ... 283
6.1.2. Восстановление несущей и тактовая синхронизация при демодуляции сигнала ... 284
6.2. ОЦЕНИВАНИЕ ФАЗЫ НЕСУЩЕЙ ... 287
6.2.1. МП оценка фазы несущей ... 288
6.2.2. Модели замкнутой ФАП ... 290
6.2.3. Влияние аддитивного шума на оценку фазы ... 292
6.2.4. Петли, управляемые решениями ... 295
6.2.5. Петли, не управляемые решениями ... 298
Квадратичная петля ... 301
Петля Костаса ... 302
Оценка несущей в системах с многопозиционными сигналами ... 303
Сравнение петель, управляемых и не управляемых решениями ... 305
6.3. ОЦЕНИВАНИЕ ПРИ СИНХРОНИЗАЦИИ СИМВОЛОВ ... 305
6.3.1. Максимально правдоподобная оценка параметра задержки ... 306
6.3.2 Оценивание параметра задержки, не управляемое решениями ... 307
Синхронизаторы с окнами на задержку–опережение ... 308
6.4. СОВМЕСТНОЕ ОЦЕНИВАНИЕ ФАЗЫ НЕСУЩЕЙ И ФАЗЫ СИНХРОНИЗАЦИИ ... 311
6.5. ХАРАКТЕРИСТИКИ КАЧЕСТВА МП ОЦЕНИВАТЕЛЕЙ ... 314
6.6. БИБЛИОГРАФИЧЕСКИЕ ЗАМЕЧАНИЯ И ССЫЛКИ ... 316
ЗАДАЧИ ... 317

7. ПРОПУСКНАЯ СПОСОБНОСТЬ КАНАЛА И КОДИРОВАНИЕ ... 319
7.1. МОДЕЛИ КАНАЛОВ И ПРОПУСКНАЯ СПОСОБНОСТЬ КАНАЛОВ ... 319
7.1.1. Модели канала ... 320
Двоичный симметричный канал ... 320
Дискретные каналы без памяти ... 321
Канал с дискретным входом и непрерывным выходом ... 322
Сигнальные каналы ... 322
7.1.2. Пропускная способность канала ... 323
Теоремы кодирования в канале с шумами ... 329
7.1.3. Пропускная способность канала, достигаемая при помощи ортогональных сигналов ... 329
7.1.4. Функции надёжности канала ... 331
7.2. СЛУЧАЙНЫЙ ВЫБОР КОДОВ ... 332
7.2.1. Случайное кодирование, основанное на использовании ансамбля из M двоичных кодовых слов ... 333
7.2.2. Случайное кодирование, основанное на использовании M кодовых слов с многоуровневыми сигналами ... 338
7.2.3. Сравнение предельной скорости с пропускной способностью канала с АБГШ ... 340
7.3. СИНТЕЗ СИСТЕМЫ СВЯЗИ, ОСНОВАННЫЙ НА ПРЕДЕЛЬНОЙ СКОРОСТИ ... 341
7.4. БИБЛИОГРАФИЧЕСКИЕ ЗАМЕЧАНИЯ И ССЫЛКИ ... 347
ЗАДАЧИ ... 347

8. БЛОКОВЫЕ И СВЁРТОЧНЫЕ КОДЫ ... 352
8.1. ЛИНЕЙНЫЕ БЛОКОВЫЕ КОДЫ ... 352
8.1.1. Порождающая и проверочная матрицы ... 355
8.1.2. Некоторые известные линейные блоковые коды ... 359
Коды Хемминга ... 359

Коды Адамара ... 359
Код Голея ... 360
8.1.3. Циклические коды ... 361
Кодеры для циклических кодов ... 367
Циклические коды Хемминга ... 369
Циклический код Голея (23, 12) ... 370
Коды сдвигового регистра максимальной длины ... 370
Коды Боуза-Чоудхури-Хоквингема (БЧХ) ... 372
8.1.4. Оптимальное декодирование мягких решений для линейных блоковых кодов ... 375
8.1.5. Декодирование жёстких решений ... 381
Декодирование по минимальному расстоянию (правило максимального правдоподобия) ... 381
Синдромное декодирование циклических кодов ... 384
Способность обнаружения и исправления ошибок ... 386
Вероятность ошибки при декодировании с исправлением ошибок ... 387
8.1.6. Сравнение качества декодирования жёстких и мягких решений ... 390
8.1.7. Границы для минимальных расстояний линейных блоковых кодов ... 394
8.1.8. Недвоичные блоковые коды и каскадные блоковые коды ... 397
Каскадные блоковые коды ... 399
8.1.9. Перемежение кодовых символов в каналах с пакетами ошибок ... 400
8.2. СВЁРТОЧНЫЕ КОДЫ ... 402
8.2.1. Передаточная функция свёрточного кода ... 409
8.2.2. Оптимальное декодирование для свёрточных кодов – алгоритм Витерби ... 413
8.2.3. Вероятность ошибки при декодирования мягких решений ... 416
8.2.4. Вероятность ошибки при декодировании жёстких решений ... 419
8.2.5. Дистанционные характеристики двоичных сверточных кодов ... 421
8.2.6. Недвоичные k- дуальные коды н каскадные коды ... 425
Качество k- дуальных кодов с M-ичной модуляцией ... 426
Каскадные коды ... 427
8.2.7. Другие алгоритмы декодирования сверточных кодов ... 428
8.2.8. Практические соображения по применению свёрточных кодов ... 434
8.3. КОДИРОВАННАЯ МОДУЛЯЦИЯ ДЛЯ ЧАСТОТНО-ОГРАНИЧЕННЫХ КАНАЛОВ ... 439
Решетчатая кодированная модуляция ... 444
8.4. БИБЛИОГРАФИЧЕСКИЕ ЗАМЕЧАНИЯ И ССЫЛКИ ... 453
ЗАДАЧИ ... 455

9. СИНТЕЗ СИГНАЛОВ ДЛЯ ОГРАНИЧЕННЫХ ПО ПОЛОСЕ КАНАЛОВ ... 461
9.1. ХАРАКТЕРИСТИКА ОГРАНИЧЕННЫХ ПО ПОЛОСЕ КАНАЛОВ ... 461
9.2. СИНТЕЗ СИГНАЛОВ ДЛЯ ОГРАНИЧЕННЫХ ПО ПОЛОСЕ КАНАЛОВ ... 465
9.2.1. Синтез ограниченных по полосе сигналов при отсутствии межсимвольной интерференции – критерий Найквиста ... 468
Теорема (Найквиста) ... 469
9.2.2. Синтез ограниченных по полосе символов с контролируемой МСИ – сигналы с парциальным откликом ... 473
Альтернативное представление сигналов с парциальным откликом ... 475
9.2.3. Детектирование данных при контролируемой МСИ ... 476
Посимвольное субоптимальное детектирование ... 476
Максимально-правдоподобное последовательное детектирование ... 479
9.2.4. Синтез сигналов для каналов с искажениями ... 480
9.3. ВЕРОЯТНОСТЬ ОШИБКИ ПРИ ДЕТЕКТИРОВАНИИ АМ ... 484
9.3.1. Вероятность ошибки при детектировании АМ с нулевой МСИ ... 484
9.3.2. Вероятность ошибки при детектировании сигналов с парциальным откликом ... 485
Посимвольный детектор ... 485
Последовательный детектор максимального правдоподобия ... 487
9.3.3. Вероятность ошибки для оптимальных сигналов в канале с искажением ... 487
9.4. МОДУЛЯЦИОННЫЕ КОДЫ ДЛЯ ФОРМИРОВАНИЯ СПЕКТРА ... 488
Коды с ограниченным разбегом ... 490
Отображение кодовых символов в сигналы ... 494

9.5. БИБЛИОГРАФИЧЕСКИЕ ЗАМЕЧАНИЯ И ССЫЛКИ ... 496
ЗАДАЧИ ... 497

10. СВЯЗЬ В ОГРАНИЧЕННЫХ ПО ПОЛОСЕ ЛИНЕЙНЫХ ФИЛЬТРОВЫХ КАНАЛАХ ... 502
10.1. ОПТИМАЛЬНЫЙ ПРИЕМНИК ДЛЯ КАНАЛА С МСИ И АБГШ ... 502
10.1.1. Оптимальный приемник максимального правдоподобия ... 503
10.1.2. Модель канала с МСИ с дискретным временем ... 505
10.1.3. Алгоритм Витерби для модели фильтра с дискретным временем и белым шумом ... 507
10.1.4. Качество алгоритма МППО для каналов с МСИ ... 511
10.2. ЛИНЕЙНОЕ ВЫРАВНИВАНИЕ ... 517
10.2.1. Критерий пикового искажения ... 518
Эквалайзер ограниченной длины ... 522
10.2.2. Критерий минимума среднеквадратичной ошибки (СКО) ... 522
Эквалайзер неограниченной длины ... 523
Эквалайзер ограниченной длины ... 525
10.2.3. Характеристики качества эквалайзера по минимуму СКО ... 525
Вероятность ошибки в линейном эквалайзере по минимуму СКО ... 527
10.2.4. Дробные эквалайзеры ... 530
10.3. ВЫРАВНИВАНИЕ С ОБРАТНОЙ СВЯЗЬЮ ПО РЕШЕНИЮ ... 532
10.3.1. Оптимизация коэффициентов ... 534
10.3.2. Качество эквалайзера с обратной связью по решению ЭОСР ... 534
Сравнение качества с МППО ... 537
10.3.3. Эквалайзер с обратной связью по решению и предсказаниям ... 538
10.4. БИБЛИОГРАФИЧЕСКИЕ ЗАМЕЧАНИЯ И ССЫЛКИ ... 539
ЗАДАЧИ ... 540

11. АДАПТИВНЫЕ ЭКВАЛАЙЗЕРЫ ... 546
11.1. АДАПТИВНЫЙ ЛИНЕЙНЫЙ ЭКВАЛАЙЗЕР ... 546
11.1.1. Алгоритм сведения к нулю ... 546
11.1.2. Алгоритм наименьших квадратов (НК) ... 548
11.1.3. Свойство сходимости алгоритма НК ... 552
11.1.4. Излишек СКО, обусловленный зашумленными оценками градиентов ... 553
11.1.5. Линейные эквалайзеры для базовых и полосовых сигналов ... 556
11.2. АДАПТИВНЫЙ ЭКВАЛАЙЗЕР С ОБРАТНОЙ СВЯЗЬЮ ПО РЕШЕНИЮ ... 558
11.2.1. Адаптивное выравнивание для решетчато-кодированных сигналов ... 558
11.3. АДАПТИВНЫЙ ОЦЕНИВАТЕЛЬ КАНАЛА ДЛЯ ПОСЛЕДОВАТЕЛЬНОГО МАКСИМАЛЬНО ПРАВДОПОДОБНОГО ДЕТЕКТИРОВАНИЯ ... 561
11.4. РЕКУРРЕНТНЫЕ АЛГОРИТМЫ МИНИМАЛЬНЫХ КВАДРАТОВ ДЛЯ АДАПТИВНОГО ВЫРАВНИВАНИЯ ... 563
11.4.1. Рекуррентный алгоритм (Калмана) минимальных квадратов ... 564
11.4.2. Линейное предсказание и лестничные фильтры ... 568
11.5. СЛЕПОЕ ВЫРАВНИВАНИЕ ... 571
11.5.1. Слепое выравнивание, основанное на критерии максимального правдоподобия ... 571
Оценка канала, основанная на усреднении последовательностей данных ... 572
Совместная оценка канала и данных ... 573
11.5.2. Стохастический градиентный алгоритм ... 575
Алгоритм Годарда ... 577
11.5.3. Алгоритмы слепого выравнивания, основанные на статистике сигнала второго и более высокого порядка ... 579
11.6. БИБЛИОГРАФИЧЕСКИЕ ЗАМЕЧАНИЯ И ССЫЛКИ ... 581
ЗАДАЧИ ... 582

12. МНОГОКАНАЛЬНЫЕ СИСТЕМЫ И СИСТЕМЫ С МНОГИМИ НЕСУЩИМИ ... 585
12.1. МНОГОКАНАЛЬНАЯ ЦИФРОВАЯ СВЯЗЬ В КАНАЛАХ С АБГШ ... 585
12.1.1. Двоичные сигналы ... 587
12.1.2. *M*-ичные ортогональные сигналы ... 588
12.2. СВЯЗЬ СО МНОГИМИ НЕСУЩИМИ ... 590
12.2.1. Пропускная способность неидеального линейного фиксированного канала ... 590
12.2.2. Система со многими поднесущими, основанная на быстром преобразовании Фурье ... 593

12.3. БИБЛИОГРАФИЧЕСКИЕ ЗАМЕЧАНИЯ И ССЫЛКИ ... 596
ЗАДАЧИ ... 596

13. ШИРОКОПОЛОСНЫЕ СИГНАЛЫ ДЛЯ ЦИФРОВОЙ СВЯЗИ (СИГНАЛЫ С РАССЕЯННЫМ СПЕКТРОМ) ... 598
13.1. МОДЕЛЬ ЦИФРОВЫХ СИСТЕМ СВЯЗИ С ШИРОКОПОЛОСНЫМИ СИГНАЛАМИ ... 599
13.2. ШИРОКОПОЛОСНЫЕ СИГНАЛЫ С ПРЯМЫМИ ПСЕВДОШУМОВЫМИ (ПШ) ПОСЛЕДОВАТЕЛЬНОСТЯМИ ... 601
13.2.1. Качество декодера ... 604
Выигрыш обработки и помехозащищённость ... 608
Кодирование широкополосных сигналов с ПП ... 609
Каскадное объединение произвольного линейного кода с двоичным кодом с повторением ... 611
Каскадное кодирование для широкополосных систем с ПП ... 612
13.2.2. Некоторые приложения широкополосных сигналов с ПП ... 613
Антипомеховое (АП) приложение ... 613
Скрытная передача сигналов ... 615
Кодовое разделение при множественном доступе ... 616
13.2.3. Влияние импульсной интерференции на широкополосные ПП системы ... 617
13.2.4. Генерирование ПШ последовательностей ... 623
13.3. ШИРОКОПОЛОСНЫЕ СИГНАЛЫ СО СКАЧКАМИ ЧАСТОТЫ ... 628
13.3.1. Качество широкополосных сигналов со скачками частоты (СЧ) в канале с АБГШ ... 630
13.3.2. Качество широкополосных сигналов с СЧ при парционально-полосовой интерференции 632
13.3.3. CDMA система, основанная на широкополосных сигналах с СЧ ... 638
13.4. ДРУГИЕ ТИПЫ ШИРОКОПОЛОСНЫХ СИГНАЛОВ ... 640
13.5. СИНХРОНИЗАЦИЯ ШИРОКОПОЛОСНЫХ СИСТЕМ ... 641
Отслеживание ... 644
13.6. БИБЛИОГРАФИЧЕСКИЕ ЗАМЕЧАНИЯ И ССЫЛКИ ... 647
ЗАДАЧИ ... 648

14. ЦИФРОВАЯ СВЯЗЬ ЧЕРЕЗ МНОГОПУТЕВЫЕ КАНАЛЫ С ЗАМИРАНИЯМИ ... 653
14.1. ХАРАКТЕРИСТИКИ МНОГОПУТЕВЫХ КАНАЛОВ С ЗАМИРАНИЯМИ ... 653
14.1.1. Корреляционная фуикция канала и спектр мощности ... 656
14.1.2. Статистические модели для каналов с замираниями ... 660
14.2. ВЛИЯНИЕ ХАРАКТЕРИСТИК СИГНАЛА НА ВЫБОР МОДЕЛИ КАНАЛА ... 664
14.3. КАНАЛ, НЕСЕЛЕКТИВНЫЙ ПО ЧАСТОТЕ С МЕДЛЕННЫМИ ЗАМИРАНИЯМИ ... 666
Релеевские замирания ... 667
Замирания по Накагами ... 669
Другая статистика замираний ... 670
14.4. ТЕХНИКА РАЗНЕСЕНИЯ ДЛЯ МНОГОПУТЕВЫХ КАНАЛОВ С ЗАМИРАНИЯМИ ... 670
14.4.1. Двоичные сигналы ... 671
Релеевские замирания ... 672
14.4.2. Многофазные сигналы ... 677
14.4.3. *M*-позиционные ортогональные сигналы ... 678
Граница Чернова ... 681
14.5. ЦИФРОВАЯ ПЕРЕДАЧА ПО ЧАСТОТНО-СЕЛЕКТИВНОМУ КАНАЛУ С МЕДЛЕННЫМИ ЗАМИРАНИЯМИ ... 684
14.5.1. Модель канала в виде линии задержки с отводами ... 684
14.5.2. RAKE демодулятор ... 686
14.5.3. Качество RAKE приёмника ... 687
14.6. КОДИРОВАННЫЕ СИГНАЛЫ ДЛЯ КАНАЛОВ С ЗАМИРАНИЯМИ ... 694
14.6.1. Вероятность ошибки при декодировании мягких решений и использовании двоичного блокового кода ... 696
14.6.2. Вероятность ошибки при декодировании жёстких решений и использовании линейных двоичных блоковых кодов ... 698
14.6.3. Верхние граиицы качества свёрточных кодов в канале с релеевскими замираниями ... 700
14.6.4. Использование кодов с постоянным весом и каскадных кодов для канала с замираниями ... 702

Метод 1: Нелинейные преобразования линейного кода ... 704
Метод 2: Вычёркивание ... 704
Метод 3: Матрица Адамара ... 705
Метод 4: Каскадирование ... 707
14.6.5. Синтез систем, основанный на предельной скорости ... 712
14.6.6. Решётчато-кодовая модуляция ... 717
14.7. БИБЛИОГРАФИЧЕСКИЕ ЗАМЕЧАНИЯ И ССЫЛКИ ... 719
ЗАДАЧИ ... 720

15. СИСТЕМЫ СВЯЗИ СО МНОГИМИ ПОЛЬЗОВАТЕЛЯМИ ... 726
15.1. ВВЕДЕНИЕ В ТЕХНИКУ МНОЖЕСТВЕННОГО ДОСТУПА ... 726
15.2. ПРОПУСКНАЯ СПОСОБНОСТЬ МЕТОДОВ МНОЖЕСТВЕННОГО ДОСТУПА ... 729
15.3. МНОЖЕСТВЕННЫЙ ДОСТУП С КОДОВЫМ РАЗДЕЛЕНИЕМ ... 734
15.3.1. Сигналы СДМА и модели канала ... 735
15.3.2. Оптимальный приёмник ... 736
Синхронная передача ... 736
Асинхронная передача ... 737
15.3.3. Субоптимальные детекторы ... 739
Общепринятый детектор для одного пользователя ... 739
Детектор с декорреляцией ... 740
Детектирование по минимуму среднего квадрата ошибки ... 742
Другие типы детекторов ... 743
15.3.4. Характеристики качества детекторов ... 744
15.4. МЕТОДЫ СЛУЧАЙНОГО ДОСТУПА ... 746
15.4.1. АЛОХА. Системы и протоколы ... 748
Щелевая АЛОХА ... 749
15.4.2. Системы, которые обнаруживают несущую, и протоколы ... 751
Ненастойчивый CSMA ... 752
1-настойчивый CSMA ... 752
p-настойчивый CSMA ... 752
15.5. БИБЛИОГРАФИЧЕСКИЕ ЗАМЕЧАНИЯ И ССЫЛКИ ... 756
ЗАДАЧИ ... 756

ПРИЛОЖЕНИЕ А. АЛГОРИТМ ЛЕВИНСОНА–ДУРБИНА ... 762
ПРИЛОЖЕНИЕ В. ВЕРОЯТНОСТЬ ОШИБКИ ДЛЯ МНОГОКАНАЛЬНЫХ ДВОИЧНЫХ СИГНАЛОВ ... 763
ПРИЛОЖЕНИЕ С. ВЕРОЯТНОСТИ ОШИБКИ ДЛЯ АДАПТИВНОГО ПРИЁМА M-ФАЗНЫХ СИГНАЛОВ ... 767
С.1. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ДЛЯ СИСТЕМЫ СВЯЗИ С M-ФАЗОВЫМ СИГНАЛОМ ... 767
С.2. ХАРАКТЕРИСТИЧЕСКАЯ ФУНКЦИЯ И ФУНКЦИЯ ПЛОТНОСТИ ВЕРОЯТНОСТИ ФАЗЫ 769
С.3. ВЕРОЯТНОСТИ ОШИБКИ ДЛЯ КАНАЛОВ С МЕДЛЕННЫМИ РЕЛЕЕВСКИМИ ЗАМИРАНИЯМИ ... 770
Функция распределения вероятности фазы ... 770
Вероятность ошибочного приёма символа ... 770
Вероятность ошибки двоичных символов ... 771
Вычисление коэффициентов взаимной корреляции ... 771
С.4. ВЕРОЯТНОСТЬ ОШИБКИ ДЛЯ НЕИЗМЕННЫХ ВО ВРЕМЕНИ КАНАЛОВ И КАНАЛОВ С РАЙСОВСКИМИ ЗАМИРАНИЯМИ ... 772
ПРИЛОЖЕНИЕ D. ФАКТОРИЗАЦИЯ МАТРИЦЫ (ПРИ РЕШЕНИИ СИСТЕМЫ ЛИНЕЙНЫХ УРАВНЕНИЙ) ... 774
ССЫЛКИ И БИБЛИОГРАФИЯ ... 776
СПИСОК ДОПОЛНИТЕЛЬНОЙ ЛИТЕРАТУРЫ ... 788

Производственное издание

Дж. Прокис

ЦИФРОВАЯ СВЯЗЬ

Редактор М.М. Лисина
Компьютерная верстка Б.И. Николаева
Технический редактор Т.Н. Зыкина
Корректор Т.Г. Тертышная
Обложка художника В.Г. Ситникова

ИБ № 2857

ЛР № 010164 от 29.01.97

Подписано в печать с оригинал-макета 30.11.2000 Формат 60x84/8 Гарнитура Таймс Печать офсетная
Усл. печ. л. 93,0 Усл. кр.-отт. 94,0 Уч.-изд. л. 63,12 Тираж 2000 экз. Изд. № 24147 Зак. тип. № 56
Издательство "Радио и связь", 103473, Москва, 2-й Щемиловский пер., д. 4/5
Типография издательства "Радио и связь", 103473, Москва, 2-й Щемиловский пер., д. 4/5

В издательстве "Радио и связь" вышли в свет и поступили в продажу книги:

Феер К.
Беспроводная цифровая связь.
Методы модуляции и расширения спектра
М.: Радио и связь, 2000
Пер. с англ. - 520 с., ил.
ISBN 5-256-01444-7.

Книга адресована специалистам в области связи, может быть полезна студентам вузов, обучающимся по направлению "Телекоммуникации".

В систематизированном виде рассматриваются основные виды цифровой модуляции радиосигналов и методы их демодуляции, получившие широкое распространение в беспроводных системах связи: сотовых персональных, спутниковых.

Обсуждаются возможности улучшения характеристик помехоустойчивости демодуляции за счет помехоустойчивости кодов с прямым исправлением ошибок и разнесения приема. Приводятся примеры беспроводных систем связи.

Глава 1 является введением.
В главе 2 описаны методы цифровой обработки сигналов в системах беспроводной телефонной связи и вещания.
В главе 3 изложены наиболее важные концепции в технике сотовой связи.
В главе 4 рассматривается работа модемов.
В главе 5 обсуждаются методы и способы обнаружения и исправления ошибок в цифровых сигналах.
В главе 6 описаны системы с расширением спектра.
В главе 7 освещены методы разнесенного приема.
В главе 8 даются сведения по системам спутниковой подвижной связи и вещания.
Глава 9 посвящена наиболее важным принципам и применениям интегрированных сотовых и беспроводных систем.
В конце некоторых глав приведены задачи, в тексте содержатся примеры с решениями.

К. Феер, профессор электротехники и компьютерной техники Калифорнийского университета, г. Дэйвис, руководит одной из наиболее продуктивных экспериментальных лабораторий в области исследования беспроводной цифровой связи. Он автор шести книг и большого числа статей в научно-технических журналах. Через Общество д-ра Феера (Digcom, Inc.) и Консорциум FQPSK он активно участвует в консультировании и обучении в области современных цифровых технологий связи.

Слепов Н.Н.
Технологии синхронных цифровых и оптических сетей
М.: Радио и связь, 2000. - 468 с., ил.
ISBN - 5-256-01516-8

Книга адресована специалистам в области связи и глобальных сетевых технологий, может быть полезна специалистам в области локальных и корпоративных сетей, желающих расширить свои познания в области глобальных сетевых технологий, а также - студентам вузов, обучающимся по направлению "Телекоммуникации", в качестве учебного или справочного пособия при изучении новых сетевых технологий.

Рассмотрены основные цифровые технологии современных глобальных сетей связи: PDH, SDH, SONET, ATM и WDM, а также оптические и солитонные технологии связи. Описаны основные особенности и модели указанных технологий, логические структуры фреймов, функциональные элементы, архитектура. Даны примеры и характеристики промышленных систем различных компаний. Изложены вопросы проектирования и синхронизации таких сетей. Приводятся основные особенности оптических функциональных элементов, используемых в оптоволоконных сетях: усилителей, кон-

верторов, модуляторов и коммутаторов, а также основные свойства, типы, параметры и маркировка промышленного оптического волокна и волоконно-оптических кабелей.

Краткое содержание книги Н.Н. Слепова:

Глава 1. Введение в технику передачи цифровых сигналов и технология PDH
Глава 2. Синхронные цифровые сети на основе технологии SDH
Глава 3. Основы синхронной технологии SONET
Глава 4. Радиорелейные и спутниковые системы SONET/SDH
Глава 5. Синхронизация цифровых сетей
Глава 6. Управление сетью: функционирование, администрирование и обслуживание
Глава 7. Основные элементы расчета сетей SDH
Глава 8. Введение в технологию ATM
Глава 9. Введение в оптические синхронные цифровые сети
Глава 10. Функциональные элементы оптических сетей
Глава 11. Новые технологии оптических сетей связи и перспективы их использования
Глава 12. Промышленные оптические волокна и кабели
Глава 13. Стандарты и терминология синхронных цифровых сетей
Список используемых обозначений
Список используемых сокращений
Толковый словарь терминов
Список литературы

ЗАКАЗАТЬ и ПРИОБРЕСТИ КНИГИ ИЗДАТЕЛЬСТВА «РАДИО и СВЯЗЬ» МОЖНО В МАГАЗИНАХ:

ТОО ТД «Библио Глобус»: Москва, Мясницкая ул., 6. Тел.: 928-87-44

ГУП ОЦ «Московский дом книги»: Москва, Новый Арбат, 8. Тел.: (095) 203-82-42

МКП магазин «Новый»: Москва, шоссе Энтузиастов, 24/43. Тел.: 361-98-34

Книжный магазин «Мир печати» ООО «Прессторг»: Москва, 2-я Тверская-Ямская, 54. Тел.: 978-60-22

ТОО «Молодая гвардия»: Москва, Б. Полянка, 28. Тел.: (095) 238-50-83

МКП «Дом технической книги»: Москва, Ленинский пр., 40. Тел.: (095) 137-60-19

ТОО Торговый дом «Москва»: Москва, Тверская ул., 8, стр.1. Тел.: (095) 229-66-43

АО «Центркнига»: Москва, 2-я Фрезерная, 14. Тел.: (095) 171-45-66

По всем интересующим Вас вопросам приобретения книг ***издательства «Радио и связь»*** *рекомендуем обращаться в отдел реализации и маркетинга по адресу:*

103473 Москва, 2-й Щемиловский пер., 4/5.
Тел./Факс: (095) 978-54-10, 978-53-51

ЦИФРОВАЯ

СВЯЗЬ



Профессор Джон Дж. Прокис известный специалист в области цифровой связи, ученый и педагог.

Он работает на факультете Северо-Восточного университета, начиная с сентября 1969 г., и в настоящее время является заведующим кафедрой электротехники и вычислительной техники. Ранее он работал в штате сотрудников лаборатории компании GTE (General Telephone and Electronics) и в лаборатории Линкольна Массачусетского технологического института (MTI).

Доктор Прокис получил степень бакалавра в университете Цинциннати, степень магистра в MTI и степень доктора философии по технике в Гарвардском университете. Его профессиональный опыт и интересы исследователя находятся в обширной области цифровой связи и цифровой обработки сигналов. По этим вопросам он постоянно публикуется.

Предлагаемое читателю русское издание книги Дж. Прокиса "Цифровая связь" сделано с 3-го издания книги, вышедшего в 1995 г. Первое ее издание появилось в 1983 г.

ISBN 5-256-01434-X

9 785256 014346

РАДИО И СВЯЗЬ