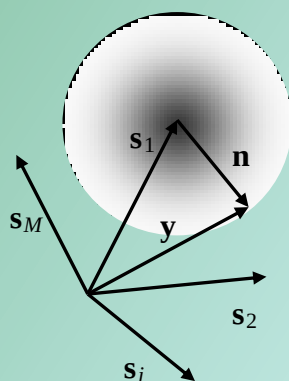


Министерство образования и науки Российской Федерации
Федеральное государственное бюджетное образовательное
учреждение высшего профессионального образования
Новгородский государственный университет
имени Ярослава Мудрого

Быстров Н.Е.

СТАТИСТИЧЕСКАЯ ТЕОРИЯ СВЯЗИ



Великий Новгород
2017 г.

УДК 621.396; (073.8)
ББК 31.811я73
Р24

Печатается по решению
РИС НовГУ

Рецензенты:

доктор технических наук, профессор
Санкт-Петербургского Электротехнического университета **В.П. Ипатов**
доктор физико-математических наук, профессор
Новгородского государственного университета **А.Ю. Захаров**

Быстров Н.Е.

Р24 . Статистическая теория связи: Учеб. пособие /
НовГУ им. Ярослава Мудрого, Великий Новгород, 2017 г. - 94 с.

В учебном пособии рассмотрены вопросы: оптимального приема и различения сигналов, эффекты распространения сигналов в каналах связи, передачи цифровой информации, оптимальные решающие правила приема сигналов, широкополосные методы борьбы с многолучевостью и оптимальный прием сигналов в каналах со случайными параметрами. Учебное пособие отвечает новым образовательным стандартам и предназначено для подготовки магистров по направлению 11.04.01 «Системы и устройства передачи, приема и обработки сигналов».

УДК 621.396; (073.8)
ББК 31.811я73

©Новгородский государственный
университет , 2017
©Н.Е Быстров, 2017

Предисловие

Главным побудительным мотивом автора пособия явилось появление нового образовательного стандарта специальности «Радиотехника», в соответствии с которым в Новгородском университете курсу «Статистическая теория связи», предшествуют дисциплины «Статистическая теория радиотехнических систем» и «Радиотехнические системы». Существующие на сегодняшний день книги учебной направленности по системам связи содержат, как правило, разделы цитированных дисциплин наряду с информацией более общего характера. В результате такие методически прекрасно написанные книги, как [1, 3], имеют весьма большой объем, который зачастую просто отпугивает студентов. Поэтому автор взял на себя смелость выбора из различных источников отдельных фрагментов и написания связывающую их часть текста, предполагая получить компактный объем настоящего пособия.

Автор выражает искреннюю признательность В.П. Ипатову за тщательное рецензирование и полезные замечания, способствовавшие улучшению настоящего пособия.

Оглавление

ВВЕДЕНИЕ.....	5
ОСНОВНЫЕ ПОНЯТИЯ И ОПРЕДЕЛЕНИЯ.....	5
1 ПРАВИЛА ОПТИМАЛЬНОГО ПРИЕМА СИГНАЛОВ	11
1.1 Постановка задачи.....	11
1.2 Оптимальные решающие правила различения сигналов	12
1.3 Оптимальные фильтры приема сигналов	16
1.4 Функция правдоподобия при различении сигналов	25
2 РАЗЛИЧЕНИЕ ДВУХ ДЕТЕРМИНИРОВАННЫХ СИГНАЛОВ	27
2.1 Определение оптимального алгоритма приема сигналов	27
2.2 Различение сигналов равных энергий.....	31
2.3 Прием когерентных двоичных сигналов	33
3 РАЗЛИЧЕНИЕ МНОЖЕСТВА ДЕТЕРМИНИРОВАННЫХ СИГНАЛОВ.....	37
3.1 Структура многоканального различителя сигналов	37
3.2 Анализ достоверности приема когерентных сигналов	38
3.3 Синтез ансамбля сигналов.....	39
4 РАЗЛИЧЕНИЕ МНОЖЕСТВА НЕКОГЕРЕНТНЫХ СИГНАЛОВ	46
4.1 Скалярное произведение некогерентных сигналов	46
4.2 Различение М-сигналов с неизвестной фазой	49
4.3 Прием некогерентных двоичных сигналов	51
4.4 Множества ортогональных сигналов	56
5 ОСНОВЫ ТЕОРИИ ПОМЕХОУСТОЙЧИВОГО КОДИРОВАНИЯ.....	65
5.1 Введение в помехоустойчивое кодирование	65
5.2 Принципы обнаружения и исправления ошибок	73
5.3 Понятие о линейных кодах и синдромное декодирование	Ошибка! Закладка не определена.
5.4 Построение ортогональных и симплексных линейных кодов	87
5.5 Систематические линейные коды.....	Ошибка! Закладка не определена.
6 МОДУЛЯЦИЯ ЦИФРОВОЙ ИНФОРМАЦИИ	92
6.1 Бинарная фазовая манипуляция сигналов	92
6.2 Приём сигналов с бинарной фазовой манипуляцией	94
6.3 Относительная бинарная фазовая манипуляция сигналов.....	96
6.4 Модуляция цифровой информации кодовой последовательностью.....	98
6.5 Помехозащищённость широкополосных систем	103
7 ЭФФЕКТЫ РАСПРОСТРАНЕНИЯ СИГНАЛОВ В КАНАЛАХ СВЯЗИ.....	111
7.1 Распространение в свободном пространстве.....	111
7.2 Затенение.....	112
7.3 Замирания вследствие многолучевого распространения	112
7.4 Разнесение.....	114
7.5 Принципы многолучевого разнесения	119
8 ПРИЕМ СИГНАЛОВ В КАНАЛАХ СО СЛУЧАЙНЫМИ ПАРАМЕТРАМИ.....	123
8.1 Импульсная характеристика канала связи	123
8.2 Корреляционный приемник при многолучевом распространении сигнала	124
8.3 Оптимальный некогерентный приемник для однолучевого канала	125
8.4 Оптимальный приемник многолучевого сигнала.	127
Случай известных запаздываний.....	127
8.5 Оптимальный приемник многолучевого сигнала.	129
Случай неизвестных запаздываний.....	129
8.6 Оптимальные приемники, зондирующие канал	132

ВВЕДЕНИЕ. ОСНОВНЫЕ ПОНЯТИЯ И ОПРЕДЕЛЕНИЯ.

Надежность передачи сообщений в реальных линиях связи зависит от помех и искажений в канале, так что проблема помехоустойчивости оказывается центральной при разработке большинства систем связи. Современная теория связи, базирующаяся на статистических методах, включает множество ветвей и направлений. Так содержание теории оптимального приема сигналов включает статистический синтез приемных устройств, анализ эффективности различных видов модуляции сигналов, расчет показателей помехоустойчивости различных систем связи и др.

Появление работ Найквиста (H.Nyquist) и Хартли (R.Hartley), посвященных исследованию канала связи и скорости передачи информации (1924, 1928 гг.) положило начало современной теории цифровой связи. Шеннон (C.Shannon) в 1947-48 гг. сформулировал основную проблему надежной передачи информации в статистических терминах, используя вероятностные модели источников и каналов связи. Шеннон также установил существование для любого канала основополагающей константы – пропускной способности, т.е. предельной скорости безошибочной передачи, зависящей от физических ресурсов, доступных передатчику, и интенсивности помех в канале. В работах Хэмминга (R.Hamming) в 1949-50 гг. впервые были разработаны методы кодирования сообщений (т.е. отображения их в последовательности некоторых символов), позволяющие обнаруживать и исправлять ошибки, вызванные канальными помехами.

Определим основные понятия и определения, наиболее часто используемые в современных системах цифровой связи.

Сообщения, знаки и символы. Источник *сообщений* может быть *аналоговым* или *дискретным*. При цифровой передаче данных *сообщение* представляет собой последовательность текстовых *знаков*, принадлежащих конечному алфавиту. Знаки могут представляться последовательностью двоичных цифр.

При передаче текстовые знаки вначале кодируются в последовательность битов, которая называется *битовым потоком*. После этого формируются группы из k бит, именуемые *символами* (*цифровое сообщение*) m_i , $i = 1...M$. Символ - это группа из k бит, рассматриваемых как единое целое, причем число всех символов конечно и равно $M = 2^k$, где k - число битов в символе.

Система, использующая символьный набор размера M , называется *M-арной*. При $k=1$ система является *бинарной*, размер набора символов $M=2$, используется один из двух различных сигналов для представления

двоичного значения «1», а другой – для представления двоичного значения «0». В этом частном случае символ и бит – это одно и то же. При $k=2$ система именуется *четверичной* или *4-х уровневой* ($M=4$).

Если информация является аналоговой, её знаковое кодирование (как в случае текстовой информации) невозможно. Вначале информацию следует перевести в цифровой формат. Для этого над сигналом осуществляется в аналого-цифровое преобразование. Исходный непрерывный сигнал дискретизируется во времени и квантуется в один из L уровней. После этого каждая квантованная выборка проходит цифровое кодирование для превращения в l -битовое ($l = \log_2 L$) кодовое слово.

Цифровой сигнал. При *низкочастотной* передаче каждый из сигналов m_i , $i=1...M$ будет представлен одним из набора видеосигналов g_i , $i=1...M$. Характеристики видеосигнала - амплитуда, длительность и положение. Для типичной *полосовой* передачи каждый видеосигнал g_i , $i=1...M$ будет представляться одним из набора полосовых радиосигналов s_i , $i=1...M$. Характеристики радиосигнала (амплитуда, частота и фаза) позволяют его идентифицировать как один из символов конечного алфавита.

Скорость передачи данных. Для цифровых систем символ m_i , $i=1, \dots, M$ посылается путем передачи цифрового сигнала g_i , $i=1...M$ в течение T секунд (T - длительность символа). Следующий символ посылается в течение следующего временного интервала, T . Скорость передачи данных (бит/с) определяется формулой

$$R = k/T = 1/T \log_2 M ,$$

где k бит определяют символ из $M = 2^k$ - символьного алфавита, а T - длительность k -битового символа.

Пропускная способность канала связи. Верхняя граница пропускной способности канала связи определяется *теоремой Шеннона-Хартли*

$$C = W \log_2 \left(1 + \frac{S}{N} \right),$$

где C - пропускная способность канала связи, бит/сек;

W - ширина полосы пропускания системы, Гц;

S - средняя мощность принятого сигнала, Вт;

N - средняя мощность шума, Вт.

Теоретически информация по каналу связи может быть передана со сколь угодно малой ошибкой при любой скорости передачи данных R , удовлетворяющей условию $R \leq C$, что достигается при помощи использования сложных методов кодирования. В случае, если $R > C$, передача данных со сколь угодно малой вероятностью ошибки невозможна.

Система передачи информации. Системой передачи информации (СПИ) называется совокупность технических средств, с помощью которых осуществляется передача сообщений от источника к получателю. Обобщенная структура СПИ приведена на рис. В.3.



Рис. В.3. Структура системы передачи информации

Источник сообщений формирует последовательность символов m_i , $i = 1...M$ в виде двоичного потока данных. Выбор величины k или M является важным первоначальным этапом проектирования любой цифровой системы связи.

Физико-электрический преобразователь - это устройство, которое преобразует сообщения источника в первичные электрические сигналы. Каждый символ m_i , $i = 1...M$, который требуется передать, преобразовывается из двоичного представления в *цифровой видеосигнал* g_i , $i = 1...M$. При использовании импульсной модуляции результирующий двоичный сигнал называется ИКМ-сигналом (ИКМ – импульсно-кодовая модуляция, в англоязычной литературе (PCM - pulse-code modulation)).

При применении импульсной модуляции к недвоичным символам результирующий сигнал именуется М-арным (М-ичным) импульсно-модулированным. Существует несколько типов подобных сигналов, основное внимание будем уделять *амплитудно-импульсной модуляции* АИМ (англ: pulse-amplitude modulation - PAM).

Модулятор осуществляет преобразование видеосигнала в радиосигнал, заключающееся в изменении одного или нескольких параметров радиосигнала в соответствии с изменением первичного

сигнала с целью наилучшего согласования с характеристиками среды распространения. При модуляции каждый видеосигнал $g_i, i=1...M$ будет представляться одним из набора полосовых сигналов $s_i, i=1...M$.

Передачик формирует сигнал в требуемом частотном диапазоне, который после соответствующего усиления через излучающую систему поступает в среду распространения.

Канал передачи. Далее сигнал $s_i t$ проходит через канал, причем связь между входным и выходным сигналами канала определяется **импульсной характеристикой канала** $h_c t$. Кроме того, в различных точках вдоль маршрута передачи дополнительные случайные шумы искажают сигнал, так что сигнал на входе приемника $y t$ отличается от переданного сигнала $s_i t$: $y(t) = s_i(t) * h_c(t) + n(t)$, $i=1,...,M$ где знак «*» представляет собой операцию свертки, а $n(t)$ - случайный процесс.

Приемник цифровой системы связи должен определить, какой из возможных M сигналов был передан. При обработке полученного сигнала в принимающем устройстве входной каскад приемника и/или демодулятор обеспечивают понижение частоты каждого полосового сигнала $y(t)$. В качестве подготовки к детектированию демодулятор восстанавливает $y t$ в виде огибающей видеосигнала $z t$. Далее демодулятор и электрофизический преобразователь осуществляют операции, обратные по отношению к модуляции и физико-электрическому преобразованию, в результате которых принимаемым сигналам сопоставляются некоторые сообщения, поступающие получателю.

Модели сигналов и каналов передачи информации. Специфика радиотехнических систем связи связана с особенностями распространения радиоволн, которая учитывается при выборе модели канала связи.

Сигнал на входе канала $S_p t, \lambda_0$ имеет параметры $\lambda_0 = \omega_0, \varphi, A$, которые, в общем случае, зависят от времени. Сигнал на выходе канала $Y_p t$ содержит искаженный входной сигнал $S_p t, \lambda_0$ и помеху $n t$

$$Y_p t = S_p t, \lambda_0 * h_c + n t = S_p t, \lambda + n t.$$

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

Канал называется *непрерывным*, если на его входе и выходе действуют непрерывные сигналы.

Непрерывные каналы можно классифицировать по виду помех и характеру преобразования $S_p(t, \lambda_0)$ в полезный принятый сигнал $S_p(t, \lambda)$. В большинстве радиотехнических систем излучаемые сигналы являются узкополосными

$$S_p(t, \lambda_0) = A(t) \cos[\omega_0 t + \varphi(t)],$$

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

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

1. Гауссовский канал - канал, в котором помеха имеет вид аддитивного нормального шума, а искажения полезного сигнала несут существенны, так как могут быть скомпенсированы. Компенсация искажений возможна, если на приемной стороне дополнительные параметры полностью известны или могут быть измерены достаточно точно (т.е. известна или доступна для измерения импульсная характеристика канала h_c). Поэтому можно считать, что

$S_p(t, \lambda_0) = S_p(t, \lambda)$, а выходной сигнал гауссовского канала есть

$$Y_p(t) = S_p(t, \lambda_0) + n(t).$$

2. Гауссовский канал с неизвестной фазой. Такой сигнал описывается выражением

$$S_p(t, \lambda_0) = A(t) \cos[\omega_0 t + \varphi(t) + \theta].$$

.Начальная фаза сигнала θ неизвестна и обычно считается равномерно распределенной в интервале $0, 2\pi$. Такой сигнал хорошо описывает процессы в линии связи на расстоянии прямой видимости.

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

$$S_p(t, \lambda_0) = \alpha(t) A(t - \tau_c) \cos[\omega_0 t + \Omega t + \varphi(t) + \theta].$$

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

Таким образом, непрерывный канал считается заданным, если определены мощность сигналов, полоса частот, дано описание моделей помех и искажений сигналов.

1 ПРАВИЛА ОПТИМАЛЬНОГО ПРИЕМА СИГНАЛОВ

1.1 Постановка задачи

Типичным для теории связи является подход, при котором анализ той или иной системы начинается с приемной стороны. Цель подобной стратегии состоит в разработке оптимального приемного устройства, которое с наилучшим качеством восстановит информацию, содержащуюся в наблюдаемом колебании.

Пусть имеется некоторый источник сообщений, вырабатывающий в каждый конкретный момент времени одно из M возможных сообщений. Каждое из M конкурирующих сообщений передается посредством специфического сигнала, так что имеется множество $S = s_i(t) : i = 1, 2, \dots, M$ из M возможных сигналов. Источник сообщений выбирает некоторый определенный сигнал $s_i(t) \in S$ и передает его на вход канала передачи информации. На приемной стороне (на выходе канала) наблюдается принятое колебание $y(t)$, которое является не точной копией переданного сигнала $s_i(t)$, а результатом трансформации $s_i(t)$, обусловленной искажающим воздействием помех, присущих любому реальному каналу. Для приемной стороны имеется M конкурирующих гипотез H_i , состоящих в том, какой из M возможных сигналов $s_i(t) \in S$ был в действительности передан и трансформирован каналом в принятое наблюдение $y(t)$, и требуется выбрать только одну из них как истинную.

Обозначим результат этого выбора, т.е. решение, через $\hat{H}_j$, означающее, что «решение принято в пользу сигнала с индексом j ». Из данной классической задачи приема вытекает следующая: *какова наилучшая стратегия принятия решения о возможном переданном сигнале, основанная лишь на наблюдении $y(t)$?*

Следует отметить, что *оптимальная процедура* принятия решения состоит в наилучшем использовании *априорной информации* и самого наблюдения. Эту информацию можно структурировать следующим образом:

- информация о сигнале (детерминированный или со случайными параметрами);
- информация о помехе (плотность вероятности или другие функциональные характеристики);
- информация об априорных вероятностях проверяемых гипотез и о последствиях, связанных с ошибочными решениями.

1.2 Оптимальные решающие правила различения сигналов

Пусть на вход канала поступает один из M сигналов, отвечающих M возможным сообщениям: $s_1(t), s_2(t), \dots, s_M(t)$. На выходе канала наблюдателю доступно колебание $y(t)$, представляющее собой переданный сигнал, искаженный канальным шумом. Располагая только колебанием $y(t)$, по какому правилу должен действовать блок решения, чтобы решение о том, какое сообщение было передано, было наиболее достоверным? Иными словами, какой из возможных алгоритмов принятия решения является оптимальным? В случае равной вероятности M входных сигналов и использовании в качестве критерия величины вероятности перепутывания передаваемых сигналов оптимальным является *правило максимального правдоподобия*: решающее устройство отдает предпочтение сигналу, наиболее «похожему» на принятое колебание.

Похожесть означает, что переходная вероятность $p(y(t)|s_i(t))$ для сигнала, в пользу которого вынесено решение, является наибольшей, иначе говоря, максимальна вероятность трансформации сигнала $s_i(t)$ в колебание $y(t)$ под действием канального шума. Таким образом, правило максимального правдоподобия (МП) можно записать в виде

$$p(y(t)|s_i(t)) \stackrel{\hat{H}_i}{\geq} p(y(t)|s_j(t)), \quad j=1, \dots, M, i \neq j, \quad (1.1)$$

где $\hat{H}_i$ – решение о том, что был передан сигнал $s_i(t)$.

В гауссовском канале переходная вероятность экспоненциально уменьшается с ростом квадрата *Евклидова расстояния* между переданным сигналом и выходным наблюдением

$$p[y(t)|s(t)] = k \exp \left[-\frac{1}{N_0} d^2(\mathbf{s}, \mathbf{y}) \right], \quad (1.2)$$

где k – константа, не зависящая от $s(t)$ и $y(t)$; N_0 – односторонняя спектральная плотность мощности белого шума.

Осуществим дискретизацию во времени непрерывного сигнала, т.е. представим сигнал $s(t)$ его отсчетами $s_i = s(iT_s)$, $i = 0, 1, \dots$, взятыми через постоянный интервал T_s . Если вся энергия сигнала сосредоточена в пределах полосы W , а $T_s \leq 1/2W$, то отсчеты s_i полностью определяют исходный непрерывный во времени сигнал $s(t)$. При длительности сигнала

T всего имеется $n = T / T_s$ подобных отсчетов и, значит, n -мерный вектор $\mathbf{s} = s_0, s_1, \dots, s_{n-1}$ полностью описывает сигнал. Прделав подобную же операцию с наблюдением $y(t)$, приходим к его n -мерному векторному эквиваленту $\mathbf{y} = y_0, y_1, \dots, y_{n-1}$, что позволяет найти Евклидово расстояние между векторами $\mathbf{s}$ и $\mathbf{y}$ согласно теореме Пифагора для n -мерного векторного пространства

$$d(\mathbf{s}, \mathbf{y}) = \sqrt{\sum_{i=0}^{n-1} [y_i - s_i]^2}. \quad (1.3)$$

Если устремить T_s к нулю, то вектора $\mathbf{s}$ и $\mathbf{y}$, являвшиеся дискретными эквивалентами сигнала и наблюдения, становятся бесконечной размерности, а значит, восстанавливают $s(t)$, $y(t)$, поскольку фактически отсутствует дискретизация во времени. В результате мы приходим к определению Евклидова расстояния для непрерывных во времени колебаний

$$d(\mathbf{s}, \mathbf{y}) = \sqrt{\int_0^T [y(t) - s(t)]^2 dt}. \quad (1.4)$$

Возвратимся вновь к правилу МП для гауссовского канала. Согласно соотношениям (1.1) и (1.2), правдоподобие сигнала (вероятность того, что он преобразован каналом в наблюдение $y(t)$) уменьшается с увеличением евклидова расстояния между $s(t)$, $y(t)$. Следовательно, правило МП для гауссовского канала может быть преобразовано в правило *минимума расстояния*

$$d(\mathbf{s}_j, \mathbf{y}) = \min_i d(\mathbf{s}_i, \mathbf{y}) \Rightarrow \hat{H}_j, \quad (1.5)$$

т.е. решение принимается в пользу сигнала $s_j(t)$, поскольку он наиболее близок (в смысле Евклидова расстояния) к наблюдению $y(t)$ среди всех M конкурирующих сигналов (см. рис.1.1).

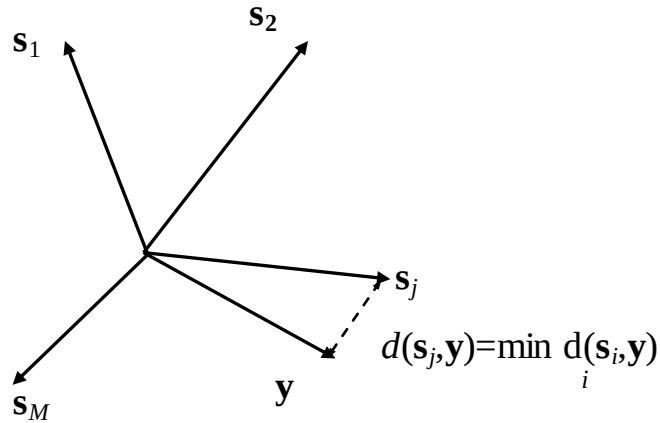


Рис. 1.1. Правило минимума расстояния.

Можно ввести длину сигнала $\|\mathbf{s}\|$ как его расстояние относительно начала координат. Тогда из (1.4) следует, что

$$\|\mathbf{s}\| = d(\mathbf{s}, \mathbf{0}) = \sqrt{E}, \quad (1.6)$$

где $E = \int_0^T s^2(t) dt$ - энергия сигнала.

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

$$(\mathbf{u}, \mathbf{v}) = \int_0^T u(t)v(t) dt, \quad (1.7)$$

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

Рассмотрим несколько иную версию правила минимального расстояния. Раскрыв скобки в (1.4), приходим к соотношению

$$d^2(\mathbf{s}_i, \mathbf{y}) = \int_0^T y^2(t) dt - 2 \int_0^T y(t) s_i(t) dt + \int_0^T s_i^2(t) dt = \|\mathbf{y}\|^2 - 2z_i + \|\mathbf{s}_i\|^2, \quad (1.8)$$

где

$$z_i = (\mathbf{y}, \mathbf{s}_i) = \int_0^T y(t) s_i(t) dt \quad (1.9)$$

корреляция между наблюдением $y(t)$ и i -м сигналом $s_i(t)$

Первое слагаемое в правой части соотношения (1.8) фиксировано для данного наблюдения и поэтому не влияет на анализируемые расстояния и решение, какой из сигналов был принят. Последний член суммы есть не что иное, как энергия i -го сигнала E_i . Учитывая это, правило минимума расстояния (1.5) может быть сформулировано как правило *максимума корреляции*

$$z_j - \frac{E_j}{2} = \max_i (z_i - \frac{E_i}{2}) \Rightarrow \hat{H}_j, \quad (1.10)$$

означающее, в частности, что из M возможных сигналов с одинаковой энергией фактически принятым считается тот, который имеет максимум корреляции с наблюдением $y(t)$.

Правило максимума корреляции поддается ясной физической трактовке. Обратимся к рис.1.2, на котором изображены сигнальные вектора.

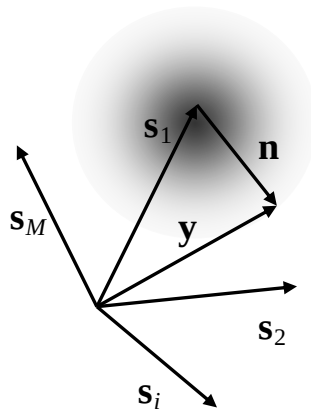


Рис. 1.2. Зашумленное наблюдение и задача выбора сигналов.

Предположим, что передавался сигнал s_1 , и что он подвергся искажению в канале, следствием чего служит добавление к s_1 вектора шума $\mathbf{n}$. Гауссовский вектор $\mathbf{n}$ характеризуется симметричным (сферическим) вероятностным распределением, экспоненциально спадающим с увеличением длины вектора $\mathbf{n}$, что очевидным образом следует из (1.2) после удаления из него сигнала (т.е. при подстановки $s(t) = 0$). Следовательно, вектор наблюдения $\mathbf{y} = \mathbf{s}_1 + \mathbf{n}$ будет случайным

образом перемещаться вокруг s_1 , как это показано на рисунке, и тогда, согласно правилу минимума расстояния (1.5), как только y окажется ближе к некоторому другому, чем s_1 , сигналу, то будет принято ошибочное решение. Для минимизации ошибок подобного рода следует располагать другие сигналы на максимально большом расстоянии от s_1 . Поскольку любой из M сигналов может передаваться равновероятно, т.е. занимать место s_1 , то, очевидно, что все расстояния $d(s_i, s_j), 1 \leq i < j \leq M$, следует делать максимально большими.

В случае достаточно большого значения M задача одновременной максимизации всех расстояний не представляется простой, поскольку они могут конфликтовать друг с другом.

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

1.3 Оптимальные фильтры приема сигналов

При решении задач оптимального приема сигналов отношение сигнал/шум (с/ш), играют ключевую роль. Под отношением с/ш на выходе линейного фильтра понимают отношение модуля значения выходного сигнала в момент t_0 к среднеквадратическому значению помехи на выходе в этот же момент времени

$$c/ш_{\text{вых}} = \frac{|s_{\text{вых}}(t_0)|}{\sqrt{D x_{\text{вых}}(t_0)}} .$$

Будем полагать, что помеха (шум) является стационарным случайным процессом. Спектральная плотность помехи $X(\omega)$ ограничена по частоте и характеризуется корреляционной функцией K_{t_1, t_2} .

Пусть фильтруемый сигнал $s(t)$ задан своей спектральной плотностью

$$\tilde{s}(j\omega) = |\tilde{s}(\omega)| \cdot \exp \left[j\varphi_s \omega \right] .$$

Частотную характеристику синтезируемого фильтра представим в показательной форме

$$K(j\omega) = |K(\omega)| \cdot \exp[j\varphi_K(\omega)],$$

где: $|K(\omega)|$ - амплитудно-частотная и $\varphi_K(\omega)$ - фазочастотная характеристика фильтра.

Для отыскания оптимальной частотной характеристики синтезируемого фильтра составим выражения для сигнала и шума на его выходе. Используя спектральный метод анализа, можно найти полезный сигнал на выходе фильтра в любой момент времени t_0

$$\begin{aligned} s_{\text{ВЫХ}}(t_0) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{s}(j\omega) \cdot K(j\omega) \exp(j\omega \cdot t_0) d\omega = \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} |\tilde{s}(\omega)| \cdot |K(\omega)| \cdot \exp[j\omega \cdot t_0 + \varphi_s(\omega) + \varphi_K(\omega)] d\omega. \end{aligned}$$

Среднеквадратическое значение шума на выходе фильтра определяется выражением

$$\sigma_{\text{ВЫХ}} = \sqrt{\frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega) \cdot |K(\omega)|^2 d\omega}.$$

При использовании частотного метода анализа отношение сигнал/шум представляется выражением

$$\begin{aligned} q_{\text{ВЫХ}} &= \frac{\left| \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{s}(\omega) K(j\omega) \exp(j\omega t_0) d\omega \right|}{\sqrt{\frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega) |K(j\omega)|^2 d\omega}} = \\ &= \frac{\left| \frac{1}{2\pi} \int_{-\infty}^{\infty} |\tilde{s}(\omega)| \cdot |K(\omega)| \exp[j\omega t_0 + \varphi_s(\omega) + \varphi_K(\omega)] d\omega \right|}{\sqrt{\frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega) |K(j\omega)|^2 d\omega}}. \quad (1.11) \end{aligned}$$

Задача нахождения оптимальной функции $K(j\omega)$ решается на основании известного неравенства Коши-Буняковского. Для того чтобы получить максимальный уровень сигнала, необходимо скомпенсировать начальные фазы спектральных компонент сигнала. Это достигается, если ФЧХ фильтра будет противоположно по знаку фазовому спектру сигнала. В этом случае, после прохождения через фильтр спектр выходного сигнала будет иметь фазовую характеристику, определяемую выражением

$$\varphi_{S \text{ Вых}}(\omega) = \varphi_S(\omega) + \varphi_K(\omega) = \varphi_S(\omega) + -\varphi_S(\omega) - \omega \cdot t_0 = -\omega \cdot t_0.$$

Как видно, фазовая характеристика оптимального фильтра $\varphi_K(\omega)$ обеспечивает когерентное сложение скорректированных по фазе всех компонентов спектра в момент $t = t_0$ и в результате этого образуется пик выходного сигнала.

Исследование функции (1.11) показывает, что максимально достижимое отношение с/ш определяется выражением

$$C_{\text{max Вых}} = \sqrt{\frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{|\tilde{s}(\omega)|^2}{X(\omega)} d\omega} \quad (1.12)$$

и достигается с помощью фильтра, имеющего коэффициент передачи

$$K_{\text{опт}}(j\omega) = c \cdot \frac{|\tilde{s}(\omega)|}{X(\omega)} \cdot \exp -j\varphi_S(\omega) \cdot \exp -j\omega \cdot t_0 = c \frac{\tilde{s}^*(\omega)}{X(\omega)} \exp -j\omega t_0, \quad (1.13)$$

где c – коэффициент, не влияющий на отношение с/ш, и в дальнейшем полагается равным единице.

Сигнал на выходе оптимального фильтра равен

$$s_{\text{Вых}}(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{|\tilde{s}(\omega)|^2}{X(\omega)} \exp j\omega(t - t_0) d\omega. \quad (1.14)$$

Автокорреляционная функция сигнала определяется преобразованием Фурье энергетического спектра сигнала

$$K_s(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} |\tilde{s}(\omega)|^2 \exp j\omega\tau d\omega.$$

Таким образом, амплитудно-частотная характеристика (АЧХ) оптимального фильтра в соответствии с (1.13) равна

$$\left| K_{\text{опт}}(j\omega) \right| = c \cdot \frac{|\tilde{s}(\omega)|}{X(\omega)}, \quad (1.15)$$

а фазо-частотная характеристика фильтра определяется соотношением

$$\arg K_{\text{опт}}(j\omega) = -\varphi_s(\omega) - \omega t_0. \quad (1.6)$$

Как видно, АЧХ оптимального фильтра производит весовую обработку спектра входного колебания, придавая большие веса тем спектральным участкам, на которых больше отношение $|\tilde{s}(\omega)|/X(\omega)$, что условно можно считать отношением с/ш на частоте ω . ФЧХ оптимального фильтра обеспечивает синфазное суммирование спектральных компонент сигнала в момент времени $t = t_0$. Предельно достижимое отношение сигнал/шум в соответствии с (1.12) определяется интегрированием отношения энергетического спектра сигнала $|\tilde{s}(\omega)|^2$ и спектральной плотности мощности помех $X(\omega)$ по всем частотам, на которых это отношение определено.

Так как спектральная плотность помехи на выходе оптимального фильтра равна

$$X_{\text{вых}}(\omega) = X(\omega) |K_{\text{опт}}(j\omega)|^2 = X(\omega) \frac{|\tilde{s}(\omega)|^2}{X^2(\omega)} = \frac{|\tilde{s}(\omega)|^2}{X(\omega)},$$

то корреляционная функция помехи на выходе фильтра определяется обратным преобразованием Фурье

$$K_{\text{вых}}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{|\tilde{s}(\omega)|^2}{X(\omega)} \exp(j\omega\tau) d\omega. \quad (1.17)$$

Импульсная характеристика оптимального фильтра может быть выражена через частотный коэффициент передачи

$$h_{\text{опт}}(t, t_0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\tilde{s}^*(\omega)}{X(\omega)} \exp(j\omega(t-t_0)) d\omega. \quad (1.18)$$

На основании равенства Парсеваля выражение (1.8) можно записать как свертку обратной корреляционной функцией помехи $K^{-1}(\tau_1 - \tau_2) = F[1/X(\omega)]$ (где F – оператор Фурье) и сигнала $s(t_0 - t)$, т.е.

$$h_{\text{опт}}(t, t_0) = \int_{-\infty}^{\infty} K^{-1}(\tau - t) \cdot s(t_0 - \tau) d\tau. \quad (1.19)$$

Так как преобразование Фурье энергетического спектра сигнала $|\tilde{s}(\omega)|^2$ есть его автокорреляционная функция $K_s(\tau)$, то сигнал на выходе есть свертка обратной корреляционной функции помехи $K^{-1} \tau$ и автокорреляционной функции сигнала, смещенной на время t_0

$$s_{\text{вых}}(t) = \int_{-\infty}^{\infty} K^{-1}(\tau) \cdot K_s(t_0 - \tau) d\tau. \quad (1.20)$$

Согласованный фильтр. Пусть помехой является белый шум $x(t) = n(t)$ со спектральной равномерной плотностью $X(\omega) = \frac{N_0}{2}$. Оптимальный фильтр для помехи в виде белого шума называется согласованным фильтром (СФ). В этом случае отношение сигнал/шум, исходя из формулы (1.12), принимает вид

$$q_{\text{max вых}} = \sqrt{\frac{2}{N_0} \frac{1}{2\pi} \int_{-\infty}^{\infty} |\tilde{s}(\omega)|^2 d\omega} = \sqrt{\frac{2E}{N_0}}. \quad (1.21)$$

Коэффициент передачи оптимального фильтра для помехи в виде белого шума, с учетом замечания о постоянном множителе, равен

$$K_{\text{сф}}(j\omega) = \tilde{s}^*(\omega) \exp -j\omega t_0. \quad (1.22)$$

АЧХ согласованного фильтра совпадает с амплитудным спектром сигнала, т.е.

$$|K_{\text{сф}}(j\omega)| = |\tilde{s}(\omega)|, \quad (1.23)$$

а ФЧХ по-прежнему определяется выражением

$$\arg K_{\text{опт}}(j\omega) = -\varphi(\omega) - \omega t_0.$$

Так как корреляционная функция белого шума равна $K(\tau) = N_0/2\delta(\tau)$, то импульсная характеристика с учетом выражения (1.19) примет вид

$$h_{\text{сф}}(t, t_0) = s(-t + t_0). \quad (1.24)$$

Построение графика функции $h_{\text{сф}}(t, t_0)$ показано на рис. 1.1.

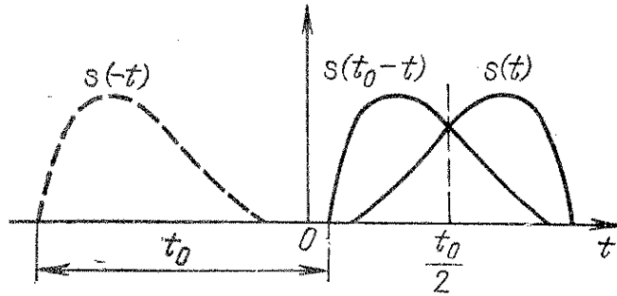


Рис. 1.1. Импульсная характеристика согласованного фильтра.

Кривая $s(-t)$ является зеркальным отражением заданного сигнала $s(t)$ с осью ординат в качестве оси симметрии. Функция же $s(t_0 - t)$, сдвинутая относительно $s(-t)$ на время t_0 вправо, также зеркальна по отношению к исходному сигналу $s(t)$, но с осью симметрии, проходящей через точку $t_0/2$ на оси абсцисс. Очевидно, что только при $t_0 \geq T_c$ может быть использована вся энергия сигнала для создания наибольшего возможного пика выходного сигнала. Ясно, что увеличение t_0 сверх T_c не влияет на пиковое значение выходного сигнала, а просто сдвигает его вправо (в сторону запаздывания).

Сигнал на выходе согласованного фильтра в соответствии с (1.20) равен

$$s_{\text{ВЫХ}}(t) = K_s [s(t - t_0)] \cdot \quad (1.25)$$

Следовательно, в момент времени $t = t_0$ пиковое значение выходного сигнала равно $s_{\text{макс}}(t_0) = R_s \cdot 0 = E$.

Спектральная плотность шума на выходе согласованного фильтра равна

$$X_{\text{ВЫХ}}(\omega) = \frac{N_0}{2} |K_{\text{сф}}(j\omega)|^2 = \frac{N_0}{2} |\tilde{s}(\omega)|^2 \quad (1.26)$$

Дисперсия шума на выходе равна

$$D\{x_{\text{ВЫХ}}\} = \frac{1}{2\pi} \int_{-\infty}^{\infty} X_{\text{ВЫХ}}(\omega) d\omega = \frac{N_0}{2} \frac{1}{2\pi} \int_{-\infty}^{\infty} |\tilde{s}(\omega)|^2 d\omega = \frac{N_0 E}{2}, \quad (1.27)$$

а корреляционная функция описывается выражением

$$\begin{aligned} K_{\text{ВЫХ}}(\tau) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} X_{\text{ВЫХ}}(\omega) \exp j\omega\tau d\omega = \\ &= \frac{N_0}{2} \frac{1}{2\pi} \int_{-\infty}^{\infty} |\tilde{s}(\omega)|^2 \exp j\omega\tau d\omega = \frac{N_0}{2} K_s(\tau) \end{aligned} \quad (1.28)$$

и по форме совпадает с автокорреляционной функцией сигнала. Совпадение форм корреляционной функции помехи на выходе СФ и автокорреляционной функции сигнала говорит о том, что помеха на выходе СФ становится «похожей» на сигнал $s(t)$ с точки зрения скорости изменения.

Так как $K_s(0) = E$, то и отношение с/ш на выходе СФ будет равно

$$q = E / \sqrt{\frac{N_0 E}{2}} = \sqrt{\frac{2E}{N_0}}, \quad (1.29)$$

что совпадает с выражением (1.21).

Корреляционный приемник. Сравним фильтр, максимизирующий отношение с/ш, с коррелятором, вычисляющим величину

$$z = \int_{-\infty}^{t_0} y(t) \cdot s_{\text{оп}}(t) dt,$$

где $y(t) = n(t) + s(t)$ принятая реализация; $n(t)$ - шум; $s(t)$ - полезный сигнал; $s_{\text{оп}}(t)$ - опорное колебание, являющееся детерминированной функцией времени; t_0 - момент времени взятия отсчета.

Так как коррелятор является линейной системой с переменными параметрами, то мы можем отдельно рассматривать прохождение через него сигнала $s(t)$ и шума $n(t)$.

Преобразование сигнала определяется выражением

$$s_{\text{ВЫХ}}(t_0) = \int_{-\infty}^{t_0} s(t) \cdot s_{\text{ОП}}(t) dt. \quad (1.30)$$

СКО случайной величины, вызванной действием шума $n(t)$ равно

$$\sigma_{\text{ВЫХ}}(t_0) = \left[\int_{-\infty}^{t_0} \int_{-\infty}^{t_0} K(t_1, t_2) s_{\text{ОП}}(t_1) s_{\text{ОП}}(t_2) dt_1 dt_2 \right]^{1/2}. \quad (1.31)$$

Отношение сигнал/шум на выходе коррелятора будет определяться выражением

$$q_{\text{ВЫХ}} = \frac{s_{\text{ВЫХ}}}{\sigma_{\text{ВЫХ}}} = \frac{\int_{-\infty}^{t_0} s(t) \cdot s_{\text{ОП}}(t) dt}{\left[\int_{-\infty}^{t_0} \int_{-\infty}^{t_0} K(t_1, t_2) s_{\text{ОП}}(t_1) s_{\text{ОП}}(t_2) dt_1 dt_2 \right]^{1/2}}. \quad (1.32)$$

Максимизируя отношение (1.32) за счет выбора $s_{\text{ОП}} t$, получим следующее выражение для оптимального опорного сигнала

$$s_{\text{ОПТ}}(t) = s(t). \quad (1.33)$$

Если момент времени t_0 соответствует окончанию сигнала, то значение сигнала на выходе будет равно

$$s_{\text{ВЫХ}}(t_0) = \int_{-\infty}^{t_0} s^2(t) dt = E.$$

Нахождение дисперсии шума на выходе коррелятора сводится к вычислению выражения

$$\sigma^2 = D\{z\} = \overline{\left\{ \int_{-\infty}^{t_0} n(t) s(t) dt \right\}^2} = \int_{-\infty}^{t_0} \int_{-\infty}^{t_0} \overline{n(t_1) n(t_2)} s(t_1) s(t_2) dt_1 dt_2. \quad (1.34)$$

Напомним, что автокорреляционная функция белого шума (статистическое среднее произведения двух различных временных отсчетов) есть дельта-функция Дирака $\overline{n(t_1) n(t_2)} = (N_0 / 2) \delta(t_1 - t_2)$. Другими словами, любые два отсчета белого шума, не совпадающие во

времени, являются некоррелированными. Используя последний результат в выражении для дисперсии, а также *фильтрующее* свойство дельта-функции $\int_0^T s(t')\delta(t'-t)dt' = s(t)$, приходим к выражению

$$\sigma^2 = \frac{N_0}{2} \int_{-\infty}^{t_0} s^2(t)dt = \frac{N_0 E}{2}. \quad (1.35)$$

Таким образом, для отношения с/ш получим величину

$$q_{\text{max вых}} = \sqrt{\frac{2E}{N_0}}, \quad (1.36)$$

совпадающую с отношением с/ш на выходе СФ.

Если опорное напряжение коррелятора отличается от оптимального, то возникают потери. Так для белого шума сигнал на выходе коррелятора будет равен

$$s_{\text{вых}}(t_0) = \int_{-\infty}^{t_0} s(t)s_{\text{оп}}(t)dt = \rho(t_0)\sqrt{EE_{\text{оп}}(t_0)}, \quad (1.37)$$

где $\rho(t_0)$ - коэффициент корреляции сигнала $s(t)$ и опорного колебания $s_{\text{оп}}(t)$.

Дисперсия шума на выходе коррелятора будет определяться выражением

$$\sigma_{\text{вых}}(t_0) = \sqrt{\frac{N_0 E_{\text{оп}}(t_0)}{2}}. \quad (1.38)$$

и отношение сигнал/шум примет вид

$$q_{\text{вых}} = \rho(t_0)\sqrt{\frac{2E}{N_0}}. \quad (1.39)$$

В завершении следует отметить, что хотя СФ инвариантен к временному положению сигнала, для реализации его возможностей надо знать момент времени, в который достигается максимальное отношение с/ш на выходе, т.е. знать временное положение сигнала. Аналогичную информацию надо иметь для реализации возможностей коррелятора.

Поэтому с точки зрения обработки сигнала СФ и коррелятор равноценны, и решение в пользу той или иной структуры следует принимать, исходя из возможностей технической реализации.

1.4 Функция правдоподобия при различении сигналов

Дальнейшее рассмотрение процедур различения сигналов будет вестись применительно к помехе $x(t) = n(t)$ в виде нормального шума с известной спектральной плотностью N_0 .

При отсутствии сигнала колебание $y(t)$, принятое на конечном интервале времени T , представляет собой случайный процесс $y(t) = x(t)$.

Рассмотрим метод дискретного наблюдения, когда отсчеты $y_i = y(t_i)$ берутся через равностоящие моменты времени $t_i - t_{i-1} = \Delta$. Определим сначала совместную плотность вероятности для случайных величин $y_i = x_i = x(t_i)$, которые имеют следующие характеристики: $\langle x_i \rangle = 0$, $\sigma_i^2 = \langle x_i^2 \rangle = N_0/2 \cdot \Delta$ и $\langle x_i \cdot x_j \rangle = 0$. Шум имеет нормальный закон распределения

$$W(x_i) = \frac{1}{\sqrt{2\pi\sigma_i}} \exp\left(-\frac{x_i^2}{2\sigma_i^2}\right). \quad (1.40)$$

В силу независимости отсчетов совместная плотность вероятности примет вид

$$W(x_1, x_2, \dots, x_m) = \prod_{i=1}^m W(x_i) = \left(\pi \frac{N_0}{\Delta}\right)^{-\frac{m}{2}} \exp\left(-\frac{1}{N_0} \sum_{i=1}^m x_i^2 \Delta\right). \quad (1.41)$$

Чтобы перейти к случаю непрерывного наблюдения, нужно $\Delta \rightarrow 0$. Тогда совместная плотность вероятности перейдет в соответствующий функционал вероятности

$$W(x(t)) = c \cdot \exp\left(-\frac{1}{N_0} \int_0^T x^2(t) dt\right), \quad (1.42)$$

где c - некоторый несущественный коэффициент одинаковый для всех реализаций.

При различении M детерминированных сигналов на фоне аддитивного шума гипотеза H_i означает, что $y(t) = x(t) + s_i(t)$. Подставляя в функционал ПВ гауссовского шума $x(t) = y(t) - s_i(t)$, получаем выражение ФП вида

$$W(y(t) | H_i) = c \cdot \exp \left(-\frac{1}{N_0} \int_0^T [y(t) - s_i(t)]^2 dt \right). \quad (1.43)$$

Приведенное выражение позволяет дать наглядную интерпретацию правила различения сигналов - для данной реализации $y(t)$ принимают решение о присутствии в ней того сигнала $s_i(t)$, который наименее уклоняется от $y(t)$.

2 РАЗЛИЧЕНИЕ ДВУХ ДЕТЕРМИНИРОВАННЫХ СИГНАЛОВ

2.1 Определение оптимального алгоритма приема сигналов

Определение оптимального алгоритма обработки, базирующегося на учете специфических свойств переданного сигнала, позволяет синтезировать оптимальным образом и сам переданный сигнал, т.е. выбрать наилучшим образом методы его модуляции. В данной главе будет исследована связь между классическими задачами приема и синтеза сигналов. Под классическими задачами здесь понимаются такие, которые базируются на использовании традиционной модели гауссовского канала.

Строгая зависимость качества приема от расстояния между сигналами наиболее наглядно проявляется в простейшей, но очень типичной связной задаче *передачи двоичной информации*, когда по каналу пересылается только одно из $M = 2$ возможных сообщений. Фактически данная ситуация отвечает передаче одного бита данных в системе цифровой связи.

Обозначая сообщения через нуль и единицу и полагая, что для их передачи используются детерминированные сигналы $s_0(t)$ и $s_1(t)$, правило решения по минимуму расстояния может быть представлено в виде

$$\begin{array}{c} \hat{H}_0 \\ d(s_0, y) < \\ > \\ \hat{H}_1 \end{array} d(s_1, y). \quad (2.1)$$

Аналогичным образом может быть переписано и решающее правило, базирующееся на учете корреляции

$$\begin{array}{c} \hat{H}_0 \\ z = z_0 - z_1 > \frac{E_0 - E_1}{2}, \\ < \\ \hat{H}_1 \end{array} \quad (2.2)$$

где $z_i = \int_0^T y(t)s_i(t)dt$, $i = 0, 1$ - корреляция каждого из сигналов и наблюдения $y(t)$; $E_i = \|s_i\|^2$, $i = 0, 1$ - энергия сигналов.

Оптимальные правила различения (2.1) и (2.2) двух сигналов могут быть наглядно объяснены с помощью их геометрической интерпретации. Два

сигнальных вектора s_0 и s_1 всегда лежат в сигнальной плоскости SP . Вектор наблюдения y обязательно попадет на эту плоскость, но близость его к одному или другому сигналу определяется близостью к ним проекции y' вектора y на плоскости SP (см. рис. 2.1, а).

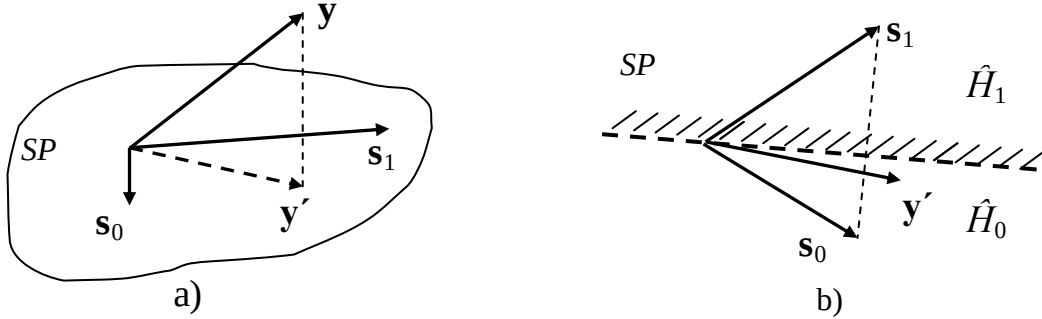


Рис. 2.1. Сигнальная плоскость и полуплоскости решения.

Плоскость SP может быть поделена на две полуплоскости прямолинейной границей, перпендикулярной отрезку, соединяющему сигнальные векторы, и делит его пополам. Тогда решения $\hat{H}_0$ и $\hat{H}_1$ фактически определяются как попадание проекции y' на соответствующую полуплоскость (см. рис. 2.1, б).

Следовательно, вероятность перепутывания сигналов (вероятность ошибки) зависит от расстояния между векторами s_0 и s_1 в сравнении с диапазоном случайных флуктуаций y' , обусловленных канальным шумом. Согласно (2.2), фактически принятый сигнал $s_0(t)$ будет ошибочно засчитан за $s_1(t)$, если и только если разность корреляций будет меньше порога $(E_0 - E_1)/2$. Следовательно, вероятность P_{01} подобной ошибки может быть найдена как

$$p_{01} = P\left(z < \frac{E_0 - E_1}{2} \middle| s_0(t)\right) = \int_{-\infty}^{\frac{E_0 - E_1}{2}} W(z | s_0(t)) dz, \quad (2.3)$$

где $W(z | s_0(t))$ - условная плотность вероятности разности корреляций z из (2.2) при условии, что сигнал $s_0(t)$ действительно был принят.

Поскольку z получено как линейное преобразование гауссовского наблюдения $y(t)$, то z также подчиняется гауссовскому распределению

$$W(z | s_0(t)) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(z - \bar{z})^2}{2\sigma^2}\right], \quad (2.4)$$

интегрирование которого, согласно (2.3), приводит к выражению

$$p_{01} = Q\left(\frac{2\bar{z} - E_0 + E_1}{2\sigma}\right), \quad (2.5)$$

где $Q(x) = 1 - \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\left(-\frac{y^2}{2}\right) dy$ - дополнительная функция ошибок; $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{y^2}{2}\right) dy$ - интеграл вероятности.

Для раскрытия выражения (2.5) найдем математическое ожидание $\bar{z}$, обусловленное принятым сигналом и дисперсию $\sigma^2 = D\{z\}$.

В случае, когда действительно был принят сигнал $s_0(t)$, т.е. $\overline{y(t)} = s_0(t)$, математическое ожидание будет равно

$$\bar{z} = \int_0^T \overline{y(t)} [s_0(t) - s_1(t)] dt = E_0 - \rho_{01} \sqrt{E_0 E_1}, \quad (2.6)$$

где

$$\rho_{ij} = \frac{(s_i, s_j)}{\|s_i\| \|s_j\|} = \frac{1}{\sqrt{E_i E_j}} \int_0^T s_i(t) \cdot s_j(t) dt \quad (2.7)$$

- коэффициент корреляции сигналов $s_i(t)$, $s_j(t)$, а E_i, E_j - энергии сигналов.

Из (2.7) следует, что в геометрической интерпретации ρ_{01} есть просто косинус угла между сигналами $s_0(t)$, $s_1(t)$ (или сигнальными векторами s_0, s_1), и, значит, характеризует близость, или *подобие* (сходство) сигналов.

При нахождении дисперсии σ^2 учтем тот факт, что на нее не оказывает влияние детерминированный компонент наблюдения $y(t)$ (в рассматриваемом случае сигнал $s_0(t)$), поскольку шум в канале аддитивен. Фактически, тогда можно удалить сигнал из выражения для $y(t)$, полагая $y(t) = n(t)$, где $n(t)$ - белый шум с двусторонней спектральной плотностью мощности $N_0/2$. После приведенных рассуждений приходим к выражению

$$\sigma^2 = \frac{N_0}{2} \int_0^T s^2(t) dt = \frac{N_0 E}{2}. \quad (2.11)$$

В рассматриваемой ситуации в (2.11) следует подставить $s(t) = s_0(t) - s_1(t)$, т.е. E представляет собой энергию E_d разностного сигнала $s_0(t) - s_1(t)$.

Учет последнего дает

$$E_d = \int_0^T [s_0(t) - s_1(t)]^2 dt = d^2(\mathbf{s}_0, \mathbf{s}_1) = E_0 + E_1 - 2\rho_{01}\sqrt{E_0 E_1}. \quad (2.12)$$

Окончательно, подстановка (2.6), (2.11) и (2.12) в (2.5) дает

$$p_{01} = Q\left(\sqrt{\frac{d^2(\mathbf{s}_0, \mathbf{s}_1)}{2N_0}}\right). \quad (2.13)$$

Поскольку задача нахождения вероятности ошибочного приема симметрична, то аналогичный результат будет получен и для вероятности p_{10} перепутывания $s_1(t)$ с $s_0(t)$. Тогда полная (безусловная) вероятность ошибочного приема P_e не зависит априори от вероятности p_0 p_0 поступления в канал сигнала $s_0(t)$ и устанавливается выражением

$$P_e = p_0 p_{01} + (1 - p_0) p_{10} = Q\left(\sqrt{\frac{d^2(\mathbf{s}_0, \mathbf{s}_1)}{2N_0}}\right). \quad (2.14)$$

Из (2.14) очевидным образом вытекает, что единственным путем достижения высокой достоверности передачи данных является увеличение расстояния $d(\mathbf{s}_0, \mathbf{s}_1)$ между сигналами до максимально возможного значения. Ясно, что увеличение $d(\mathbf{s}_0, \mathbf{s}_1)$ может быть достигнуто за счет большей энергии сигналов, или длины соответствующих векторов. Однако какова будет *оптимальная* пара сигналов, если ресурс прямого решения лимитирован, т.е. энергия сигналов ограничена заранее?

2.2 Различение сигналов равных энергий

Рассмотрим важный случай различения двух сигналов равных энергий $E_0 = E_1 = E$. В этом случае правило принятия решения (2.2) сводится к сравнению значений корреляций z_0 и z_1 или, что эквивалентно, определению знака их разности

$$z = z_0 - z_1 \underset{\hat{H}_1}{\overset{\hat{H}_0}{>}} 0. \quad (2.15)$$

Таким образом, чтобы различить два детерминированных сигнала, достаточно вычислить единственную корреляцию $z = \int_0^T y(t) \cdot s(t) dt$ с разностным сигналом $s(t) = s_0(t) - s_1(t)$, причем какое из решений будет принято, зависит только от знака z . Структура, реализующая данное правило и использующая коррелятор, показана на рис. 2.2.

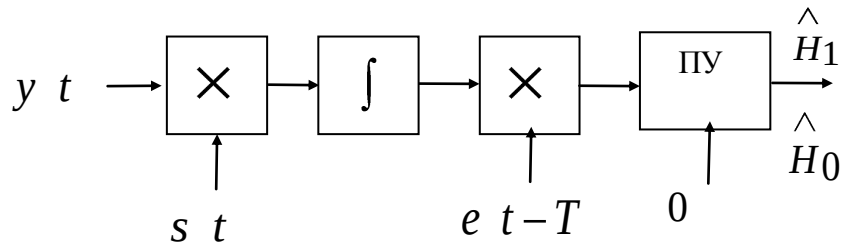


Рис. 2.2. Структурная схема различителя двух сигналов

Как всегда, коррелятор можно заменить фильтром, который должен быть согласованным с разностным сигналом $s(t)$.

Энергия E_d разностного сигнала на основании (2.12) будет определяться выражением

$$E_d = \int_0^T [s_0(t) - s_1(t)]^2 dt = d_{01}^2 = 2E(1 - \rho), \quad (2.16)$$

где $\rho = 1/E \int_0^T s_0(t)s_1(t)dt$ - устанавливает коэффициент корреляции сигналов.

Теперь выражение (2.13) можно представить в более традиционном виде

$$p_{01} = Q\left(\sqrt{\frac{2E}{N_0}\left(\frac{1-\rho}{2}\right)}\right) = Q\left(q\sqrt{\frac{1-\rho}{2}}\right), \quad (2.17)$$

где $q = \sqrt{\frac{2E}{N_0}}$ - параметр различителя, характеризующий отношение сигнал/шум на выходе фильтра.

В силу симметрии векторов s_0 и s_1 вероятности перепутывания p_{01} и p_{10} одинаковы и полная вероятность ошибки при равновероятных сигналах $p_0 = p_1 = 1/2$ будет равна

$$P_e = \frac{p_{01} + p_{10}}{2} = Q\left(q\sqrt{\frac{1-\rho}{2}}\right). \quad (2.18)$$

Выражение (2.18) отвечают на вопрос о наилучшей в смысле минимума P_e паре сигналов фиксированной энергии E .

Для этого произведем анализ вероятности полной ошибки от корреляции $\rho = \frac{1}{E} \int s_0(t)s_1(t)dt$ пары сигналов. Очевидно, что для максимизации расстояния между двумя векторами фиксированной длины следует выбирать их *противоположными*, как это показано на рис. 2.3, а). Тогда угол между векторами s_0 и s_1 составит величину $\alpha = \pi$, $\cos \alpha = \rho_{01} = -1$ и, значит, расстояние между двумя векторами равно $d(s_0, s_1) = 2\sqrt{E}$.

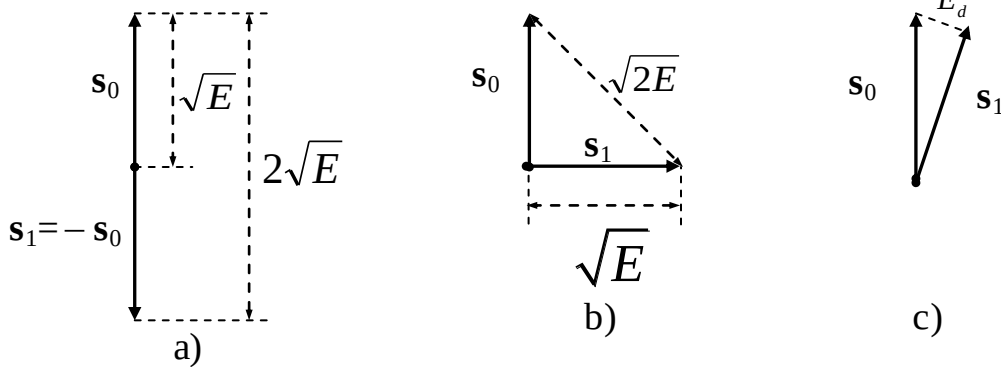


Рис. 2.3. Варианты выбора пар двоичных сигналов.

Для ортогональных сигналов $\cos \alpha = \rho_{01} = 0$ следовательно, расстояние между двумя векторами равно $d(s_0, s_1) = \sqrt{2E}$ (см. рис. 2.3, б).

Следует отметить, что с увеличением коэффициента корреляции $\rho_{01} \rightarrow 1$ энергия разностного сигнала $E_d \rightarrow 0$ (см. рис. 2.3, с), а вероятность полной ошибки $P_e \rightarrow 0.5$.

2.3 Прием когерентных двоичных сигналов

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

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

Временная диаграмма сигнала с бинарной фазовой манипуляцией данных изображена на рис.2.4.

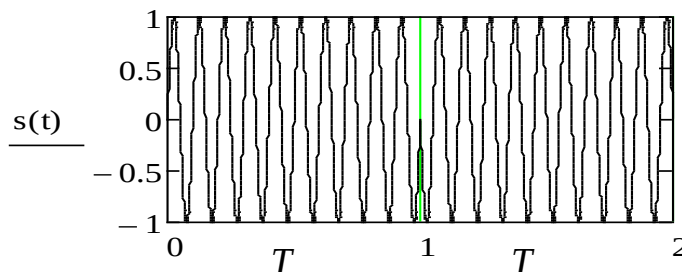


Рис. 2.4. Сигнал с бинарной фазовой манипуляцией данных.

Энергии сигналов $s_1(t)$ и $s_0(t)$ на интервале наблюдения одинаковые $E_1 = E_2 = E$. Поскольку угол между векторами s_0 и s_1 равен π , то коэффициент корреляции $\rho_{01} = -1$, что преобразует соотношение (2.18) в следующее

$$P_e = Q \left(\sqrt{\frac{2E}{N_0}} \right). \quad (2.19)$$

Вид условных плотностей вероятности при модуляции данных сигналами с бинарной фазовой манипуляцией представлен на рис. 2.6.

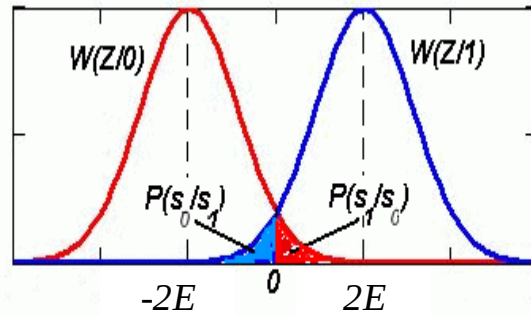


Рис.2.6. Вид условных плотностей вероятности при бинарной фазовой манипуляции данных

Хотя БФМ является наилучшим из возможных способов передачи двоичной информации, его применение основано на фазовом различии двух сигналов, носителей информации, и, следовательно, требует точного знания на приемной стороне текущего значения фазы несущей частоты. Практическая реализация этого достигается путем применения специального устройства восстановления несущей, что иногда рассматривается, как нежелательное усложнение приемной схемы. Избежать подобного усложнения удастся при использовании другого способа манипуляции - *частотной* ЧМ (FSK), при которой сообщения 0 и 1 передаются сигналами на различных частотах

$$s(t) = \begin{cases} s_1(t) = S_0 \sin \omega_1 t + \varphi_0, & \text{если передается 1} \\ s_0(t) = S_0 \sin \omega_2 t + \varphi_0, & \text{если передается 0} \end{cases}$$

Временная диаграмма сигнала с частотной манипуляцией данных изображена на рис. 2.6.

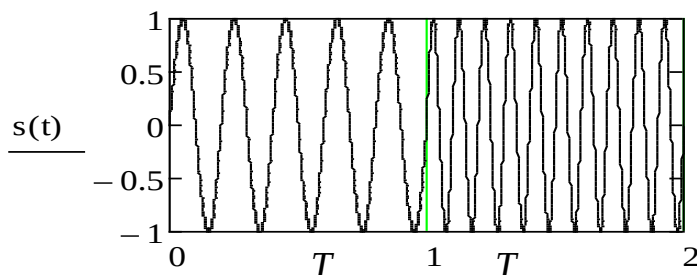


Рис. 2.6. Сигнал с частотной фазовой манипуляцией данных

Типичным является такой выбор несущих частот, при котором сигналы оказываются *ортогональными*, т.е. $\cos \alpha = \rho_{01} = 0$. Для ортогональных сигналов вероятность ошибки будет равна

$$P_e = Q\left(\frac{q}{\sqrt{2}}\right) = Q\left(\sqrt{\frac{E}{N_0}}\right). \quad (2.20)$$

Вид условных плотностей вероятности при модуляции данных сигналами с частотной манипуляцией представлен на рис. 2.7.

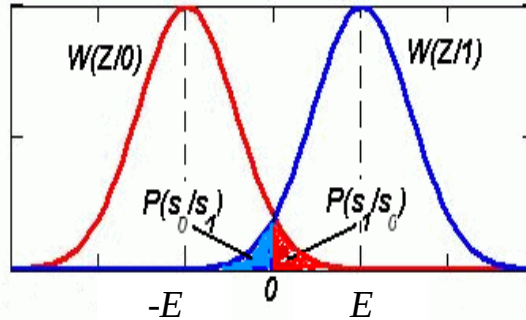


Рис.2.7. Вид условных плотностей вероятности при частотной манипуляции данных

Существует еще один, достаточно старый, способ передачи двоичной информации, до сих пор использующийся на практике: *амплитудная модуляция* АМ (ASK), при которой символ данных «1» передается сигналом $s_1(t) = s(t)$ с энергией $E_1 = E$, а символ «0» - паузой, т.е. $s_0(t) = 0, E_0 = 0$. Типовая диаграмма сигналов с пассивной паузой представлена на рис.2.8.

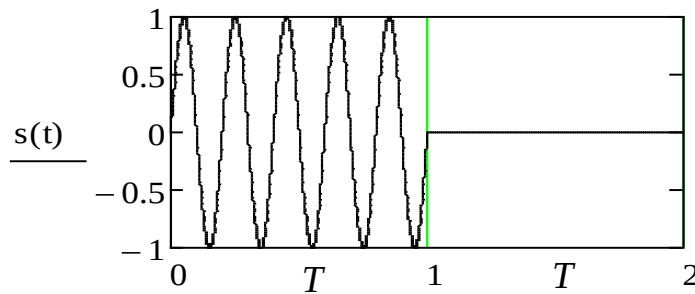


Рис.2.8. Сигнал с амплитудной манипуляцией данных

В этом случае $d(s_0, s_1) = \sqrt{E}$, а вероятность ошибки становится равной

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

Вид условных плотностей вероятности при амплитудной манипуляции данных представлен на рис. 2.9.

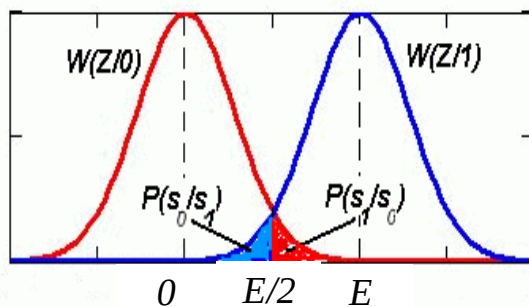


Рис. 2.9. Вид условных плотностей вероятности при амплитудной модуляции данных

На рис. 2.10 приведены графики вероятности полной ошибки для рассмотренных видов модуляции в зависимости от отношения сигнал/шум.

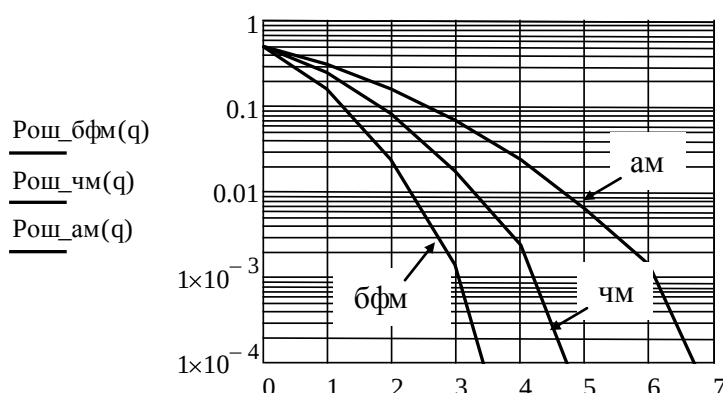


Рис. 2.10. Вероятности полной ошибки в зависимости от видов модуляции двоичных данных

Как видно, для обеспечения заданной вероятности полной ошибки требуется необходимое отношение сигнал шум в зависимости от вида модуляции двоичных данных. Наилучшей помехоустойчивостью обладают сигналы с бинарной фазовой манипуляцией. Сигналы с амплитудной манипуляцией характеризуются относительно низкой помехоустойчивостью. Сигналы с частотной манипуляцией занимают промежуточное значение. Так сравнивая результат (2.21) с (2.19), можно сделать вывод о том, что АМ требует в 4 раза (6 дБ) большей энергии, чем в случае БФМ, для обеспечения одинаковой достоверности приема. Последнее утверждение справедливо тогда, когда накладывается ограничение на пиковую энергию. На практике, как правило, ограничение накладывается на среднюю энергию. А поскольку при АМ энергия не излучается при передаче нулевых символов, то при равновероятных сообщениях «0» и «1» средняя энергия определится, как $(E_0 + E_1) / 2 = E / 2$. Тогда при условии одинаковой вероятности ошибки АМ проигрывает БФМ в средней энергии только в два раза. Значит, для АМ по отношению к БФМ характерны такие же энергетические потери, как и в случае применения ЧМ, т.е. 3 дБ.

3 РАЗЛИЧЕНИЕ МНОЖЕСТВА ДЕТЕРМИНИРОВАННЫХ СИГНАЛОВ

3.1 Структура многоканального различителя сигналов

При различении множества M детерминированных сигналов $S = s_i(t) : i = 1, 2, \dots, M$ на приемной стороне наблюдается принятое колебание $y(t)$, которое является результатом трансформации некоторого сигнала $s_i(t)$ искаженного канальным шумом. При произвольном числе сигналов M различитель, придерживающийся правила *максимального правдоподобия*, по-прежнему считает присутствующим в принятом колебании $y(t)$ сигнал $s_i(t)$, наименее удаленный от $y(t)$ в смысле евклидова расстояния

$$d_{ys} = \sqrt{\int [y(t) - s_i(t)]^2 dt}. \quad (3.1)$$

При одинаковых энергиях сигналов различитель считает присутствующим в принятом колебании сигнал $s_i(t)$, имеющий с $y(t)$ максимальную корреляцию

$$\hat{H}_k \geq z_i, \quad i = 0, 1, \dots, M-1 \quad (3.2)$$

Структура многоканального различителя сигналов изображена на рис. 3.1.

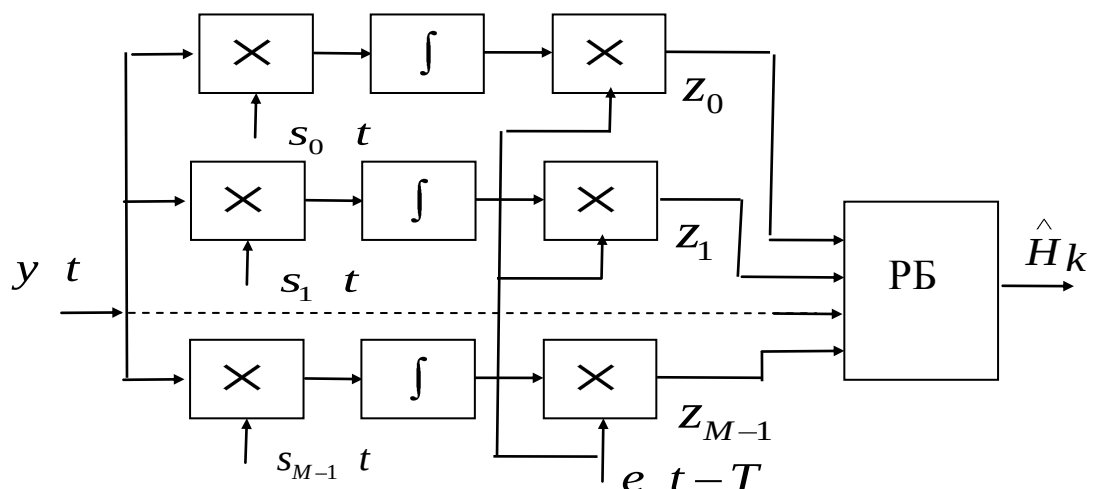


Рис. 3.1. Структура различителя M -детерминированных сигналов

Различитель содержит M каналов (по числу сигналов), вычисляющих корреляции наблюдаемой реализации $y(t)$ со всеми сигналами. Решающий блок (РБ) отыскивает наибольшую из корреляций. Сигнал, поданный в качестве опорного на коррелятор, на выходе которого зафиксирован максимум z_i , объявляется присутствующим в реализации $y(t)$. Вместо корреляторов можно использовать M фильтров, каждый из которых согласован со своим сигналом $s_i(t)$.

Оптимальный выбор M детерминированных сигналов сводится к поиску такой конфигурации пучка M векторов, в которой минимальное евклидово расстояние между парой векторов было бы максимальным $\min d_{ik} = \max$. Так при равенстве энергий сигналов евклидово расстояние определяется выражением

$$d_{ik} = \sqrt{\int [s_i(t) - s_k(t)]^2 dt} = \sqrt{2E(1 - \rho_{ik})}, \quad (3.3)$$

где ρ_{ik} - коэффициент корреляции i -го и k -го сигналов.

Таким образом, правило максимума минимального расстояния тождественно условию минимума максимального коэффициента корреляции во множестве (ансамбле) сигналов

$$\max_{i \neq k} \rho_{ik} = \min. \quad (3.4)$$

3.2 Анализ достоверности приема когерентных сигналов

Вероятность $p_{1,e}$ перепутывания действительно принятого $s_1(t)$ с одним из $M-1$ ошибочных сигналов $s_i(t)$, $i = 2, 3, \dots, M$, определяется, согласно решающим правилам, как

$$\begin{aligned} p_{1,e} &= P(d^2(s_1, y) \neq \min_i d^2(s_i, y) | s_1(t)) = \\ &= 1 - P\left(z_1 - \frac{E_1}{2} = \max_i (z_i - \frac{E_i}{2}) | s_1(t)\right) \end{aligned} \quad (3.5)$$

Весьма полезная и точная оценка сверху вероятности перепутывания $p_{1,e}$ может быть получена с использованием аддитивной границы. Так с учетом равной априорной вероятности всех M сигналов, аддитивная верхняя граница для полной (безусловной) вероятности ошибок примет вид

$$P_e = \frac{1}{M} \sum_{i=1}^M p_{i,e} \leq \frac{1}{M} \sum_{i=1}^M \sum_{\substack{j=1 \\ j \neq i}}^M Q \left(\sqrt{\frac{d^2(s_i, s_j)}{2N_0}} \right). \quad (3.6)$$

Первым заслуживающим внимания фактом, который непосредственно следует из (3.6), является то, что при $M=2$ данное неравенство превращается в равенство (2.14). Другое замечание связано с поведением выражения (3.6) в асимптотике при росте отношения сигнал-шум. При достаточно большом отношении сигнал-шум только близко расположенные сигнальные пары могут вносить заметный вклад в значение суммы в неравенстве (3.6). Пусть $d_{\min}$ есть минимальное расстояние на множестве всех возможных пар, встречающееся $n_{\min}$ раз, то при $d_{\min}^2/2N_0 \gg 1$ оценка (3.6) асимптотически трансформируется к виду

$$P_e \approx \frac{n_{\min}}{M} Q \left(\sqrt{\frac{d_{\min}^2}{2N_0}} \right). \quad (3.7)$$

Приближение (3.7) указывает, прежде всего, на асимптотическую сходимость аддитивной границы к истинной вероятности ошибки при увеличении отношения сигнал-шум.

3.3 Синтез ансамбля сигналов

Соотношение (3.7) лежит в основе одной из возможных и наиболее важных постановок задачи синтеза ансамбля сигналов, требующей *максимизации минимума расстояния во множестве M сигналов*. Подобная задача в геометрической интерпретации эквивалентна упаковке M векторов таким образом, чтобы ближайшая их пара находилась на максимально достижимом расстоянии, т.е. $d_{\min} = \max$. При этом на созвездие векторов (сигналов) могут быть наложены различные ограничения. Прежде всего, следует предусмотреть определенные

энергетические ограничения, учитывая доступные ресурсы мощности или энергии.

Если фиксировать только среднюю энергию сигналов $\bar{E} = \sum_{i=1}^M E_i / M = \text{const}$, то сигнальные вектора могут иметь различную длину и процедуру выбора сигналов можно назвать *объемной упаковкой*. Тем не менее, часто при решении задачи отображения сообщений в сигналы выдвигается требование исключения влияния энергетического параметра, т.е. требование одинаковой $E = E_i = \text{const}$, $i = 1, 2, \dots, M$ энергии всех сигналов. В этом случае все сигнальные вектора имеют одинаковую длину, и значит, лежат на сферической поверхности, откуда и следует название задачи - *сферическая упаковка*.

Другим типичным ограничением при синтезе сигналов является размерность n_s сигнального пространства, внутри которого и осуществляется их упаковка.

Рассмотрим первоначально видеосигналы и предположим, что полный (двусторонний) диапазон частот и временной интервал, отведенные для всех M сигналов, ограничены значениями W_t и T_t соответственно. Первое из этих ограничений учитывает экономию полосы, тогда как второе отражает желание передавать данные с приемлемой скоростью $R = \log M / T_t$. Тогда имеется около $W_t T_t$ независимых отсчетов, доступных для построения M сигналов, причем каждый из сигналов трактуется как вектор в пространстве размерности $n_s = W_t T_t$.

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

$$s(t) = S(t) \cos[2\pi f_0 t + \gamma(t)] = S_I(t) \cos 2\pi f_0 t - S_Q \sin 2\pi f_0 t, \quad (3.8)$$

где $S(t)$ - вещественная огибающая сигнала, отражающая закон модуляции амплитуды;

$S_I(t) = S(t) \cos \gamma(t)$ и $S_Q(t) = S(t) \sin \gamma(t)$ - квадратурные компоненты сигнала;

$\gamma(t)$ - описывает закон фазовой модуляции; f_0 - несущая частота.

При переносе на несущую частоту любой радиосигнал исчерпывающе описывается двумя независимыми низкочастотными квадратурными компонентами. Следовательно, при синтезе полосных сигналов могут использоваться в два раза больше независимых координат (отсчетов), чем в случае видеосигналов при одинаковых значениях частотно-временного произведения и, значит, $n_s = 2W_t T_t$.

Теперь в общем виде задача выбора множества сигналов может быть сформулирована следующим образом: *найти в пространстве заданной размерности n_s созвездие из M точек или векторов, удовлетворяющее энергетическим ограничениям и обладающее максимально возможным минимумом расстояния между точками $d_{\min} = \max$.*

Данная задача может быть переформулирована в дуальную: *найти в пространстве заданной размерности n_s созвездие из M точек или векторов с гарантированным минимальным расстоянием $d_{\min}$, которое обеспечивает минимизацию энергетических затрат.*

Простейшая версия этой задачи $n_s=1$ отвечает случаю амплитудной (АМ) модуляции. Альтернативным наименованием АМ служит *импульсно-кодовая модуляция* (ИКМ)). В этом случае все сигнальные точки располагаются на одной прямой линии. Не составляет труда убедиться, что при заданном $d_{\min}$ оптимальное созвездие, минимизирующее среднюю энергию, образуется в результате равномерного и симметричного расположения сигнальных точек с разнесением соседних на величину, в точности равную $d_{\min}$ (см. рис. 3.2, а).

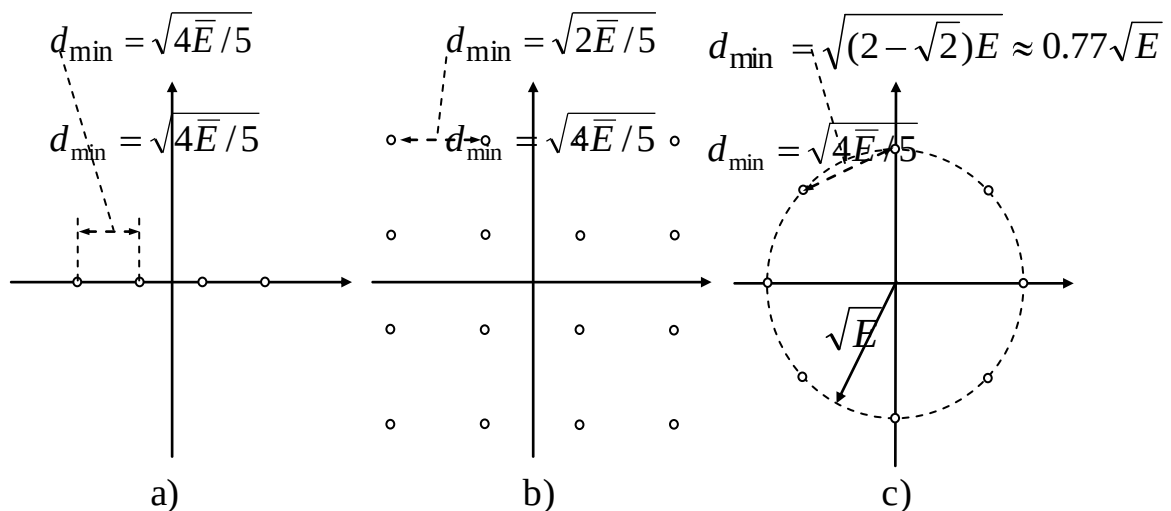


Рис. 3.2. Одно- и двумерные созвездия: 4-АМ (а), 16-КАМ (b), и 8-ФМ

Широко используемая в современной цифровой связи M -ичная *квадратурная амплитудная модуляция* (КАМ) служит примером двумерного, симметричного объемно-упакованного созвездия, которое, не являясь оптимальным теоретически, удобно с точки зрения аппаратной реализации (рис. 3.2, b). С другой стороны, традиционная M -ичная *фазовая манипуляция* (М-ФМ) отвечает созвездиям с равномерно расположенными на окружностях точками, оптимальным с точки зрения сферической упаковки (рис. 3.2, с).

Задача оптимального расположения точек в пространстве большей размерности ($n_s > 2$) является чрезвычайно сложной и не имеет до сих пор общего математического решения.

Определим верхний предел минимума расстояния в отсутствие предварительного ограничения на размерность сигнального пространства n_s . Ограничиваясь только случаем сферической упаковки ($E_i = E, i = 1, 2, \dots, M$), вычислим сумму всех M^2 возможных квадратов расстояний, включая и тривиальные (т.е. расстояния между любым сигналом с ним самим). Тогда на основании теоремы косинусов получаем

$$\sum_{i,j=1}^M d^2(\mathbf{s}_i, \mathbf{s}_j) = 2M^2E - 2E \sum_{i,j=1}^M \rho_{ij}, \quad (3.9)$$

где ρ_{ij} - коэффициент корреляции между i -м и j -м сигналами.

Для оценки суммы всех коэффициентов корреляции воспользуемся определением коэффициента корреляции

$$\rho_{ij} = \frac{1}{E} \int_0^T s_i(t) \cdot s_j(t) dt. \quad (3.10)$$

Поменяем порядок интегрирования и суммирования и отметим, что двойное суммирование под знаком интеграла осуществляется по отдельным индексам i и j , а значит, его можно заменить произведением двух идентичных сумм

$$E \sum_{i,j=1}^M \rho_{ij} = \int_0^T \left(\sum_{i,j=1}^M s_i(t) s_j(t) \right) dt = \int_0^T \left(\sum_{i=1}^M s_i(t) \right)^2 dt. \quad (3.11)$$

Поскольку интеграл от квадрата никогда не будет отрицательным, из (3.11) вытекает, что

$$\sum_{i,j=1}^M d^2(\mathbf{s}_i, \mathbf{s}_j) \leq 2M^2E. \quad (3.12)$$

В то же время приведенная выше сумма не меньше, чем $M(M-1)d_{\min}^2$.

Учитывая оба указанных неравенства, можно получить следующую верхнюю границу для минимума расстояния

$$d_{\min}^2 \leq \frac{2M}{M-1} E. \quad (3.13)$$

Сигналы, достигающие этой верхней границы, будем называть оптимальными по критерию минимума расстояния. Сигналы, удовлетворяющие границе (3.13), действительно существуют. Данный класс сигналов широко известен под специальным наименованием *симплексных сигналов*.

Симплексные сигналы. Свойства симплексных сигналов влекут за собой равенство коэффициентов корреляции ρ_{ij} любой их пары

$$\rho_{ij} = \frac{(\mathbf{v}_i, \mathbf{v}_j)}{\|\mathbf{v}_i\| \|\mathbf{v}_j\|} = -\frac{1}{M-1}, i \neq j, i, j = 1, 2, \dots, M, \quad (3.14)$$

демонстрирующему, что угол между любыми двумя симплексными сигналами один и тот же и больше, чем $\pi/2$. Диаграммы для простейших множеств из $M = 2, 3, 4$ симплексных сигналов, представлены на рис. 3.3.

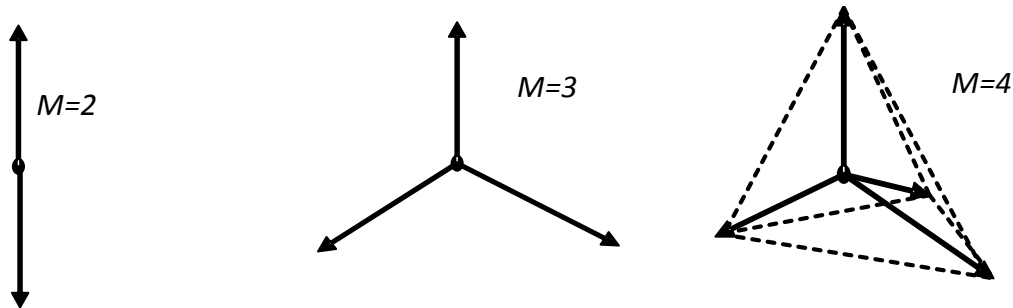


Рис. 3.3. Примеры симплексных сигналов.

Для приведенных множеств величина коэффициента корреляции принимает значения, равные -1 (противоположные сигналы), -1/2 и -1/3 соответственно, что в свою очередь отвечает углам в 180° , 120° и примерно 110° . При $M = 4$ симплексные вектора образуют простейший правильный многогранник (тетраэдр), что объясняет наименование сигналов: «симплекс» на латинском означает «простой».

При использовании симплексных сигналов асимптотическая оценка вероятности ошибки описывается неравенством

$$P_{e \min} \leq (M-1) Q \left(\sqrt{\frac{ME}{(M-1)N_0}} \right), \quad \frac{2E}{N_0} \gg 1. \quad (3.15)$$

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

Ортогональные сигналы практически столь же эффективны, как и симплексные, при достаточно большом значении M . Говоря об M -ичной ортогональной передаче вспомним, что максимальное число ортогональных сигналов в точности совпадает с размерностью сигнального пространства, т.е. $M = n_s$. Поэтому в пределах общей полосы

W_t и длительности T_t могут быть расположены $W_t T_t$ видео или $2W_t T_t$ полосных ортогональных сигналов.

Отметим, что реализация ансамбля M ортогональных сигналов как путем временных или частотных сдвигов между отдельными сигналами, так и путем кодирования их формы требует расширения общей полосы частот в M раз. Построение n_s ортогональных полосных сигналов вида

$s(t) = S(t) \cos[2\pi f_0 t + \gamma(t)]$ предполагает добавление к ним еще n_s сигналов, получаемых за счет сдвига фазы несущей частоты на угол $\pi/2$. Эта возможность осуществима на практике только в случае, когда все сигналы детерминированы или *когерентны*, что означает управление фазами несущей и возможность их использования для идентификации сообщений.

Приведем оценку вероятности ошибки при использовании ортогональных сигналов. Отметим, что коэффициент корреляции для ортогональных сигналов равен нулю, а расстояние между любой их парой $d(\mathbf{s}_i, \mathbf{s}_j) = d_{\min} = \sqrt{2E}$. Используя этот факт, получим асимптотическую оценку вероятности ошибки для M ортогональных сигналов, которая ограничивает сверху точное значение вероятности ошибки

$$P_{e \text{ ort}} \leq (M-1) Q \left(\sqrt{\frac{E}{N_0}} \right), \quad \frac{2E}{N_0} \gg 1. \quad (3.16)$$

Сравнение (3.16) с (3.15) демонстрирует, что для выравнивания в обоих случаях вероятностей ошибок необходимо увеличить энергию ортогональных сигналов в $M / (M-1)$ раз по сравнению с симплексными,

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

4 РАЗЛИЧЕНИЕ МНОЖЕСТВА НЕКОГЕРЕНТНЫХ СИГНАЛОВ

Как указывалось ранее, в реальной ситуации высока вероятность того, что передатчик, либо канал не смогут сохранить когерентность радиосигналов и, значит, на приемной стороне последние будут характеризоваться случайностью фазы. В этом случае значения начальных фаз не могут быть использованы для различения передаваемых сообщений, и отличительный признак сигналов должен строиться не на основе фазового сдвига. Ситуации, отвечающие подобному положению, известны под названием некогерентного приема.

4.1 Скалярное произведение некогерентных сигналов

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

$$s_i(t; \varphi_i) = \operatorname{Re}[\dot{S}_i(t; \varphi_i) \exp(j2\pi f_0 t)], \quad (4.1)$$

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

Введем теперь обобщенную версию скалярного произведения, которая могла бы использоваться не только в случае вещественных (реальных) сигналов $u(t), v(t)$, но и их комплексных заменителей - аналитических сигналов $\dot{u}(t), \dot{v}(t)$ или комплексных огибающих $\dot{U}(t), \dot{V}(t)$.

Модифицированное скалярное произведение может быть определено как

$$(\dot{\mathbf{u}}, \dot{\mathbf{v}}) = \int_0^T \dot{u}(t) \cdot \dot{v}^*(t) dt = \int_0^T \dot{U}(t) \cdot \dot{V}^*(t) dt = (\dot{\mathbf{U}}, \dot{\mathbf{V}}), \quad (4.2)$$

где комплексное сопряжение используется для сохранения равенства между скалярным произведением вектора с самим собой и квадратом длины вектора (всегда вещественного и не отрицательного).

Совпадение скалярного произведения аналитических сигналов и комплексных огибающих следует из определения аналитического сигнала

$$\dot{s}(t) = \dot{S}(t) \exp(j2\pi f_0 t) = s(t) + js_{\perp}(t), \quad (4.3)$$

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

В частности, для сигнала $s(t)$ с энергией E согласно соотношениям (4.2) и (4.3) получаем

$$(\dot{S}, \dot{S}) = \|\dot{S}\|^2 = \int_0^T |\dot{S}(t)|^2 dt = \int_0^T S^2(t) dt = \int_0^T s^2(t) dt + \int_0^T s_{\perp}^2(t) dt = 2E, \quad (4.4)$$

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

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

$$\begin{aligned} d^2(\dot{S}_i, \dot{S}_j) &= \|\dot{S}_i - \dot{S}_j\|^2 = (\dot{S}_i - \dot{S}_j, \dot{S}_i - \dot{S}_j) = \\ &= 2E_i + 2E_j - 4\text{Re}\left(\dot{\rho}_{ij} \sqrt{E_i E_j}\right), \end{aligned} \quad (4.5)$$

где

$$\dot{\rho}_{ij} = \frac{(\dot{S}_i, \dot{S}_j)}{\|\dot{S}_i\| \|\dot{S}_j\|} = \frac{1}{2\sqrt{E_i E_j}} \int_0^T \dot{S}_i(t) \dot{S}_j^*(t) dt \quad (4.6)$$

- коэффициент корреляции, но адаптированный к комплексным огибающим.

Так как корреляция комплексных огибающих равна

$$\int_0^T \dot{S}_i(t) \dot{S}_j^*(t) dt = \int_0^T (s_i(t) + js_{i\perp}(t))(s_j(t) - js_{j\perp}(t)) dt = 2(s_i, s_j) + 2j(s_{i\perp}, s_{j\perp}),$$

то $\text{Re}(\dot{\rho}_{ij}) = \rho_{ij}$, т.е. совпадает с обычным коэффициентом корреляции сигналов $s_i(t)$ и $s_j(t)$. Учитывая это в (4.5), получаем соотношение, связывающее расстояние между комплексными огибающими и собственно сигналами

$$d^2(\dot{\mathbf{S}}_i, \dot{\mathbf{S}}_j) = 2d^2(\mathbf{s}_i, \mathbf{s}_j). \quad (4.7)$$

Теперь вместо непосредственного вычисления расстояния между сигналами $s_i(t; \varphi_i)$ и $s_j(t; \varphi_j)$ можно воспользоваться соотношением (4.5) и оценить расстояние между комплексными огибающими $\dot{S}_i(t; \varphi_i)$ и $\dot{S}_j(t; \varphi_j)$. Предполагая равенство энергий E всех сигналов, получаем

$$d^2(\dot{\mathbf{S}}_{i\varphi}, \dot{\mathbf{S}}_{j\varphi}) = 4E [1 - \text{Re } \dot{\rho}_{ij}(\varphi)], \quad (4.8)$$

где дополнительный подстрочный индекс φ подчеркивает соответствие вектора $\dot{\mathbf{S}}_{i\varphi}$ полной комплексной огибающей $\dot{S}_i(t; \varphi_i)$.

Принимая в расчет независимость энергии сигнала E от значения начальной фазы φ_i , получаем, что

$$\dot{\rho}_{ij}(\varphi) = \frac{(\dot{\mathbf{S}}_{i\varphi}, \dot{\mathbf{S}}_{j\varphi})}{2E} = \frac{1}{2E} \int_0^T \dot{S}_i(t; \varphi_i) \dot{S}_j^*(t; \varphi_j) dt, \quad (4.9)$$

является коэффициентом корреляции полных комплексных огибающих $\dot{S}_i(t; \varphi_i)$ и $\dot{S}_j(t; \varphi_j)$. Учитывая неизменность значений фаз во времени, последнее соотношение может быть переписано в виде

$$\dot{\rho}_{ij}(\varphi) = \dot{\rho}_{ij} \exp(j\varphi_i - j\varphi_j).$$

Корреляция $\dot{\rho}_{ij}$ связана только с детерминированными комплексными огибающими $\dot{S}_i(t; \varphi_i)$ и $\dot{S}_j(t; \varphi_j)$ сигналов и представима в виде

$$\dot{\rho}_{ij} = \frac{1}{2E} \int_0^T \dot{S}_i(t) \dot{S}_j^*(t) dt = |\dot{\rho}_{ij}| \exp(j\varphi_{ij}), \quad (4.10)$$

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

Теперь квадрат расстояния, приведенный выше, будет записан, как

$$d^2(\dot{\mathbf{s}}_i, \dot{\mathbf{s}}_j) = 4E \left[1 - |\rho_{ij}| \cos(\varphi_{ij} + \varphi_i - \varphi_j) \right]. \quad (4.11)$$

Единственная трудность, возникающая при его вычислении, заключается в его зависимости от неизвестных фаз φ_i, φ_j . Вследствие этого, существует множество расстояний для заданных детерминированных законов модуляции $\dot{\mathbf{s}}_i(t)$ и $\dot{\mathbf{s}}_j(t)$, отвечающих случайной разности фаз $\varphi_i - \varphi_j$.

При значительном уровне модуля коэффициента корреляции $|\rho_{ij}|$ всегда присутствует вероятность, что при неудачной комбинации фаз (т.е. при малости $\varphi_{ij} + \varphi_i - \varphi_j$) расстояние (4.11) окажется малым. Для исключения подобной ситуации необходимо, чтобы модуль коэффициента корреляции был минимально возможным, а наилучшее множество сигналов должно удовлетворять условию

$$\dot{\rho}_{ij} = \frac{(\dot{\mathbf{s}}_i, \dot{\mathbf{s}}_j)}{2E} = 0, \quad i \neq j, \quad i, j = 1, 2, \dots, M. \quad (4.12)$$

Вновь приходим к необходимости использования ортогональных сигналов. Однако теперь условие ортогональности является более обязывающим, обуславливая ортогональность комплексных огибающих, или, другими словами, законов модуляции, сигналов, а не только самих сигналов. Вследствие этого полосные сигналы сохраняют ортогональность при любых сочетаниях фаз, поскольку $\dot{\rho}_{ij} = 0$ влечет и $\dot{\rho}_{ij}(\varphi) = 0$. С другой стороны, условие (4.12) исключает возможность обеспечения ортогональности за счет квадратурного сдвига фазы несущей, что было приемлемым в случае когерентных полосных сигналов. Прямым следствием этого является сокращение вдвое размерности сигнального пространства с величины $2W_t T_t$, как это было в случае видеосигналов, до значения $W_t T_t$ при заданной полосе W_t .

4.2 Различение М-сигналов с неизвестной фазой

Следует упомянуть, что в отличие от ситуации, которая имела место для

M детерминированных сигналов, при некогерентном приеме ортогональные сигналы являются строго оптимальными вне зависимости от их числа. Например, при некогерентном приеме оптимальной является пара ортогональных сигналов, тогда как пара противоположных сигналов не удовлетворяет данному утверждению, поскольку случайность фазы исключает возможность их различения.

Еще один момент, о котором следует упомянуть, касается выяснения оптимальной стратегии решения при некогерентном приеме. Наблюдение $y(t)$, рассматриваемое как полосный сигнал, может быть выражено в терминах комплексной огибающей или закона модуляции $\dot{Y}(t)$.

Очевидно, что $\dot{Y}(t)$ является случайным процессом. При решении, какой из M возможных сигналов был принят, необходимо сравнить расстояния между наблюдением и сигналами, что эквивалентно сравнению квадратов расстояний между соответствующими комплексными огибающими

$$d^2(\dot{S}_i, \dot{Y}) = \|\dot{Y}\|^2 + 2E - 2\operatorname{Re}[\dot{z}_i(\varphi_i)], \quad (4.13)$$

где $\dot{z}_i(\varphi_i)$ - значение корреляции комплексных огибающих наблюдения

$y(t)$ и i -го сигнала со случайной фазой $s_i\left(t; \varphi_i\right)$.

Для исключения зависимости от случайной фазы логично выбрать для сравнения только минимальное по всем возможным значениям φ_i расстояние для каждого i . Опуская детали, окончательное решающее правило выражается в терминах модуля корреляции

$$Z_i = \left| \int_0^T \dot{Y}(t) \dot{S}_i^*(t) dt \right|, i = 1, 2, \dots, M, \quad (4.14)$$

и означает, что j -й сигнал считается принятым, если модуль корреляции между его полностью известной комплексной огибающей и аналогичной характеристикой наблюдения принимает максимальное значение. Другими словами, используется старое и неоднократно используемое правило: выбирается такой сигнал, закон модуляции которого в наибольшей степени подобен закону модуляции наблюдения. Структура различителя, в основу которой положено правило (4.14), изображена на рис. 4.1, в которой Z_i находится как огибающая на выходе фильтра, согласованного

с i -м сигналом $s_i(t, \varphi)$ при некотором фиксированном значении φ .

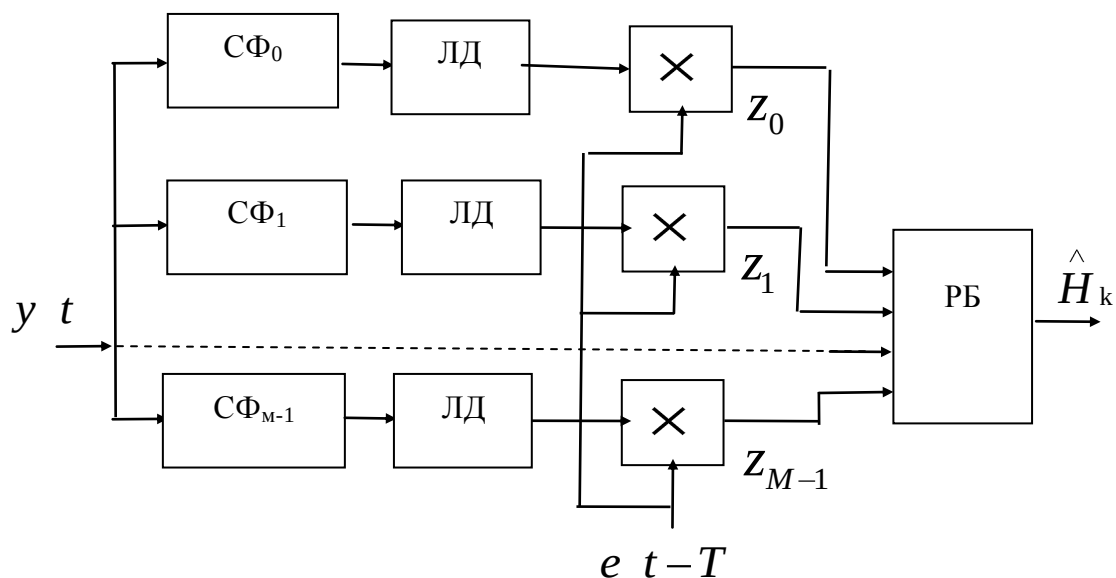


Рис. 4.1. Схема различителя М-сигналов с неизвестной фазой на согласованных фильтрах

Вместо линейных детекторов ЛД в этой схеме можно применить детекторы огибающей с любыми монотонными характеристиками, так как логика работы решающего блока РБ сводится к отбору максимальной из величин Z_i и вынесению решения о присутствии на входе того сигнала, с которым согласован фильтр, заметнее других прореагировавший на $y(t)$.

4.3 Прием некогерентных двоичных сигналов

4.3.1 Амплитудно-манипулированные сигналы

Структурная схема некогерентного приема двоичных данных с пассивной паузой (единица - импульс, ноль - пауза) приведена на рис. 4.2. Она содержит полосовой фильтр, детектор огибающей и пороговое устройство принятия решения.



Рис. 4.2. Схема приема некогерентных двоичных сигналов

Огибающая сигнала с БАМ представлена на рис. 4.3, а условные плотности вероятностей на выходе детектора огибающей приведены на рис. 4.4.

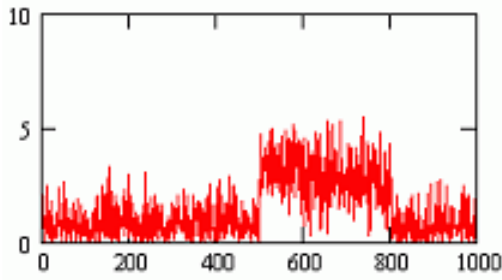


Рис. 4.3. Огибающая сигнала БАМ.

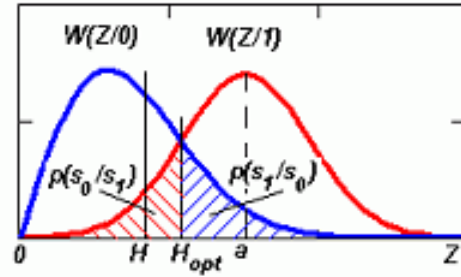


Рис. 4.4. Распределения огибающей БАМ.

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

$$W(U_{\text{сш}}) = \frac{U_{\text{сш}}}{\sigma^2} \exp\left(-\frac{U_{\text{сш}}^2 + S_0^2}{2\sigma^2}\right) I_0\left(\frac{S_0 U_{\text{сш}}}{\sigma^2}\right), \quad (4.15)$$

где S_0 - амплитуда полезного сигнала, I_0 - модифицированная функция Бесселя нулевого порядка.

При отсутствии сигнала получаем закон Релея

$$W(U_{\text{ш}}) = \frac{U_{\text{ш}}}{\sigma^2} \exp -U_{\text{ш}}^2 / 2\sigma^2. \quad (4.16)$$

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

$$W(Z/1) = Z \exp - Z^2 + a^2 / 2 I_0(aZ); \quad (4.17)$$

$$W(Z/0) = Z \exp -Z^2/2. \quad (4.18)$$

Теперь, даже при равенстве априорных вероятностей $P_{s_0} = P_{s_1}$, канал несимметричен и вероятности перепутывания символов определяются

$$P_{s_0/s_1} = \int_0^H W(Z/1) dZ; \quad (4.19)$$

$$P_{s_1/s_0} = \int_H^{\infty} W(Z/0) dZ,$$

где $H = U_{\text{пор}}/\sigma$ - нормированный порог.

Как видно из рис. 4.4, оптимальное значение порога, минимизирующее вероятность полной ошибки P_e , находится как абсцисса точки пересечения кривых $W(Z|0)$ и $W(Z|1)$.

Решая уравнение

$$W(Z|0)|_{Z=H_{\text{opt}}} = W(Z|1)|_{Z=H_{\text{opt}}} \quad (4.20)$$

относительно Z , получим

$$I_0(aH_{\text{opt}}) = \exp(a^2/2). \quad (4.21)$$

Решение этого трансцендентного уравнения представлено графиком на рис. 4.5, где $h = \sqrt{E/N_0}$ - отношения сигнал/шум.

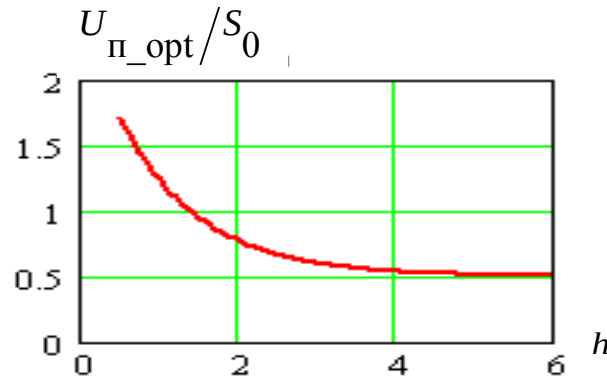


Рис. 4.7. Зависимость оптимального порога от отношения сигнал/шум

Заметим, что

$$H_{\text{opt}} = \frac{U_{\text{п_опт}}}{\sigma} = \frac{U_{\text{п_опт}}}{S_0} \cdot \frac{S_0}{\sigma} = \frac{U_{\text{п_опт}}}{S_0} \sqrt{2}h. \quad (4.22)$$

При достаточно большом отношении сигнал/шум $h \geq 3$ величина оптимального порога $U_{\text{п_опт}}/S_0$ стремится к 0,5, т.е. $H_{\text{opt}} \approx a/2 = h/\sqrt{2}$

.При $h \geq 3$ закон Релея-Райса достаточно хорошо аппроксимируется нормальным законом

$$W(Z/1) \approx \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(Z-a)^2}{2}\right], \quad (4.23)$$

Используем эту аппроксимацию для вычисления условной вероятности перепутывания «1» с «0» (пропуск сигнала)

$$\begin{aligned} P_{s_0/s_1} &= 1 - \frac{1}{\sqrt{2\pi}} \int_{H_{\text{opt}}}^{\infty} \exp\left[-\frac{(Z-a)^2}{2}\right] dZ = 1 - \frac{1}{\sqrt{2\pi}} \int_{H_{\text{opt}}-a}^{\infty} \exp(-y^2/2) dy = \\ &= 1 - \frac{1}{\sqrt{2\pi}} \int_{-H_{\text{opt}}}^{\infty} \exp(-y^2/2) dy = 0,5 - \Phi(H_{\text{opt}}) = Q(H_{\text{opt}}). \end{aligned} \quad (4.24)$$

Вероятность перепутывания сигналов P_{s_1/s_0} определится, как

$$\begin{aligned} P_{s_1/s_0} &= \int_{H_{\text{opt}}}^{\infty} z \exp\left(\frac{-z^2}{2}\right) dz = \\ &= \int_{H_{\text{opt}}}^{\infty} \exp\left(\frac{-z^2}{2}\right) d\frac{z^2}{2} = \exp\left(-H_{\text{opt}}^2/2\right). \end{aligned} \quad (4.25)$$

Полная вероятность ошибки

$$P_e = 0,5 P_{s_0/s_1} + P_{s_1/s_0} = 0,5 \left\{ \exp\left(-H_{\text{opt}}^2/2\right) + Q(H_{\text{opt}}) \right\} \quad (4.26)$$

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

$$P_e = 0,5 \exp(-h^2/4) + Q(h/\sqrt{2}). \quad (4.27)$$

4.3.2 Частотно-манипулированные сигналы

Рассмотрим частотно-манипулированные сигналы в качестве ортогональных. Аналитическая запись такого сигнала имеет вид

$$s(t) = \begin{cases} s_1(t) = S_0 \cos(\omega_1 t + \varphi_1) \\ s_0(t) = S_0 \cos(\omega_2 t + \varphi_2) \end{cases} \quad (4.28)$$

Некогерентный линейный приемник для этих сигналов можно строить с двумя полосовыми фильтрами, как показано на рис.4.8.

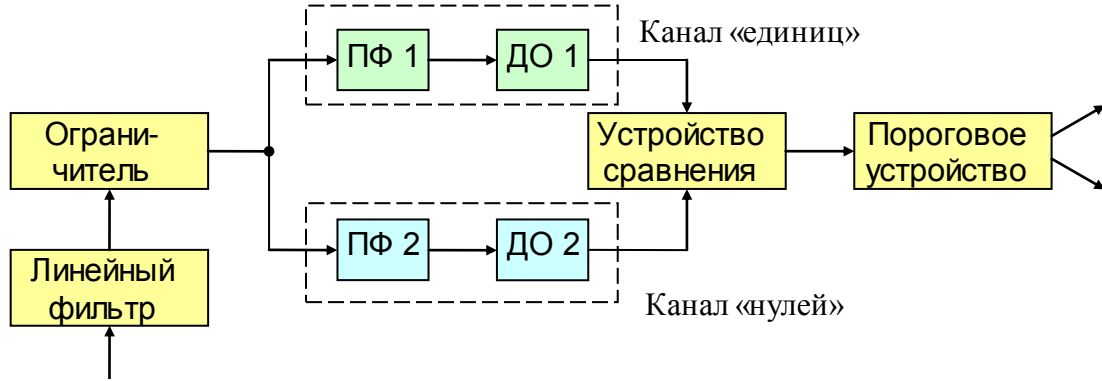


Рис. 4.8. Приемник ЧМ-сигналов с двумя полосовыми фильтрами

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

$$P_{U_{\text{ш}} \geq U_{\text{сш}}} = \int_{U_{\text{сш}}}^{\infty} W(U_{\text{ш}}) dU_{\text{ш}}. \quad (4.29)$$

$U_{\text{сш}}$ - случайная величина, поэтому необходимо произвести усреднение по всем возможным значениям

$$P_{s_0/s_1} = \int_0^{\infty} W(U_{\text{сш}}) P(U_{\text{ш}} \geq U_{\text{сш}}) dU_{\text{сш}} \quad (4.30)$$

или

$$P_{s_0/s_1} = \int_0^{\infty} \int_0^{\infty} W(U_{\text{сш}}) W(U_{\text{ш}}) dU_{\text{ш}} dU_{\text{сш}}, \quad (4.31)$$

где $W_{U_{\text{сш}}}$ - распределение Релея-Райса, $W_{U_{\text{ш}}}$ - распределение Релея. Используя обозначения $v = U_{\text{сш}}/\sigma$, $z = U_{\text{ш}}/\sigma$, $s = S_0/\sigma$, получим

$$\begin{aligned}
P_{s_0/s_1} &= \int_0^\infty \int_0^\infty \frac{U_{\text{сш}}}{\sigma^2} \exp\left(-\frac{U_{\text{сш}}^2 + S_0^2}{2\sigma^2}\right) \times \\
&\times I_0\left(\frac{U_{\text{сш}} S_0}{\sigma^2}\right) \frac{U_{\text{ш}}}{\sigma^2} \exp\left(-\frac{U_{\text{ш}}^2}{2\sigma^2}\right) dU_{\text{ш}} dU_{\text{сш}} = \\
&= \int_0^\infty \int_0^\infty v \exp\left(-\frac{v^2 + a^2}{2}\right) I_0(av) z \exp\left(-\frac{z^2}{2}\right) dv dz. \quad (4.32)
\end{aligned}$$

Двойной интеграл легко разделяется, причем

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

и окончательно получим

$$P_{s_0/s_1} = 0.5 \exp\left(-\frac{a^2}{4}\right) = 0.5 \exp\left(-\frac{h^2}{2}\right). \quad (4.33)$$

Итак, вероятность полной ошибки при некогерентном приеме сигналов БЧМ

$$P_e = 0.5 \exp\left(-\frac{h^2}{2}\right). \quad (4.34)$$

4.4 Множества ортогональных сигналов

В данном разделе будет уделено основное внимание зависимости между M и эквивалентным частотно-временным ресурсом $W_t T_t$ системы. Продемонстрируем возможность построения простейших множеств ортогональных сигналов за счет дробления доступного ресурса.

4.4.1 Кодирование путем временного сдвига

Очевидно, что скалярное произведение любых двух неперекрывающихся во времени сигналов равно нулю. Пусть общий частотно-временной ресурс

представлен в виде прямоугольника на координатной плоскости время-частота со сторонами, равными соответственно T_t и W_t . Тогда кодирование путем временного сдвига означает лишь разрезание этого ресурса на M вертикальных полосок, каждая из которых приписывается некоторому конкретному сигналу (см. рис. 4.9). Ортогональность при передаче обеспечивается строгим распределением временного ресурса между сигналами, каждый из которых занимает весь отведенный частотный ресурс.

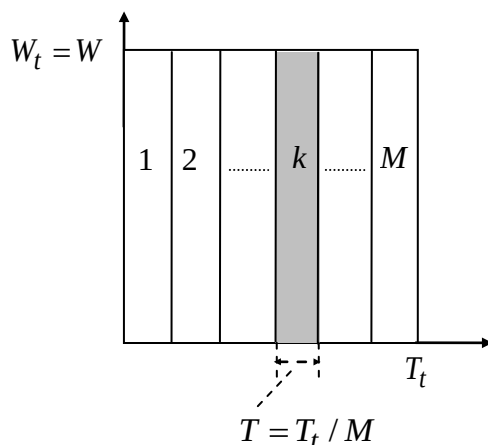


Рис. 4.9. Распределение ресурса при временном ортогональном кодировании.

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

4.4.2 Кодирование путем частотного сдвига

Другим непосредственным способом обеспечения ортогональности служит кодирование путем частотного сдвига. На основании дуальности

времени и частоты или теоремы Парсеваля скалярное произведение сигналов $u(t), v(t)$ и их спектров $\tilde{u}(f), \tilde{v}(f)$ совпадает

$$(\mathbf{u}, \mathbf{v}) = \int_{-\infty}^{\infty} u(t) v(t) dt = \int_{-\infty}^{\infty} \tilde{u}(f) \tilde{v}^*(f) df = (\tilde{\mathbf{u}}, \tilde{\mathbf{v}}), \quad (4.35)$$

что позволяет перевести ранее представленный метод в частотную область (см. рис. 4.10).

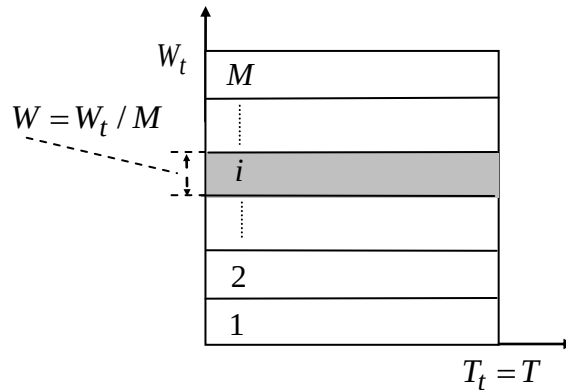


Рис. 4.10. Распределение ресурса при частотном ортогональном кодировании.

При полном перекрытии сигналов во времени ($T = T_t$) каждый из них занимает полосу не менее чем $W = 1/T_t$. Тогда максимальное число ортогональных сигналов, образованных сдвигом спектра, снова будет $M = W_t / W = W_t T_t = n_s$. Как и в предыдущем случае, общий ресурс снова нарезается на полосы, однако иным образом. Полоски располагаются горизонтально, означая, что любой сигнал использует весь временной ресурс T_t и только M -ю часть общего частотного ресурса W_t (рис. 4.10). Очевидно, что каждый индивидуальный сигнал снова не является сигналом с распределенным спектром, поскольку отвечающее ему частотно-временное произведение $WT = (W_t / M)T = 1$.

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

4.4.3 Ортогональное кодирование с распределением спектра

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

При таких обстоятельствах метод построения ортогональных сигналов путем распределения спектра может оказаться весьма конкурентоспособным, поскольку предполагает совместное использование всеми сигналами общего частотно-временного ресурса без распределения или нарезания последнего.

Рассмотрим простой пример воплощения указанной идеи для дискретных бинарных фазоманипулированных сигналов. Образует любой из M сигналов в виде последовательности N примыкающих друг другу элементарных импульсов (или чипов), каждый из которых обладает прямоугольной формой и длительностью Δ . Пусть полярность чипов сигнала под номером k изменяется в соответствие с *кодовой последовательностью* (или просто *кодом*) бинарных символов $a_{k,i} = \pm 1$, где $k = 1, 2, \dots, M$, а второй подстрочный символ отвечает номеру чипа (дискретному времени) $i = 0, 1, \dots, N-1$. Тогда версия подобного сигнала на видеочастоте может быть записана в виде

$$s_k(t) = \sum_{i=0}^{N-1} a_{k,i} s_0(t - i\Delta), \quad (4.36)$$

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

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

$$(s_k, s_l) = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} a_{k,i} a_{l,j} \int_0^T s_0(t - i\Delta) s_0(t - j\Delta) dt. \quad (4.37)$$

В последнем соотношении интеграл представляет собой скалярное произведение двух сдвинутых во времени друг относительно друга чипов

на величину, равную $(i - j)\Delta$. При $i \neq j$ интеграл равен нулю, поскольку импульсы не перекрываются во времени. Таким образом

$$\int_0^T s_0(t - i\Delta) s_0(t - j\Delta) dt = E_0 \delta_{ij}, \quad (4.38)$$

где E_0 - энергия элементарного импульса. Использование полученного выражения в (4.37) дает

$$(\mathbf{s}_k, \mathbf{s}_l) = E_0 \sum_{i=0}^{N-1} a_{k,i} a_{l,i} = E_0 (\mathbf{a}_k, \mathbf{a}_l). \quad (4.39)$$

Последнее соотношение связывает скалярное произведение сигналов со скалярным произведением N -мерных векторов, отвечающих кодовым последовательностям $\mathbf{a}_k = (a_{k,0}, a_{k,1}, \dots, a_{k,N-1})$. Очевидно, что M ортогональных кодовых последовательностей автоматически формируют M ортогональных сигналов типа (4.36). При $M \leq N$ существует множество способов конструирования подобных последовательностей, поскольку рассматриваемая ситуация отвечает задаче нахождения $M \leq N$ ортогональных векторов размерности N . В нашем примере упомянутые вектора являются бинарными, т.е. их компоненты принимают значения из множества ± 1 . При $M = N$ ортогональные бинарные вектора, рассматриваемые в качестве строк, образуют квадратную матрицу, называемую матрицей Адамара. Не составляет труда убедиться, что существуют лишь матрицы Адамара размерности, кратной четырем, т.е. $M \equiv 0 \pmod{4}$, где используемое сравнение $a \equiv b \pmod{c}$ означает равенство остатков от деления целых чисел a, b на целое c . До сих пор не найдено доказательство достаточности этого необходимого условия.

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

$$\mathbf{H}_{2M} = \begin{bmatrix} \mathbf{H}_M & \mathbf{H}_M \\ \mathbf{H}_M & -\mathbf{H}_M \end{bmatrix} = \begin{bmatrix} \mathbf{1} & \mathbf{1} \\ \mathbf{1} & -\mathbf{1} \end{bmatrix} \otimes \mathbf{H}_M, \quad (4.40)$$

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

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

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

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

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

Рис. 4.11 иллюстрирует тот факт, что в рамках данного метода построения ортогональных сигналов, не происходит деления общего ресурса: все сигналы располагаются в нем, полностью перекрываясь как во временной, так и в частотной области.

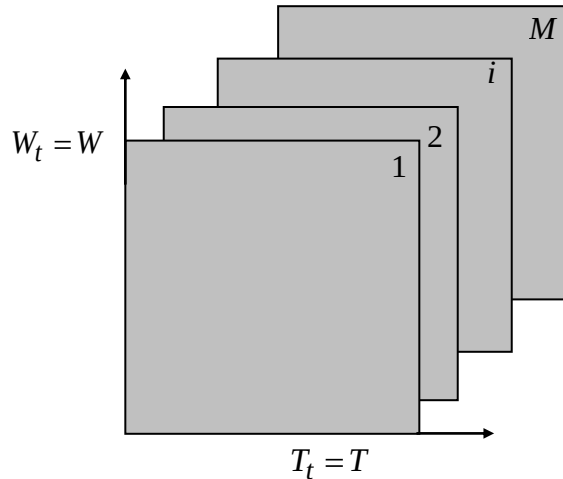


Рис. 4.11. Распределение ресурса при ортогонализации.

Действительно, полоса, занимаемая каждым сигналом, может быть оценена как $W = 1/\Delta$, тогда как длительность составляет величину $T = M\Delta$, откуда $WT = M = W_t T_t$. Ортогональность же в данном методе достигается не путем деления временного интервала или полосы, а за счет соответствующего выбора закона модуляции сигнала.

Анализируя достоинства ортогональности за счет распределения спектра, можно отметить, что методы формирования и обработки сигналов хорошо сопрягаются с современными средствами цифровой микроэлектроники. Еще одной особенностью служит автоматическое приобретение тех преимуществ технологии распределенного спектра, которые незаметны в рамках классической структуры приема, но многочисленны и очень ценны на практике.

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

4.5 Обмен между выигрышем от ортогонального кодирования и шириной полосы

Предыдущее рассмотрение очевидным образом высветило особую роль, которую играют ортогональные сигналы. Они практически оптимальны в случае M –ичной передачи данных как для когерентных сигналов, так и для некогерентного случая. Оценим теперь выигрыш, который сопровождает применение ортогональных сигналов со случаем отсутствия кодирования, т.е. непосредственной передачей потока информационных бит источника.

Сначала предположим, что энергетический ресурс позволяет передавать каждый бит данных с энергией E_b и каждый информационный бит передается оптимальным образом с помощью пары противоположных сигналов, другими словами используется БФМ. Среди блоков из m последовательных бит возможны любые комбинации, в том числе и такие, которые отличаются друг от друга значением только одного бита. Поэтому минимальная величина квадрата расстояния между сигналами, отвечающими неидентичным m –битовым блокам, совпадает со значением квадрата расстояния между однобитовыми противоположными сигналами

$$d_{\min}^2 = 4E_b. \quad (4.41)$$

Рассмотрим теперь другую систему, в которой все различные m –битовые блоки передаются ортогональными сигналами. Каждый такой сигнал, сохраняя неизменным ее значение на бит E_b , характеризуется энергией, равной mE_b . Тогда квадрат расстояния (одинаковый для любой пары сигналов, поскольку ортогональные сигналы эквидистантны) между сигналами снова легко может быть определен

$$d_{\min, \text{ort}}^2 = 2mE_b. \quad (4.42)$$

Выигрыш от ортогонального кодирования G_a определяется сравнением квадратов минимумов расстояния при кодировании и без него

$$G_a = \frac{d_{\min, \text{ort}}^2}{d_{\min, u}^2} = \frac{m}{2}. \quad (4.43)$$

Выигрыш от ортогонального кодирования приобретает за счет расширения полосы, поскольку размерность сигнального пространства n_s , т.е. число ортогональных сигналов $M = n_s$, непосредственно определяется

общим частотно-временным ресурсом $W_t T_t$ системы.

Пусть необходимая скорость передачи в системе составляет R бит/с, что соответствует передаче $m = RT_t$ бит за выделенный интервал времени T_t . Ясно, что ортогональное кодирование битовых блоков такой длины может быть осуществлено с помощью $M = 2^m = 2^{RT_t}$ сигналов, обеспечивая асимптотический выигрыш от кодирования

$$G_a = m / 2 = RT_t / 2. \quad (4.44)$$

Тогда спектральная эффективность R / W_t системы, т.е. скорость, приходящаяся на 1 Гц полосы, использующей ортогональные сигналы, будет равна

$$\frac{R}{W_t} = \frac{RT_t}{2^{RT_t}} = \frac{2G_a}{2^{2G_a}}, \quad (4.45)$$

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

При альтернативном подходе для обеспечения одинакового минимума расстояния передача данных без кодирования требует энергетических затрат в $G_a = m / 2$ больше, чем в случае ортогонального кодирования.

Принимая во внимание соотношение (2.29), минимум расстояния асимптотически определяет вероятность ошибки для любых сигналов, используемых в M -ичной передаче. Следовательно, требование высокой достоверности приема, автоматически влекущее за собой необходимость высокого значения отношения сигнал-шум, означает, что одинаковая надежность двух рассмотренных систем возможна только в случае G_a раз больших энергетических затрат для варианта без кодирования. Таким образом, *асимптотический* выигрыш G_a от кодирования ортогональными сигналами является адекватным показателем преимущества ортогональной сигнализации в пределе, т.е. при отношении сигнал-шум, стремящемся к бесконечности.

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

5 ОСНОВЫ ТЕОРИИ ПОМЕХОУСТОЙЧИВОГО КОДИРОВАНИЯ

5.1 Введение в помехоустойчивое кодирование

Повышение требований к скорости и достоверности передачи информации, увеличение протяженности линий связи приводит к необходимости принятия специальных мер, направленных на уменьшение вероятности возникновения ошибок в процессе передачи. Одним из возможных решений указанной задачи служит помехоустойчивое кодирование. Под *помехоустойчивыми* понимаются коды, позволяющие обнаруживать и исправлять ошибки, возникающие при передаче из-за воздействия помех.

Суть данной процедуры состоит во введении специальным образом дополнительных символов в информационный поток, в результате чего каждому блоку из k информационных бит сопоставляется n символьная последовательность $\mathbf{X}_i = (x^{(1)}, x^{(2)}, \dots, x^{(n)})$, $i = 1, 2, \dots, M$, где $M = 2^k$ – число возможных сообщений. Поскольку $n > k$, то не все последовательности длины n используются при кодировании M сообщений. Комбинации символов, используемые для отображения информационных блоков или сообщений, называют *разрешенными* комбинациями или *кодowymi* последовательностями (словами), тогда как остальные – *запрещенными*.

Вся совокупность кодовых слов $\mathbf{X}_i$, $i = 1, \dots, M$ образует *код*, для обозначения которого обычно говорят «код объема M длины n ». Множество символов, из которых составляются кодовые слова, называется *алфавитом кода*, а число различных символов в алфавите – *основанием кода*, или объемом (мощностью) алфавита.

Именно введение дополнительных символов и позволяет осуществить нейтрализацию влияния канальных помех. Появление подобной возможности можно пояснить простейшим примером.

Предположим, что по двоичному симметричному каналу (ДСК) передается $M = 4$ сообщений. В простейшем варианте для передачи этого количества сообщений достаточно двух бит, т.е. сопоставить сообщениям x_1, x_2, x_3 и x_4 двоичные кодовые слова вида 00, 10, 01 и 11 соответственно. Хотя подобное сопоставление может рассматриваться как примитивное кодирование (каждое сообщение отображается в двоичный вектор размерности $k = 2$), оно не предполагает никаких мер борьбы с канальными искажениями. Данный вариант отображения может быть назван безызбыточным кодированием. Достаточно исказиться одному символу k – мерного вектора (например, входная последовательность 10 трансформируется в 11), как на приемной стороне будет принято

неправильное решение о переданном сообщении (вместо x_2 будет принято решение в пользу x_4). Если же осуществить специальное кодирование, предусматривающее введение избыточных символов, появляется возможность защитить передаваемую информацию от искажения в канале. Например, сопоставим упомянутым выше 4-м сообщениям n – мерные двоичные вектора, состоящие из 5-ти символов: 00000, 10110, 01101 и 11011. В этом случае искажение одного символа последовательности не приведет к неправильному решению о переданном сообщении, если принятие решения основывается на «близости» (минимальном отличии в числе символов) принятой последовательности от кодового слова. Так, при векторе на выходе канала вида 10111, отличающемся от кодовых слов в 4-х, 1-й, 3-х и 2-х позициях соответственно, принимается решение о передаче сообщения x_2 . Не трудно убедиться, что данный вариант кодирования позволяет исправлять любую однократную ошибку в передаваемой последовательности. Появление указанной способности объясняется введение добавочных (*проверочных*) символов в кодовом слове, т.е. за счет введения избыточности.

В зависимости от позиций, с которых рассматривается процедура кодирования, классификация кодов может осуществляться различным образом. Простейшим вариантом может служить классификация *по размеру алфавита кода*. Если символы кода $x_i \in 0,1$ или $x_i \in 1,-1$, то код называется *двоичным* или *бинарным* соответственно. Если же алфавит кода содержит q символов, соответствующий ему код носит наименование *q-ичного*. В дальнейшем, вследствие простоты и наибольшего распространения, основное внимание будет сконцентрировано на двоичных или бинарных кодах.

Иной вариант классификации кодов может основываться *на способе вложения информации* в кодовое слово. По этому признаку различают *блоковые* (или *блочные*) и *древовидные* (или *решетчатые*) коды. При построении блокового кода последовательность непересекающихся блоков k информационных символов преобразуется в последовательность n -символьных кодовых слов, причем каждое кодовое слово отвечает только своему блоку данных. В отличие от блоковых древовидные коды строятся таким образом, что текущее кодовое слово определяется, в принципе, бесконечным набором предшествующих информационных блоков. Учитывая ограничения технологического характера при реализации кодера, кодовое слово зависит от конечного числа блоков данных, которое получило название *кодového ограничения*. Очевидно, что блоковые коды являются частным случаем древовидных с кодовым ограничением, равным единице. Наименование решетчатые коды, как аналог древовидных, связано со способом описания кода с помощью графа или решетки.

Классификация кодов может быть осуществлена и *по возможности выделения информационных символов* в кодовом слове. Коды, в которых, как правило, первые k позиций занимают информационные символы, называются *систематическими* кодами. В противном случае - *несистематическими*.

Коды можно классифицировать и по способу противодействия искажениям в канале распространения. Коды, позволяющие исправлять ошибки, получили наименование *исправляющих ошибки*, тогда как коды только их фиксирующие, называются кодами, *обнаруживающими ошибки*. Не редко коды, обнаруживающие и исправляющие ошибки, называют *контролирующими ошибки*.

5.1.1. Евклидово и Хэммингово расстояние.

Геометрическое толкование процедуры различения M сигналов $s_i(t)$ конечной энергии E_i предполагает их представление в виде векторов $\mathbf{S}_i$, длина которых может быть выражена соотношением

$$\|\mathbf{S}_i\| = \sqrt{\int_{-\infty}^{\infty} s_i^2(t) dt}.$$

Тогда *Евклидово расстояние* между сигналами $s_1(t)$ и $s_2(t)$ определяется как

$$d_E \mathbf{S}_1, \mathbf{S}_2 = \|\mathbf{S}_1 - \mathbf{S}_2\| = \sqrt{\int_{-\infty}^{\infty} [s_1(t) - s_2(t)]^2 dt}. \quad (5.1)$$

Данная характеристика показывает насколько удалены сигналы друг от друга, что играет определяющую роль в теории информации в целом. Чем больше расстояние между сигналами, тем меньше вероятность перепутывания переносимой ими информации.

Альтернативным вариантом определения расстояния, широко используемым при рассмотрении дискретного канала, служит *расстояние Хэмминга* $d_H \mathbf{X}_1, \mathbf{X}_2$, определяемое как число позиций, в которых кодовые символы двух слов отличаются друг от друга.

Для двух двоичных векторов $\mathbf{X}_1 = (10110)$ и $\mathbf{X}_2 = (11011)$ расстояние Хэмминга $d_H(\mathbf{X}_1, \mathbf{X}_2) = 3$, поскольку символы, стоящие на второй, третьей и пятой позиций различаются, а на первой и четвертой - совпадают. В свою очередь вес Хэмминга для указанных векторов составляет величину $w(\mathbf{X}_1) = 3$ и $w(\mathbf{X}_2) = 4$.

5.1.2. Связь кодового расстояния с корректирующей способностью кода.

Говорят, что код исправляет все α кратные ошибки, если декодирование по правилу минимума расстояния любого кодового слова с α (и менее) ошибочными символами завершается правильным решением. Параметром, определяющим исправляющую способность кода, служит его *кодвое расстояние*.

Кодовым расстоянием кода X называется минимальное расстояние Хэмминга между любой парой несовпадающих векторов кода

$$d = \min_{\substack{i, j \\ i \neq j}} d_H(X_i, X_j), \quad i, j = \overline{1, M}. \quad (5.2)$$

Код исправляет любые ошибки кратности α и менее в том и только в том случае, если кодовое расстояние удовлетворяет неравенству

$$d \geq 2\alpha + 1. \quad (5.3)$$

Из рассмотренного видно, что основными параметрами блочного кода являются: кодовое расстояние d , его объем M и длина n . Часто при описании характеристик кода вместо объема M используют число информационных символов в кодовом слове $k = \log M$

5.1.3. Важнейшие границы теории кодирования.

Значительная часть работ по теории кодирования посвящена установлению и исследованию различных границ для кодов и, в частности, границ для кодового расстояния. Указанные границы представляют собой ориентиры, позволяющие объективно судить, насколько успешно решены задачи построения кодов. Наиболее важными и полезными границами кодового расстояния являются *границы Хэмминга, Плоткина и Гильберта*.

Первые две из упомянутых границ указывают предел максимизации кодового расстояния d при заданных длине n и скорости R кода. Граница же Гильберта является границей существования и дает нижнюю оценку кодового расстояния «наилучшего» кода.

Граница Хэмминга. Для любого двоичного кода, содержащего M кодовых слов длины n и исправляющего α (и менее) ошибок, выполняется соотношение

$$n - k \geq \log \left[\sum_{i=0}^{\alpha} C_n^i \right], \quad (5.4)$$

устанавливающее минимальное число проверочных символов, необходимое для обеспечения требуемой корректирующей способности,

где $C_n^i = \frac{n!}{(n-i)!i!}$ - биномиальный коэффициент.

Граница Плоткина. Для любого двоичного блочного кода с параметрами $M = 2^k$, n и d выполняется неравенство

$$d \leq \frac{Mn}{2(M-1)} = \frac{n2^{k-1}}{2^k - 1}. \quad (5.5)$$

Граница Гильберта. Безусловно существует двоичный блочный код, параметры которого удовлетворяют неравенству

$$(M-1) \sum_{i=0}^{d-1} C_n^i < 2^n. \quad (5.6)$$

Если $k = \log M$ целое число, то существует более строгий вариант границы (5.6), называемый *границей Варшавова–Гильберта*

$$\sum_{i=0}^{d-2} C_{n-1}^i < 2^{n(1-R)}. \quad (5.7)$$

Рис. 5.1 служит графической иллюстрацией асимптотических границ (5.5-5.7).

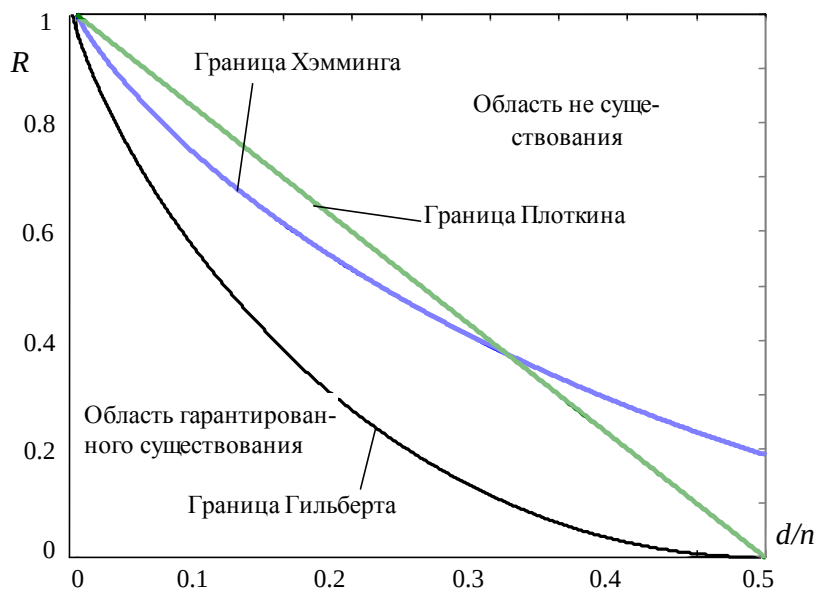


Рис. 5.1. Границы кодирования.

Область, лежащая выше хотя бы одной из границ (5.6-5.7), отвечает тем значениям M, n, d , при которых не существуют коды с указанными параметрами. Не составляет труда определить, что точка пересечения границ Хэмминга и Плоткина, устанавливающая диапазон использования соответствующей границы, отвечает значению $d/n = 0,312$. С другой стороны, те значения M, n, d , которые соответствуют области, лежащей ниже границы Гильберта, гарантируют существование кода. Область, лежащая между двумя упомянутыми зонами, отвечает неопределенной ситуации, для которой точный ответ о существовании кода не может быть дан на основании приведенных границ.

5.1.4 Скорость небредачи кода

Важной характеристикой кода является *скорость кода*, определяемая соотношением

$$R = \frac{\Delta \log M}{n} . \quad (5.8)$$

под которой понимается количество информации, приходящееся на один символ равновероятных кодовых слов. Из (5.1) следует, что скорость кода представляет собой удельную энтропию множества M равновероятных кодовых слов длины n и измеряется в битах на символ. При заданной длительности символа кода T_0 скорость кода легко пересчитывается в *скорость передачи*

$$R' = \frac{\Delta \log M}{nT_0} = \frac{R}{T_0} . \quad (5.9)$$

Параметр (5.2) показывает, какое количество бит информации передается за единицу времени, и измеряется в бит/сек.

Скорость кода является важнейшей характеристикой, так как входит в условия прямой и обратной теорем помехоустойчивого кодирования. Она определяет ту избыточность, которая вводится в кодовые слова для противодействия помехам в канале связи. Для двоичных кодов, в словах которых символы могут принимать только 2 значения, скорость лежит в диапазоне от 0 до 1. При $R=1$ в коде отсутствует избыточность и, следовательно, он не способен ни обнаруживать, ни исправлять ошибки. Избыточность кода уменьшает скорость, но, как показывает пример 5.1, позволяет понизить вероятность ошибочного приема.

5.1.5 Энергетический выигрыш от кодирования

Под термином «*выигрыш от кодирования*» понимают количественный показатель повышения достоверности передачи информации в цифровой системе связи при использовании кодирования со

случае его отсутствия (т.е. безызбыточного кодирования). Точное его значение показывает, насколько может быть уменьшена энергия сигнала при кодировании в сравнении со случаем передачи не кодированного потока данных. При этом должна обеспечиваться одинаковая достоверность (оцениваемая по вероятности ошибочного решения) и скорость передачи в бит/сек. Простейший вариант оценки выигрыша от кодирования базируется на сравнении минимума евклидова расстояния между блоками некодированных данных и аналогичной характеристикой для ближайших кодовых слов рассматриваемого блокового кода. Следует отметить, что результат сравнения расстояний зависит от физического сигнала, используемого для отображения двоичного информационного символа.

Безызбыточное кодирование.

Пусть требуется передавать $M = 2^k$ сообщений при длительности кодового слова, равной T . Учитывая безызбыточность кодирования, длительность одного кодового символа (бита) составляет

$$\Delta_{\delta} = \frac{T}{k} = \frac{1}{R'}, \quad (5.10)$$

где $R' = \frac{\log M}{T} = \frac{k}{T} [\text{бит} / \text{с}]$ - скорость передачи информации.

Предположим, что мощность передатчика равна P , тогда энергия, приходящаяся на один бит информации, будет

$$E_{\delta} = P \Delta_{\delta} = \frac{P}{R'}. \quad (5.11)$$

Пусть отображение информационного потока осуществляется с помощью бинарной фазовой модуляции, тогда минимальное евклидово расстояние будет между блоками данных, отличающихся только одним информационным символом. Например, кодовыми блоками вида – 01000 и 00000, иллюстрацией чему служит диаграмма на рис. 5.2, а. Поскольку на $k-1$ позициях вектора совпадают, а отличие заключается только в одной позиции, которой соответствуют противоположные сигналы, то

$$d_{\min E, 1} = 2\sqrt{E_{\delta}}, \quad (5.12)$$

где индекс «1» отвечает первому варианту кодирования, и значит,

$$d_{\min E, 1}^2 = 4E_{\delta} = \frac{4P}{R'} . \quad (5.13)$$

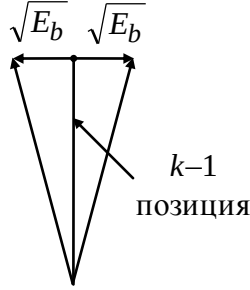


Рис. 5.2., а.

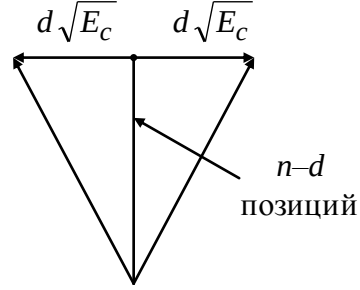


Рис. 5.2, б.

Избыточное кодирование.

Рассмотрим теперь вариант, отвечающий введению дополнительной избыточности. В этом случае для сохранения прежней скорости передачи информации R' длительность кодового символа уменьшается $\Delta_c < \Delta_{\delta}$, поскольку за время T передаются не k , а n символов кода ($n > k$). При неизменной мощности передатчика энергия, приходящаяся на один кодовый символ, определится как

$$E_c = P\Delta_c = P \frac{\Delta_c}{\Delta_{\delta}} \Delta_{\delta} = P \frac{T k}{n T} \Delta_{\delta} = PR\Delta_{\delta}, \quad (5.14)$$

где $R = k / n$ - скорость используемого блочного кода.

При кодировании с избыточностью минимальное евклидово расстояние определяется кодовым расстоянием d , иллюстрацией чего служит диаграмма на рис. 5.2, б, на которой ближайшие вектора совпадают на $n - d$ и отличаются на d позициях. Тогда при использовании бинарной фазовой модуляции

$$d_{\min E, 2} = \sqrt{\sum_{i=1}^n (s_{1,i} - s_{2,i})^2} = \sqrt{d(\sqrt{E_c} - (-\sqrt{E_c}))^2} = 2\sqrt{dE_c}, \quad (5.15)$$

где индекс «2» отвечает второму варианту кодирования, а

$$d_{\min E, 2}^2 = 4dE_c = 4dPR\Delta_{\delta} = \frac{4dPR}{R'}. \quad (5.16)$$

В итоге выигрыш от кодирования определится как

$$G = \frac{d_{\min E, 2}^2}{d_{\min E, 1}^2} = d \cdot R. \quad (5.17)$$

В заключении оценим предельную величину энергетического выигрыша от кодирования, воспользовавшись границей для кодового расстояния. Граница Плоткина (5.10) устанавливает, что

$$d \leq \frac{Mn}{2(M-1)}, \quad (5.18)$$

тогда, с учетом того, что $R = \frac{\log M}{n} = \frac{k}{n}$, получаем

$$G \leq \frac{k}{2} \frac{M}{M-1}. \quad (5.19)$$

5.2 Принципы обнаружения и исправления ошибок

Пусть для передачи сообщений используется некоторый код длины n и объема M . Это означает, что на вход дискретного канала поступает одна из последовательностей

$$\mathbf{X}_i = (x^{(1)}, x^{(2)}, \dots, x^{(n)}), \quad i = 1, 2, \dots, M; \quad \mathbf{X}_i \in X^n,$$

называемая кодовым словом.

На приемной стороне наблюдается некоторая выходная последовательность (вектор наблюдений)

$$\mathbf{Y} = (y^{(1)}, y^{(2)}, \dots, y^{(n)}); \quad \mathbf{Y} \in Y^n.$$

Процедуру решения о том, произошла ли ошибка при передаче кодовой последовательности или нет, можно описать на следующем языке. Все множество Y^n разбивается на две области G_0 и G_1 , причем область G_0 образована разрешенными (иначе кодовыми) последовательностями, а G_1 - запрещенными комбинациями.

Очевидно, что если наблюдаемая последовательность $\mathbf{Y}$, полученная в результате трансформации каналом некоторой кодовой комбинации $\mathbf{X}$,

оказывается в области G_1 , то принимается решение об обнаружении ошибок в принятой последовательности. Если же наблюдаемая последовательность окажется в области G_0 , то принимается решение о безошибочной передаче информации.

Естественно возникает вопрос о том, а все ли ошибки могут быть обнаружены?

Предположим, что алфавит входных $x^{(i)} \in X$ и выходных $y^{(i)} \in Y$ символов одинаков, и его объем равен $|X| = |Y| = q$. Тогда объем кода, а значит и мощность множества G_0 , составляют величину $M = q^k$, а число запрещенных комбинаций (или мощность множества G_1) определится как $M_3 = q^n - q^k$. При передаче по каналу связи M кодовых комбинаций возможны $q^k q^n$ их переходов в принятые наблюдения Y , из которых только q^k будут правильными, тогда как остальные $q^k q^n - q^k = q^k (q^n - 1)$ переходов сопровождаются искажениями. Как уже указывалось, решение об обнаружении ошибок в принятой последовательности принимается всякий раз, когда она оказывается в области G_1 . Подобной ситуации отвечают $MM_3 = q^k (q^n - q^k)$ переходов и, значит, общее число обнаруживаемых ошибок составляет величину $N_{\text{обн}} = q^k (q^n - q^k)$, что еще раз свидетельствует о возможности обнаружения ошибок только при условии $n > k$, т.е. при введении избыточных символов. Сравнение же общего количества ошибок $q^k (q^n - 1)$ с числом обнаруживаемых $N_{\text{обн}} = q^k (q^n - q^k)$ демонстрирует, что не все ошибки удастся зафиксировать. Последнее объясняется тем, что переход одной кодовой комбинации в другую под действием канальных помех невозможно обнаружить, причем общее количество подобных переходов составит величину $q^k (q^k - 1)$.

Аналогичным образом реализуется и процедура исправления ошибок. Отличие заключается лишь в том, что все множество Y^n разбивается теперь на M (по числу передаваемых сообщений) решающих областей G_i , $i = 1, 2, \dots, M$, $Y^n = \bigcup_{i=1}^M G_i$, причем G_i и G_j при $i \neq j$ не пересекаются, а в каждую область решения G_i включается только одна кодовая последовательность. Если оказывается, что вектор наблюдений Y принадлежит j -й области, т.е. $Y \in G_j$, то принимается решение о том, что было передано слово X_j и, значит, канальные ошибки, вызвавшие

трансформацию X_j в Y , будут исправлены. Поскольку области решений не перекрываются, то общее число исправляемых ошибок определяется числом запрещенных комбинаций, распределяемых между M решающими областями. Следовательно, $N_{\text{исп}} = M_3 = q^n - q^k$ и, значит, как и в случае обнаружения ошибок, их исправление возможно лишь при $n > k$, т.е. при введении избыточности. При передаче фиксированного кодового слова X_i ошибочное решение будет приниматься всякий раз, когда вектор наблюдений не принадлежит G_i . Поэтому вероятность ошибки при передаче i -го слова X_i (i -го сообщения) легко определить с помощью переходных вероятностей канала

$$p_e X_i = p Y \notin G_i = \sum_{Y \notin G_i} p Y | X_i ,$$

или в другой записи

$$p_e X = \sum_{Y \notin G(X)} p Y | X ,$$

где индекс i опущен без ущерба для понимания: $p_e X$ - вероятность ошибки при условии передачи фиксированного кодового вектора X , $G(X)$ - решающая область, отвечающая кодовому вектору X .

Поскольку такие вероятности ошибок составляют набор из M величин (по числу кодовых слов), для интегрального описания надежности передачи необходимо свести их к какому-либо единому показателю. Чаще всего это делается введением:

$$p_{e \max}^{\Delta} = \max_X p_e(X) \quad (5.20)$$

- максимальной вероятности ошибки декодирования, оценивающей влияние помех в канале в наихудшем случае, т.е. для наиболее уязвимого сообщения, или

$$p_e^{\Delta} = \sum_{X \in \{X_M\}} p X p_e X \quad (5.21)$$

- средней (полной) вероятности ошибки декодирования, рассчитываемой, как видно, согласно обычной формуле полной вероятности.

Из приведенных обсуждений очевидным образом следует, что основная задача, возникающая при осуществлении процедуры помехоустойчивого кодирования, состоит в выборе кодовых комбинаций и разбиении множества запрещенных комбинаций между областями решений, поскольку от правильного их осуществления зависит надежность передачи информации. Перед тем как перейти к рассмотрению данного вопроса введем некоторые необходимые для этого характеристики сигналов.

Как указывалось ранее, наблюдатель по принятой реализации $y(t)$ должен определить какой из M возможных сигналов $s_i(t)$, $i = 1, \dots, M$, отвечающих различным кодовым комбинациям X_i , в действительности был передан. Очевидно, что данная задача аналогична различению M сигналов, причем вероятность перепутывания сигналов, обусловленная наличием шума, в значительной степени определяется тем, насколько «далеко» друг от друга разнесены передаваемые сигналы в результате введения указанной избыточности, т.е. операции кодирования. В зависимости от модели канала распространения (непрерывный или дискретный) адекватной характеристикой, оценивающей упомянутое расстояние, а значит, и корректирующую способность метода кодирования (или кода), служат *Евклидово* или *Хэммингово расстояние*.

5.2.1. Декодирование по правилу максимального правдоподобия.

Рассмотрим задачу, необходимую решать любой системе передачи информации. На вход канала поступает один из M сигналов, отвечающих M возможным сообщениям: $s_1(t), s_2(t), \dots, s_M(t)$. На выходе канала наблюдателю доступно колебание $y(t)$, представляющее собой переданный сигнал, искаженный канальным шумом (см. Рис. 5.3). Располагая только колебанием $y(t)$, по какому правилу должен действовать блок решения, чтобы решение о том, какое сообщение было передано, было наиболее достоверным? Иными словами, какой из возможных алгоритмов принятия решения является оптимальным? В случае равной вероятности M входных сигналов и использовании в качестве критерия величины вероятности перепутывания передаваемых сигналов оптимальным является *правило максимального правдоподобия*: решающее устройство отдает предпочтение сигналу, наиболее «похожему» на принятое колебание.

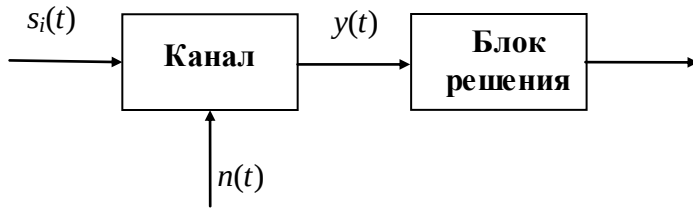


Рис. 5.3.

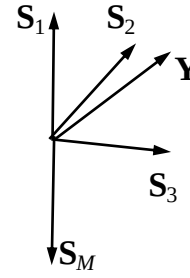


Рис. 5.4.

Похожесть означает, что переходная вероятность $p(y(t)|s_i(t))$ для сигнала, в пользу которого вынесено решение, является наибольшей, иначе говоря, максимальна вероятность трансформации сигнала $s_i(t)$ в колебание $y(t)$ под действием канального шума. Таким образом, правило максимального правдоподобия можно записать в виде

$$p(y(t)|s_i(t)) \stackrel{\hat{H}_i}{\geq} p(y(t)|s_j(t)), \quad j=1, \dots, M, i \neq j, \quad (5.22)$$

где $\hat{H}_i$ - решение о том, что был передан сигнал $s_i(t)$.

Для гауссовского канала наибольшая похожесть сигнала $s_i(t)$ принятому колебанию $y(t)$ эквивалентна минимуму евклидова расстояния между ними, поскольку переходная вероятность $p(y(t)|s_i(t))$ экспоненциально падает с ростом квадрата расстояния $d_E(\mathbf{Y}, \mathbf{S}_i)$. Тогда, как следует из иллюстрации на рис. 5.4, наиболее похожим на принятое колебание является сигнал $s_2(t)$, поскольку $d_E(\mathbf{Y}, \mathbf{S}_2) < d_E(\mathbf{Y}, \mathbf{S}_i), i=1, 3, \dots, M$. Следовательно, правило (5.22) эквивалентно правилу

$$d_E(\mathbf{Y}, \mathbf{S}_i) \stackrel{\hat{H}_i}{\leq} d_E(\mathbf{Y}, \mathbf{S}_j), \quad j=1, \dots, M, i \neq j. \quad (5.23)$$

В случае ДСК правило максимального правдоподобия записывается в виде

$$p(\mathbf{Y}|\mathbf{X}_i) \stackrel{\hat{H}_i}{\geq} p(\mathbf{Y}|\mathbf{X}_j), \quad j=1, \dots, M, i \neq j \quad (5.24)$$

где $p(\mathbf{Y}|\mathbf{X}_i)$ - переходная вероятность, т.е. вероятность того, что

двоичный кодовый вектор $\mathbf{X}_i$ трансформируется в наблюдение $\mathbf{Y}$, определяемая соотношением

$$P(\mathbf{Y}|\mathbf{X}_i) = p^{d_i} (1-p)^{n-d_i}, \quad (5.25)$$

где $d_i = d_H(\mathbf{Y}, \mathbf{X}_i)$ - расстояние Хэмминга между наблюдаемой последовательностью $\mathbf{Y}$ и передаваемым кодовым словом $\mathbf{X}_i$, p - вероятность искажения символа в ДСК. Поскольку, как правило $p < 0.5$, то $P(\mathbf{Y}|\mathbf{X}_i)$ является убывающей функцией d_i и, значит, максимально похожим на принятую последовательность $\mathbf{Y}$ будет то кодовое слово $\mathbf{X}_i$, расстояние Хэмминга для которого будет минимально. Очевидно, что правило (5.12) может быть переформулировано в виде

$$d_i(\mathbf{Y}, \mathbf{X}_i) \leq d_j(\mathbf{Y}, \mathbf{X}_j), \quad j=1, \dots, M, i \neq j. \quad (5.26)$$

Таким образом, для рассмотренных моделей канала декодирование по правилу максимального правдоподобия и минимума расстояния эквивалентны. Единственный факт, который необходимо учитывать, это различие в определении расстояния для непрерывного и дискретного каналов.

На основании приведенного рассмотрения очевиден следующий алгоритм построения корректирующего кода, а также способ формирования областей решения. С одной стороны, поскольку вероятность перепутывания сигналов падает с ростом расстояния между ними, то в качестве кодовых комбинаций следует брать последовательности, находящиеся на максимально возможном расстоянии друг от друга. С другой стороны, вероятность трансформации некоторого сигнала $s_i(t)$ в наблюдаемое колебание $y(t)$, либо некоторой кодовой последовательности $\mathbf{X}_i$ в вектор наблюдения $\mathbf{Y}$, тем меньше, чем в большем количестве позиций они отличаются друг от друга. Тогда для уменьшения вероятности принятия ошибочного решения в решающую область следует включать такие запрещенные комбинации, которые находятся на наименьшем расстоянии от кодового слова, содержащегося в данной области решения.

5.2.2 Мягкое и жесткое декодирование

Рассмотрим ситуацию, когда двоичный сигнал передается за отрезок времени $(0, T)$, причем двоичная единица представляется сигналом $s_1(t)$, а двоичный нуль - сигналом $s_2(t)$. Принятый сигнал имеет вид $y(t) = s_i(t) + n(t)$, где $n(t)$ представляет собой вклад гауссового шума с нулевым средним. Детектирование $y(t)$ в два основных этапа.

На первом этапе принятый сигнал переводится в число $z(T) = a_i + n_0$, где a_i - это компонент сигнала $z(T)$, а n_0 - компонент шума. Компонент шума n_0 - это *случайная переменная*, значения которой имеют *гауссово* распределение с нулевым средним. Следовательно, $z(T)$ также будет *случайной гауссовой величиной* со средним a_1 или a_2 , в зависимости от того, какая величина была отправлена - двоичная единица или двоичный нуль.

На втором этапе процесса детектирования принимается решение о том, какой сигнал был передан. Это решение принимается на основе сравнения $z(T)$ с порогом. Условные вероятности $z(T)$, $p(z|s_1)$ и $p(z|s_2)$, показанные на рис. 5.5, обозначены как правдоподобие s_1 и s_2 . Демодулятор преобразует упорядоченный по времени набор случайных переменных $\{z(T)\}$ в кодовую последовательность Z и подает ее на декодер.

Выход демодулятора можно настроить по-разному. Можно реализовать его в виде *жесткой схемы принятия решений* относительно того, представляет ли $z(T)$ единицу или нуль. В этом случае выход демодулятора квантуется на два уровня, нулевой и единичный, и соединяется с декодером. Поскольку декодер работает в режиме жесткой схемы принятия решений, принятых демодулятором, такое декодирование называется *жестким*.

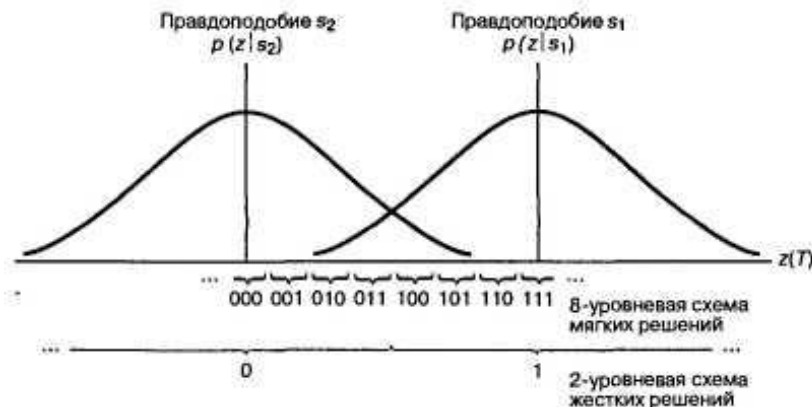


Рис. 5.5. Жесткая и мягкая схемы декодирования.

Аналогично демодулятор можно настроить так, чтобы он подавал на декодер значение $z(T)$, квантованное более чем на два уровня. Такая схема обеспечивает декодер большим количеством информации, чем жесткая схема решений. Если выход демодулятора имеет более двух уровней квантования, то декодирование называется *мягким*. На рис. 5.5 на оси

абсцисс изображено восемь (3-битовых) уровней квантования. Если в демодуляторе реализована жесткая схема принятия двоичных решений, он отправляет на декодер только один двоичный символ. Если в демодуляторе реализована мягкая двоичная схема принятия решений, квантованная на восемь уровней, он отправляет на декодер 3-битовое слово, описывающее интервал, соответствующий $z(T)$. По сути, поступление такого 3-битового слова, вместо одного двоичного символа, эквивалентно передаче декодеру *меры достоверности* вместе с решением относительно кодового символа. Согласно рис. 5.5, если с демодулятора поступила на декодер последовательность 111, это равносильно утверждению, что с очень высокой степенью достоверности кодовым символом была 1, в то время как переданная последовательность 100 равносильна утверждению, что с очень низкой степенью достоверности кодовым символом была 1. Совершенно ясно, что, в конечном счете, каждое решение, принятое декодером и касающееся сообщения, должно быть жестким; в противном случае на распечатках компьютера можно было бы увидеть нечто, подобное следующему: "думаю, это 1", "думаю, это 0" и т.д. То, что после демодулятора *не принимается жесткое решение* и на декодер поступает больше данных (мягкое принятие решений), можно понимать как промежуточный этап, необходимый для того, чтобы на декодер поступило больше информации, с помощью которой он затем сможет восстановить последовательность сообщения (с более высокой достоверностью передачи сообщения по сравнению с декодированием в рамках жесткой схемы принятия решений).

5.3 Построение линейных помехоустойчивых кодов

Рассмотрим множество V_F , состоящее из 2^n всех возможных n -компонентных векторов $\mathbf{X} = x_1, x_2, \dots, x_n$, элементы которого $x_i \in GF(2)$. Очевидно, что V_F образует n -мерное векторное пространство. Выберем в этом пространстве $k < n$ линейно независимых векторов $g_1, g_2, \dots, g_k$, что всегда возможно, поскольку в n -мерном пространстве всегда существуют n линейно независимых векторов.

Построим множество U , содержащее 2^k векторов, образованных как линейная комбинация вид

$$U = \sum_{i=1}^k a_i g_i = a_1 g_1 + a_2 g_2 + \dots + a_k g_k, \quad a_i \in GF(2). \quad (5.27)$$

Это подпространство имеет размерность k и непосредственно является той конструкцией, которую назовем *линейным кодом* U .

5.3.1 Порождающая матрица линейного кода

Введем в рассмотрение *порождающую (производящую)* матрицу $\mathbf{G}$ линейного кода (n, k) , строками которой служат вектора $\mathbf{g}_i = (g_{i1}, g_{i2}, \dots, g_{in}), i = \overline{1, k}$:

$$\mathbf{G} = \begin{bmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \\ \dots \\ \mathbf{g}_k \end{bmatrix} = \begin{bmatrix} g_{11} & g_{12} & \dots & g_{1n} \\ g_{21} & g_{22} & \dots & g_{2n} \\ \dots & \dots & \dots & \dots \\ g_{k1} & g_{k2} & \dots & g_{kn} \end{bmatrix}.$$

Рассмотрим передаваемое сообщение из k бит: $\mathbf{m} = m_1, m_2, \dots, m_k$. Тогда формируемое кодовое слово можно представить в следующем виде

$$\mathbf{U} = m_1 \mathbf{g}_1 + m_2 \mathbf{g}_2 + \dots + m_k \mathbf{g}_k = m_1, m_2, \dots, m_k \begin{bmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \\ \dots \\ \mathbf{g}_k \end{bmatrix} = \mathbf{mG} \quad (5.28)$$

Следует отметить, что любая сумма слов линейного кода вновь дает кодовое слово, принадлежащее данному коду: если $\mathbf{U}_1, \mathbf{U}_2 \in \mathbf{U}$, то $\mathbf{U}_1 + \mathbf{U}_2 \in \mathbf{U}$.

Минимальное расстояние линейного кода равно наименьшему из весов ненулевых слов кода

$$d = \min_{\mathbf{U}_i \neq \mathbf{0}} w \mathbf{U}_i.$$

Следовательно, для определения кодового расстояния линейных кодов достаточно определить веса $M - 1$ ненулевых векторов, а не осуществлять перебор $M(M - 1)/2 \gg M$ пар кодовых слов.

Любой линейный код всегда может быть преобразован в *эквивалентный* (т.е. обладающий аналогичными параметрами M, n, d), которому отвечает *каноническая* (стандартная) порождающая матрица

$$\mathbf{G} = \left[\mathbf{P} : \mathbf{I}_k \right] = \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1 \ n-k} & 1 & 0 & \dots & 0 \\ p_{21} & p_{22} & \dots & p_{2 \ n-k} & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ p_{k1} & p_{k2} & \dots & p_{k \ n-k} & 0 & 0 & \dots & 1 \end{bmatrix}. \quad (5.29)$$

Здесь $\mathbf{P}$ - массив четности, содержащий элементы $p_{ij} = 0,1, i=1,...,k, j=1,..., n-k$; I_k - единичная матрица размерностью $k \times k$.

Использование канонической порождающей матрицы (5.29) позволяет построить систематический линейный код. Систематический линейный код n, k - это такое отображение k - мерного вектора сообщения $m = m_1, m_2, ..., m_k$ в n -мерное кодовое слово $U = u_1, u_2, ..., u_n$, в котором часть формируемой кодовой последовательности совмещается с k символами сообщения. Остальные $n-k$ бит - это биты четности: $p_1, p_2, ..., p_{n-k}, m_1, m_2, ..., m_k$. Систематические кодовые слова иногда записываются так, чтобы биты сообщения занимали левую часть кодового слова, а биты четности - правую: $m_1, m_2, ..., m_k, p_1, p_2, ..., p_{n-k}$. Такая перестановка не влияет на свойства кода, связанные с процедурами обнаружения и исправления ошибок.

Рассмотрим передаваемое сообщение из k бит: $m = m_1, m_2, ..., m_k$.

Тогда используя формулу (5.29), формируемое кодовое слово можно представить в следующем виде:

$$U = u_1, u_2, ..., u_n = [m_1, m_2, ..., m_k] \cdot \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1, n-k} & 1 & 0 & \dots & 0 \\ p_{21} & p_{22} & \dots & p_{2, n-k} & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ p_{k1} & p_{k2} & \dots & p_{k, n-k} & 0 & 0 & \dots & 1 \end{bmatrix},$$

где
$$u_i = \begin{cases} m_1 p_{1i} + m_2 p_{2i} + \dots + m_k p_{ki}, & \text{для } i = 1, \dots, n-k \\ m_{i-n-k}, & \text{для } i = n-k+1, \dots, n \end{cases}$$

Систематический кодовый вектор можно записать в следующем виде

где
$$U = \underbrace{p_1, p_2, \dots, p_{n-k}}_{\text{биты четности}}, \underbrace{m_1, m_2, \dots, m_k}_{\text{биты сообщения}},$$

$$p_1 = m_1 p_{11} + m_2 p_{21} + \dots + m_k p_{k1}$$

$$p_2 = m_1 p_{12} + m_2 p_{22} + \dots + m_k p_{k2}$$

$$\dots$$

$$p_{n-k} = m_1 p_{1, n-k} + m_2 p_{2, n-k} + \dots + m_k p_{k, n-k}$$

В качестве примера воспользуемся кодом (6,3) с порождающей матрицей вида

$$G = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}.$$

и рассмотрим структуру кодового слова. Соответствие векторов сообщения $\mathbf{m}$ и кодовых слов $\mathbf{U}$ представлены в таблице 5.3.1

Табл. 5.3.1.

$\mathbf{m}_0 = (0 \ 0 \ 0)$	$\mathbf{U}_0 = (0 \ 0 \ 0 \ 0 \ 0 \ 0)$
$\mathbf{m}_1 = (1 \ 0 \ 0)$	$\mathbf{U}_1 = (1 \ 1 \ 0 \ 1 \ 0 \ 0)$
$\mathbf{m}_2 = (0 \ 1 \ 0)$	$\mathbf{U}_2 = (0 \ 1 \ 1 \ 0 \ 1 \ 0)$
$\mathbf{m}_3 = (1 \ 1 \ 0)$	$\mathbf{U}_3 = (1 \ 0 \ 1 \ 1 \ 1 \ 0)$
$\mathbf{m}_4 = (0 \ 0 \ 1)$	$\mathbf{U}_4 = (1 \ 0 \ 1 \ 0 \ 0 \ 1)$
$\mathbf{m}_5 = (1 \ 0 \ 1)$	$\mathbf{U}_5 = (0 \ 1 \ 1 \ 1 \ 0 \ 1)$
$\mathbf{m}_6 = (0 \ 1 \ 1)$	$\mathbf{U}_6 = (1 \ 1 \ 0 \ 0 \ 1 \ 1)$
$\mathbf{m}_7 = (1 \ 1 \ 1)$	$\mathbf{U}_7 = (0 \ 0 \ 0 \ 1 \ 1 \ 1)$

Как отмечалось ранее, код исправляет любые ошибки кратности α и менее в том и только в том случае, если кодовое расстояние удовлетворяет неравенству $d \geq 2\alpha + 1$. В приведенном примере расстояние Хэмминга кодовых слов равно $d=3$. Следовательно, представленные кодовые слова могут обнаруживать и исправлять все одиночные канальные ошибки.

5.3.2 Проверочная матрица линейного кода

Проверочная матрица $\mathbf{H}$ позволит нам декодировать полученные вектора. Для каждой порождающей матрицы $\mathbf{G}$ размерности $k \times n$ существует проверочная матрица $\mathbf{H}$ размером $k \times n - k$, такая, что строки матрицы $\mathbf{G}$ ортогональны к строкам матрицы $\mathbf{H}$.

Чтобы матрица $\mathbf{H}$ удовлетворяла требованиям ортогональности $\mathbf{U} \cdot \mathbf{H}^T = 0$ систематического кода, ее компоненты записываются в следующем виде

$$\mathbf{H} = \begin{bmatrix} \mathbf{I}_{n-k} & \mathbf{P}^T \end{bmatrix}. \quad (5.30)$$

Следовательно, проверочная матрица $\mathbf{H}^T$ имеет следующий вид

$$\mathbf{H}^T = \begin{bmatrix} \mathbf{I}_{n-k} \\ \mathbf{P} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \\ p_{11}p_{12}\dots p_{1\ n-k} \\ p_{21}p_{22}\dots p_{2\ n-k} \\ \dots & \dots & \dots & \dots \\ p_{k1}p_{k1}\dots p_{k\ n-k} \end{bmatrix}.$$

Рассматривая структуру матриц $\mathbf{H}$ и $\mathbf{G}$ можно вынести следующее заключение. Обе матрицы состоят из множества линейно независимых векторов, поскольку наличие в их структуре единичной матрицы делает невозможным существование линейной комбинации строк с нулевой суммой.

Учитывая сказанное выше, можно их «поменять ролями» и использовать $\mathbf{H}$ в качестве порождающей, а $\mathbf{G}$ – проверочной матрицы некоторого другого кода. Коды, связанные с таким преобразованием, называются *дуальными* друг другу. Таким образом, если исходным являлся (n, k) кодом, то дуальным ему будет $(n, n-k)$ код.

Кроме того, из заданного (n, k) кода можно получить новый *эквивалентный* код, обладающий теми же параметрами. Два кода эквивалентны тогда и только тогда, когда их порождающие матрицы получаются одна из другой посредством перестановки столбцов, либо в результате элементарных операций над строками.

Минимальное расстояние (минимальный вес) любого линейного (n, k) кода удовлетворяет неравенству:

$$d \leq n - k + 1.$$

Для приведенного примера проверочная матрица может иметь вид

$$\mathbf{H} = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}.$$

Модель одиночных канальных ошибок и соответствующие им синдромы представлены в таблице 5.3.2.

Табл. 5.3.2.

$e_0 = (0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0)$	$S_0 = (0 \ 0 \ 0)$
$e_1 = (0 \ 0 \ 0 \ 0 \ 0 \ 1 \ 0)$	$S_1 = (1 \ 0 \ 1)$
$e_2 = (0 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0)$	$S_2 = (0 \ 1 \ 1)$
$e_3 = (0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0)$	$S_3 = (1 \ 1 \ 0)$
$e_4 = (0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 0)$	$S_4 = (0 \ 0 \ 1)$
$e_5 = (0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0)$	$S_5 = (0 \ 1 \ 0)$
$e_6 = (1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0)$	$S_6 = (1 \ 0 \ 0)$

5.3.3 Декодирование и исправление ошибок

Пусть $E = e_1, e_2, \dots, e_n$ - вектор ошибки или модель ошибки, внесенная каналом. Тогда вектора наблюдений принятого сигнала при воздействии канальной ошибки можно представить в виде

$$U_r = U + E. \quad (5.31)$$

Одной из основных причин большой популярности линейных кодов служит возможность применения метода декодирования по минимуму расстояния не путем непосредственного вычисления расстояния Хэмминга между принятой последовательностью U_r и всеми кодовыми словами U , а в особой форме, используя такое понятие, как *синдром*.

Под синдромом понимается $r = (n - k)$ - компонентный вектор вида

$$S_r = U_r \cdot H^T. \quad (5.32)$$

Очевидно, что для любого двоичного кода существует 2^r различных значений синдрома S_r .

Синдромное детектирование позволяет построить декодер с минимальными аппаратными или временными затратами.

Используя уравнения (5.31) и (5.32), мы можем представить синдром вектора наблюдений U_r в следующем виде

$$S_r = U_r \cdot H^T = U \cdot H^T + E \cdot H^T. \quad (5.33)$$

Для всех элементов набора кодовых слов имеем $U \cdot H^T = 0$ и поэтому синдром вектора наблюдений U_r равен

$$S_r = E \cdot H^T = S. \quad (5.34)$$

Из сказанного следует, что по вектору наблюдения U_r можно обнаружить ошибку и установить синдром канальной помехи S .

Затем, используя таблицу соответствия синдромов, находим соответствующую модель ошибки, которая является оценкой ошибки (далее будем обозначать ее через E_r). Затем декодер прибавляет оценку E_r к вектору наблюдения U_r и оценивает переданное кодовое слово

$$U_c = U_r + E_r = U + E + E_r = U + E + E_r \quad (5.35)$$

Если правильно вычислили ошибку: $E = E_r$, тогда оценка кодового слова U_c совпадает с переданным кодовым словом U . С другой стороны, если оценка ошибки неверна, декодер неверно определит переданное кодовое слово, и мы получим необнаружимую ошибку декодирования.

Рассмотрим процедуру обнаружения и исправления ошибок на рассматриваемом примере, если принимаются пять первых кодовых слов при воздействии случайной канальной ошибки. Исходные вектора кодовых слов U , вид канальной ошибки и вектора наблюдений принятых кодов U_r представлены в таблице (5.3.3)

Табл. 5.3.3.

$U_0 = (0 \ 0 \ 0 \ 0 \ 0 \ 0)$	$er_0 = (0 \ 0 \ 0 \ 0 \ 0 \ 0)$	$U_r_0 = (0 \ 0 \ 0 \ 0 \ 0 \ 0)$
$U_1 = (1 \ 1 \ 0 \ 1 \ 0 \ 0)$	$er_1 = (0 \ 0 \ 0 \ 0 \ 1 \ 0)$	$U_r_1 = (1 \ 1 \ 0 \ 1 \ 1 \ 0)$
$U_2 = (0 \ 1 \ 1 \ 0 \ 1 \ 0)$	$er_2 = (0 \ 0 \ 0 \ 0 \ 0 \ 0)$	$U_r_2 = (0 \ 1 \ 1 \ 0 \ 1 \ 0)$
$U_3 = (1 \ 0 \ 1 \ 1 \ 1 \ 0)$	$er_3 = (0 \ 0 \ 1 \ 0 \ 0 \ 0)$	$U_r_3 = (1 \ 0 \ 0 \ 1 \ 1 \ 0)$
$U_4 = (1 \ 0 \ 1 \ 0 \ 0 \ 1)$	$er_4 = (0 \ 0 \ 0 \ 1 \ 0 \ 0)$	$U_r_4 = (1 \ 0 \ 1 \ 1 \ 0 \ 1)$

Принятые вектора наблюдений U_r позволяют определить синдром ошибки S_r и по таблице соответствия синдромов (5.3.2) установить вектор ошибки. Результаты представлены в таблице 5.3.4.

Табл. 5.3.4.

$Sr_0 = (0 \ 0 \ 0)$		$er_0 = (0 \ 0 \ 0 \ 0 \ 0 \ 0)$
$Sr_1 = (0 \ 1 \ 1)$		$er_1 = (0 \ 0 \ 0 \ 0 \ 1 \ 0)$
$Sr_2 = (0 \ 0 \ 0)$	$\Rightarrow$	$er_2 = (0 \ 0 \ 0 \ 0 \ 0 \ 0)$
$Sr_3 = (0 \ 0 \ 1)$		$er_3 = (0 \ 0 \ 1 \ 0 \ 0 \ 0)$
$Sr_4 = (1 \ 1 \ 0)$		$er_4 = (0 \ 0 \ 0 \ 1 \ 0 \ 0)$

Далее прибавляя оценку $\mathbf{E}_r$ к вектору наблюдения $\mathbf{U}_r$ получаем принятое кодовое слово $\mathbf{U}_c$. Операции исправления ошибок приведены в таблице 5.3.5.

Табл. 5.3.5.

$\mathbf{U}_r = (0 \ 0 \ 0 \ 0 \ 0 \ 0)$	$\mathbf{e}_r = (0 \ 0 \ 0 \ 0 \ 0 \ 0)$	$\mathbf{U}_c = (0 \ 0 \ 0 \ 0 \ 0 \ 0)$
$\mathbf{U}_r = (1 \ 1 \ 0 \ 1 \ 1 \ 0)$	$\mathbf{e}_r = (0 \ 0 \ 0 \ 0 \ 1 \ 0)$	$\mathbf{U}_c = (1 \ 1 \ 0 \ 1 \ 0 \ 0)$
$\mathbf{U}_r = (0 \ 1 \ 1 \ 0 \ 1 \ 0)$	$\mathbf{e}_r = (0 \ 0 \ 0 \ 0 \ 0 \ 0)$	$\mathbf{U}_c = (0 \ 1 \ 1 \ 0 \ 1 \ 0)$
$\mathbf{U}_r = (1 \ 0 \ 0 \ 1 \ 1 \ 0)$	$\mathbf{e}_r = (0 \ 0 \ 1 \ 0 \ 0 \ 0)$	$\mathbf{U}_c = (1 \ 0 \ 1 \ 1 \ 1 \ 0)$
$\mathbf{U}_r = (1 \ 0 \ 1 \ 1 \ 0 \ 1)$	$\mathbf{e}_r = (0 \ 0 \ 0 \ 1 \ 0 \ 0)$	$\mathbf{U}_c = (1 \ 0 \ 1 \ 0 \ 0 \ 1)$

Поскольку оцененная модель ошибки совпадает с действительной моделью ошибки, процедура исправления ошибки дает $\mathbf{U}_c = \mathbf{U}$.

5.4 Построение ортогональных и симплексных линейных кодов.

5.4.1 Ортогональные линейные коды

Для построения ортогональных помехоустойчивых кодов построим матрицу, столбцы которой представляют собой все m -разрядные двоичные числа, расположенные в порядке их возрастания. Полученная таким образом $m \times 2^m$ матрица $\mathbf{G}_{\text{орт}}$ является порождающей матрицей 2^m , m ортогонального кода длины $n = 2^m$ с $k = m$.

Исходные информационные слова с объемом $M = 2^m = n$ представляют собой m -разрядные двоичные числа. Рассмотрим передаваемое сообщение из k бит: $m = m_1, m_2, \dots, m_k$. Тогда формируемое кодовое слово можно представить в следующем виде $\mathbf{U}_0 \in m \cdot \mathbf{G}_0$. Основанием для подобного названия кода служит тот факт, что при замене двоичных 0,1 символов кода на ± 1 каждое кодовое становится одной из ортогональных функций Уолша.

Очевидно, что веса всех ненулевых слов ортогонального кода одинаковы и равны $n/2$, следовательно, код длины $n = 2^m$ имеет минимальное расстояние $d = n/2 = 2^{m-1}$. Это означает возможность исправлять все ошибки кратности до значений $\alpha = (n/4) - 1 = 2^{m-2} - 1$ включительно.

Пример 5.4.1. Рассмотрим ортогональный код (8,3) с параметрами $k=m=3$, $n=2^3-1=8$ и $r=n-k=4$.

Для построения систематического линейного кода (8,3) составим таблицу двоичных чисел

$$\begin{array}{c} (0 \ 1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7) \\ \begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix} \end{array}$$

и выделим из неё (столбцы 1,2,4) единичную матрицу $\mathbf{I}$ размерности $k \times k$ и матрицу четности $\mathbf{P}$ размерности $r \times k$

$$\mathbf{I} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \mathbf{P} := \begin{pmatrix} 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{pmatrix}$$

Порождающая матрица ортогонального систематического линейного кода, построенная согласно приведенному правилу (5.29), примет вид:

$$\mathbf{G}_{орт} = \begin{pmatrix} 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 & 1 \end{pmatrix}$$

Проверочная матрица в соответствии с правилом (5.30) имеет форму

$$\mathbf{H}_{орт} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$$

Рассмотрим структуру кодовых слов $\mathbf{U}$. Для этого воспользуемся кодом (8,3) из примера 5.4.1. Соответствие векторов сообщения $\mathbf{m}$ и кодовых слов $\mathbf{U}$ представлены в таблице 5.4.1.

Табл. 5.4.1.

$\mathbf{m}_0 = (0 \ 0 \ 0)$	$\mathbf{U}_0 = (0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0)$
$\mathbf{m}_1 = (1 \ 0 \ 0)$	$\mathbf{U}_1 = (0 \ 1 \ 1 \ 0 \ 1 \ 1 \ 0 \ 0)$
$\mathbf{m}_2 = (0 \ 1 \ 0)$	$\mathbf{U}_2 = (0 \ 1 \ 0 \ 1 \ 1 \ 0 \ 1 \ 0)$
$\mathbf{m}_3 = (1 \ 1 \ 0)$	$\mathbf{U}_3 = (0 \ 0 \ 1 \ 1 \ 0 \ 1 \ 1 \ 0)$
$\mathbf{m}_4 = (0 \ 0 \ 1)$	$\mathbf{U}_4 = (0 \ 0 \ 1 \ 1 \ 1 \ 0 \ 0 \ 1)$
$\mathbf{m}_5 = (1 \ 0 \ 1)$	$\mathbf{U}_5 = (0 \ 1 \ 0 \ 1 \ 0 \ 1 \ 0 \ 1)$
$\mathbf{m}_6 = (0 \ 1 \ 1)$	$\mathbf{U}_6 = (0 \ 1 \ 1 \ 0 \ 0 \ 0 \ 1 \ 1)$
$\mathbf{m}_7 = (1 \ 1 \ 1)$	$\mathbf{U}_7 = (0 \ 0 \ 0 \ 0 \ 1 \ 1 \ 1 \ 1)$

Можно видеть, что вес любого ненулевого слова равен 4 и код способен исправлять не только любую однократную ошибку, но и обнаруживать любую двукратную.

Модель одиночных канальных ошибок для кода (8.3) и синдром вектора ошибок приведены в таблице 5.4.2.

Табл. 5.4.2.

$\mathbf{e}_0 = (0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0)$	$\mathbf{S}_0 = (0 \ 0 \ 0 \ 0 \ 0)$
$\mathbf{e}_1 = (1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0)$	$\mathbf{S}_1 = (1 \ 0 \ 0 \ 0 \ 0)$
$\mathbf{e}_2 = (0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0)$	$\mathbf{S}_2 = (0 \ 1 \ 0 \ 0 \ 0)$
$\mathbf{e}_3 = (0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0)$	$\mathbf{S}_3 = (0 \ 0 \ 1 \ 0 \ 0)$
$\mathbf{e}_4 = (0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 0)$	$\mathbf{S}_4 = (0 \ 0 \ 0 \ 1 \ 0)$
$\mathbf{e}_5 = (0 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0)$	$\mathbf{S}_5 = (0 \ 0 \ 0 \ 0 \ 1)$
$\mathbf{e}_6 = (0 \ 0 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0)$	$\mathbf{S}_6 = (0 \ 1 \ 1 \ 0 \ 1)$
$\mathbf{e}_7 = (0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1 \ 0)$	$\mathbf{S}_7 = (0 \ 1 \ 0 \ 1 \ 1)$
$\mathbf{e}_8 = (0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1)$	$\mathbf{S}_8 = (0 \ 0 \ 1 \ 1 \ 1)$

Дальнейшая процедура синдромного обнаружения и исправления ошибок описана в предыдущем разделе.

5.4.2 Симплексные линейные коды.

Учитывая тот факт, что первый символ всех слов ортогонального кода является нулевым, а значит, не несет никакой информации, существует возможность его исключения. Отражением этого шага является вычеркивание первого столбца в порождающей матрице ортогонального кода. В результате получаем порождающую матрицу

$2^m - 1, m$ симплексного кода:

$$\mathbf{G}_{\text{симп}} = \left[\begin{array}{c|c} \mathbf{0} & \mathbf{G}_{\text{орт}} \end{array} \right]$$

Название обусловлено симплексными сигналами, в которые данный код трансформируется при замене алфавита $\{0,1\}$ на $\{\pm 1\}$.

Длина помехоустойчивого кода $n = 2^m - 1$, длина информационного кодового расстояния и исправляющей способности кода, так что $d = 2^{m-1} = (n+1)/2$. Однако приводит к незначительному возрастанию скорости кода за счет уменьшения на единицу его длины. Следует отметить, что с точки зрения исправляющей способности симплексные коды являются наилучшими при заданных значениях n и $M = 2^m = n+1$.

Пример 5.4.2. Воспользовавшись результатами примера 5.4.1 и вычеркнув первый столбец матрицы $\mathbf{G}$ ортогонального кода, определяем порождающую матрицу симплексного $(7,3)$ кода в виде:

$$\mathbf{G}_{\text{симп}} = \begin{pmatrix} 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 \end{pmatrix}$$

Объединение транспонированной матрицы четности $\mathbf{P}$ матрице и единичной матрицы $\mathbf{I}$ размерности $r \times r$ определяет вид проверочной матрицы

$$\mathbf{H}_{\text{орт}} = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$$

Соответствие векторов сообщения $\mathbf{m}$ и кодовых слов $\mathbf{U}$ представлены в таблице 5.4.5

Табл. 5.4.5.

$\mathbf{m}_0 = (0 \ 0 \ 0)$	$\mathbf{U}_0 = (0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0)$
$\mathbf{m}_1 = (1 \ 0 \ 0)$	$\mathbf{U}_1 = (1 \ 1 \ 0 \ 1 \ 1 \ 0 \ 0)$
$\mathbf{m}_2 = (0 \ 1 \ 0)$	$\mathbf{U}_2 = (1 \ 0 \ 1 \ 1 \ 0 \ 1 \ 0)$
$\mathbf{m}_3 = (1 \ 1 \ 0)$	$\mathbf{U}_3 = (0 \ 1 \ 1 \ 0 \ 1 \ 1 \ 0)$
$\mathbf{m}_4 = (0 \ 0 \ 1)$	$\mathbf{U}_4 = (0 \ 1 \ 1 \ 1 \ 0 \ 0 \ 1)$
$\mathbf{m}_5 = (1 \ 0 \ 1)$	$\mathbf{U}_5 = (1 \ 0 \ 1 \ 0 \ 1 \ 0 \ 1)$
$\mathbf{m}_6 = (0 \ 1 \ 1)$	$\mathbf{U}_6 = (1 \ 1 \ 0 \ 0 \ 0 \ 1 \ 1)$
$\mathbf{m}_7 = (1 \ 1 \ 1)$	$\mathbf{U}_7 = (0 \ 0 \ 0 \ 1 \ 1 \ 1 \ 1)$

Модель одиночных канальных ошибок для кода (7.3) и синдром вектора ошибок приведены в таблице 5.4.6.

Табл. 5.4.6

$\mathbf{e}_0 = (1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0)$	$\mathbf{S}_0 = (1 \ 0 \ 0 \ 0)$
$\mathbf{e}_1 = (0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0)$	$\mathbf{S}_1 = (0 \ 1 \ 0 \ 0)$
$\mathbf{e}_2 = (0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0)$	$\mathbf{S}_2 = (0 \ 0 \ 1 \ 0)$
$\mathbf{e}_3 = (0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 0)$	$\mathbf{S}_3 = (0 \ 0 \ 0 \ 1)$
$\mathbf{e}_4 = (0 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0)$	$\mathbf{S}_4 = (1 \ 1 \ 0 \ 1)$
$\mathbf{e}_5 = (0 \ 0 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0)$	$\mathbf{S}_5 = (1 \ 0 \ 1 \ 1)$
$\mathbf{e}_6 = (0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1 \ 0)$	$\mathbf{S}_6 = (0 \ 1 \ 1 \ 1)$
$\mathbf{e}_7 = (0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1)$	$\mathbf{S}_7 = (0 \ 0 \ 0 \ 0)$

Дальнейшая процедура обнаружения и исправление ошибок полностью определяется методом синдромного декадирования.

6 МОДУЛЯЦИЯ ЦИФРОВОЙ ИНФОРМАЦИИ

6.1 Бинарная фазовая манипуляция сигналов

Пусть $B(t)$ информационный сигнал, в котором импульсы положительной и отрицательной полярности отвечают передаче бита информации, равного 0 или 1 соответственно: $d_i = 0, 1$. Пусть T_b длительность одного информационного бита и N_b число информационных бит в потоке сообщения $B(t)$. Тогда длительность информационного сообщения будет равна $T_{bs} = N_b \cdot T_b$. Информационное сообщение $B(t)$ можно описать выражением

$$B(t) = \sum_{i=0}^{N_b-1} b_i \cdot B_0 \cdot t - i \cdot T_b, \quad (6.1)$$

где $b_i = 1 - 2d_i$, $b_i \in \pm 1$ - информационные символы сообщения; $B_0(t)$ - одиночный импульс.

Модуляция с использованием бинарной фазовой модуляцией просто означает умножение информационного сигнала $B(t)$ на непрерывное гармоническое колебание с несущей частотой f_0 , позволяя прийти к передаваемому модулированному сигналу (см. рис. 6.1)

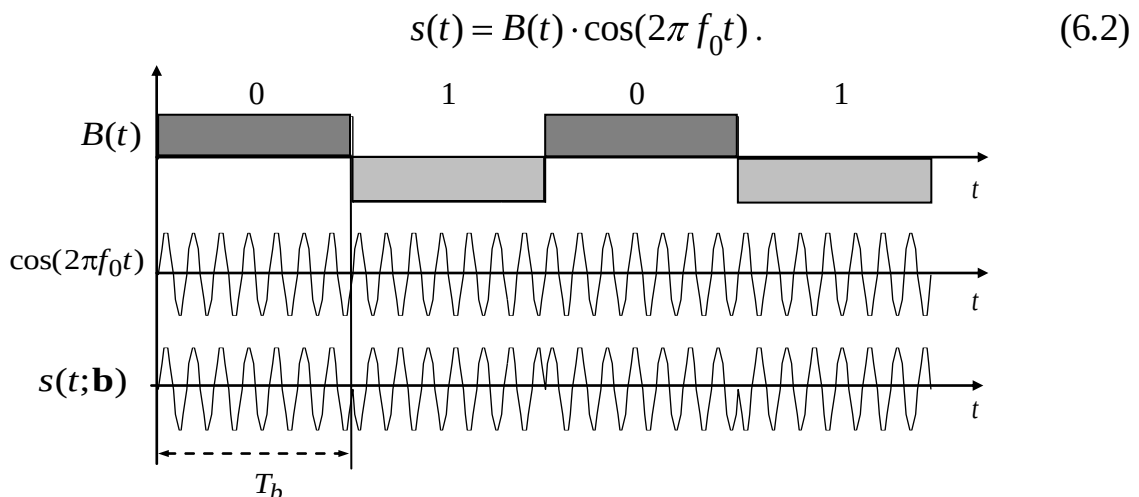


Рис. 6. 1. Иллюстрация передачи данных с помощью бинарной ФМ.

После распространения по каналу сигнал приобретает временную задержку τ_b и начальную фазу φ_b , а также испытывает ослабление, которое не учитывается в дальнейшем рассмотрении. Тогда принятый

полезный сигнал имеет вид

$$s_{\text{пр}}(t) = B(t - \tau_b) \cdot \cos(2\pi f_0 t + \varphi_b). \quad (6.3)$$

Типичный приемник сигналов с бинарной ФМ передаваемой информации содержит схемы слежения за задержкой и фазой несущей частоты. Поэтому приемнику известны «истинные» значения τ_b и φ_b .

Если рассматриваемый сигнал искажен шумом, то оптимальная (максимально правдоподобная) процедура восстановления i -го переданного бита состоит в вычислении корреляции наблюдения $y(t) = s_{\text{пр}}(t) + n(t)$ с разностью сигналов, отвечающих значениям информационного бита 0 и 1 соответственно, которая в рассматриваемом случае есть просто $2\cos(2\pi f_0 t + \varphi_b)$. Поскольку для решения о принятом бите используется только полярность корреляции и поскольку i -й бит на выходе канала занимает временной интервал $[(i-1)T_b + \tau_b, iT_b + \tau_b]$, то обсуждаемая корреляция представляет собой выражение

$$z_k = \int_{(i-1)T_b + \tau_b}^{iT_b + \tau_b} y(t) \cos(2\pi f_0 t + \varphi_b) dt, \quad (6.4)$$

а решение $\hat{b}_i = 0$ или $\hat{b}_i = 1$ принимается в зависимости от положительного или отрицательного знака z_k .

Практическая реализация бинарного ФМ модулятора представляет собой умножитель, представленный на рис. 6.2, а), а широко распространенная структура демодулятора, реализующего данное правило, представлена на рис. 6.2, б).

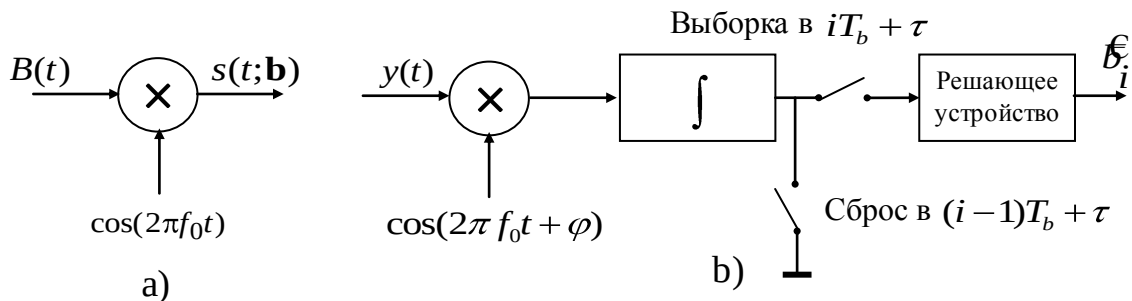


Рис. 6.2. Бинарный ФМ модулятор (а) и демодулятор (б).

Она содержит коррелятор, выполненный в виде перемножителя наблюдения $y(t)$ на генерируемое гармоническое колебание $\cos(2\pi f_0 t + \varphi)$ и интегратора со сбросом. В конце каждого последовательного битового

интервала с выхода интегратора берется отсчет, решение о текущем бите принимается в соответствии с его полярностью и после этого осуществляется обнуление интегратора, подготавливая его к работе на следующем битовом интервале.

6.2 Приём сигналов с бинарной фазовой манипуляцией

Основная трудность реализации демодулятора БФМ-сигналов состоит в создании синхронного с ними опорного колебания, имеющего неизменную начальную фазу.

Возможны два пути решения этой проблемы:

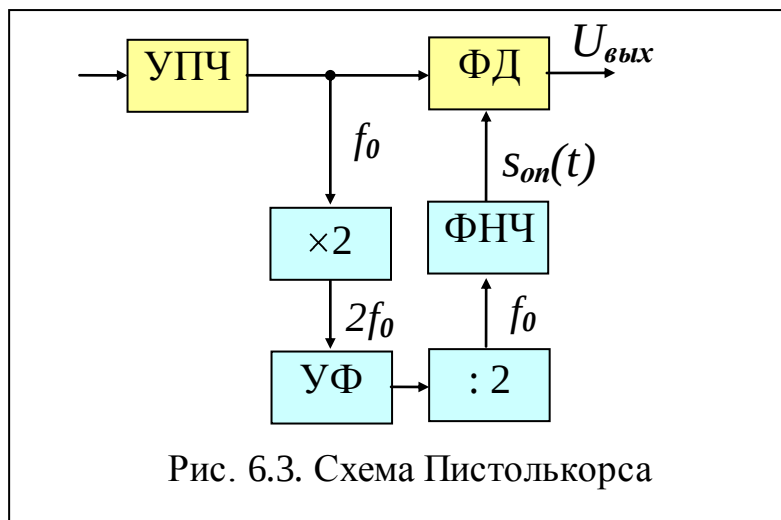
использование высокостабильного высокочастотного генератора:

формирование опорного колебания с помощью обработки входного сигнала.

Элементарные расчеты показывают, что для обеспечения длительного (более 20 с) сеанса связи с хранением в приемнике начальной фазы требуется генератор с чрезвычайно высокой относительной стабильностью - 10^{-9} . Кроме того, из-за нестабильности среды распространения начальная фаза принимаемого колебания испытывает флуктуации, которые невозможно учесть в данном случае.

Второй путь - создание опорного колебания из входного сигнала - также содержит проблемы. Основной является отсутствие несущей в спектре фазоманипулированного колебания при равновероятных символах. Тем не менее, было предпринято множество попыток создания схем формирования опорного колебания из входного. Далее кратко рассматриваются основные методы.

Одной из первых в 1933 году советским ученым А. А. Пистолькорсом была предложена схема (рис. 6.3), носящая его имя.

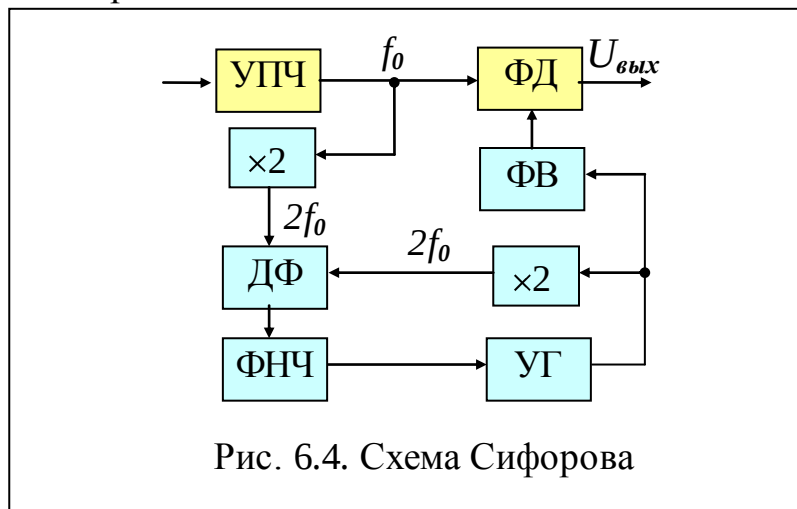


Работа схемы происходит на промежуточной частоте f_0 сигнала. С помощью умножителя частоты на два ($\times 2$) колебание на его выходе не

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

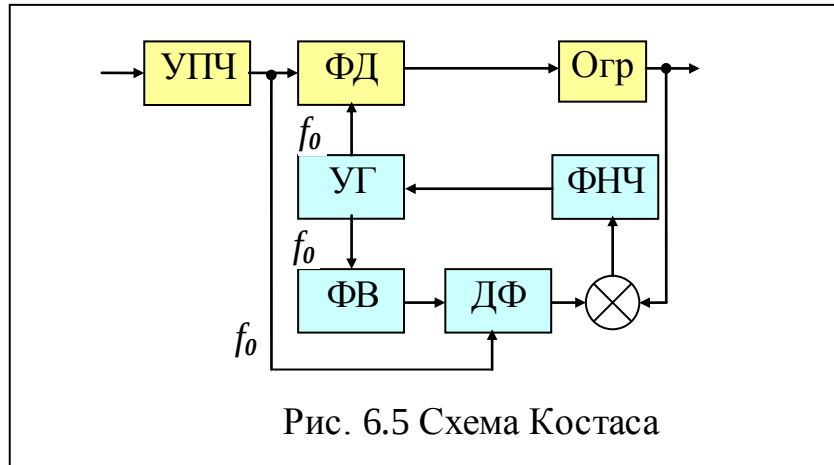
Однако, в зависимости от начальных условий, фаза сформированного опорного колебания может принимать два значения относительно фазы входного сигнала - "0" и " π ", что приводит к обратной работе ФД. Кроме того, интенсивные помехи также могут вызывать перескок фазы опорного напряжения на π . Наличие делителя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

В 1937 году советским ученым В.И. Сифоровым была предложена схема, приведенная на рис. 6.4.



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

Предложенная в 1957 году американским ученым Д. Костасом схема формирования опорного напряжения показана на рис. 6.5



Все элементы схемы работают на частоте f_0 , а фазовая манипуляция снимается умножением протектированного сигнала на входной в умножителе, выход которого после низкочастотной фильтрации управляет частотой генератора УГ. Эта схема проще в реализации, но, как и предыдущие, обладает склонностью к обратной работе. Выход состоит в использовании относительной фазовой манипуляции - ОФМ.

6.3 Относительная бинарная фазовая манипуляция сигналов

На приемной стороне для демодуляции сообщения используется фазовый детектор (ФД), в котором реализуется умножение принятого сигнала (4.3) на опорный сигнал $s_{\text{оп}} t = \cos \omega_0 t + \varphi_{\text{оп}}$ и последующее интегрирование на интервале T_b длительности информационного символа. После выполнения указанных операций происходит выделение постоянной составляющей U_b на интервале T_b

$$u_{\text{фд } i} = b_i \cdot U_b \cos \varphi_{\text{оп}} - \varphi_0 \quad (6.5)$$

Отсюда получаем значение информационного символа

$$b_i = u_{\text{фд } i} / U_b \cdot \cos \varphi_{\text{оп}} - \varphi_0 \quad (6.6)$$

Если фаза опорного сигнала $\varphi_{\text{оп}}$ неизменна на всем протяжении сеанса приема, то будет выделена однозначная последовательность информационных символов b_i . Но, если по какой-либо причине произойдет скачок фазы опорного сигнала на $\pm \pi$, то знак выделяемых символов b_i изменится на обратный. Иными словами, потеря синхронизма

фазы приводит к совершенно недопустимому явлению, так как последовательность информационных символов будет приниматься инверсно до следующего скачка.

От описанного недостатка свободен метод *относительной фазовой модуляции* (ОФМ). Суть метода заключается в том, что знак фазы каждого информационного символа определяется не относительно начальной для данного сеанса приема фазы, а относительно фазы предыдущего символа. Пусть φ_i - фаза текущего символа, а φ_{i-1} - фаза предыдущего символа.

Так как для ФМ сигнала $\varphi = 0$ или π , то разность фаз $\Delta\varphi$ между двумя соседними символами может также принимать только два значения

$$\Delta\varphi = |\varphi_i - \varphi_{i-1}| = \begin{cases} 0 & \text{если } \varphi_i = \varphi_{i-1}, \\ \pi & \text{если } \varphi_i = \varphi_{i-1} + \pi. \end{cases} \quad (6.7)$$

Из выражения (6.7) видно, что если скачок фазы произойдет точно на границе двух символов, то неправильно будет принят лишь один символ. Таким образом, перескок фазы опорного сигнала на $\pm \pi$ при приеме

ОФМ-сигнала приводит к появлению только локальных ошибок, а не к переходу на обратную работу в течение длительного времени, как это имеет место при приеме ФМ сигнала.

Относительную фазовую манипуляцию можно рассматривать как дополнительную перекодировку (относительное кодирование) последовательности бинарных информационных символов сообщения $b_i = -1, 1$. Пусть $\hat{b}_i$ - последовательность символов, полученная после перекодирования. Тогда в соответствии с принципом формирования ОФМ совпадение двух символов $\hat{b}_i$ и $\hat{b}_{i-1}$ означает передачу исходного символа $b_i = 1$, а в противном случае, когда символы $\hat{b}_i$ и $\hat{b}_{i-1}$ различны, означает передачу символа $b_i = -1$. Это правило можно записать в виде следующих операций

$$\hat{b}_i = b_i \times \hat{b}_{i-1} \quad (6.8, \text{ а})$$

$$b_i = \hat{b}_{i-1} \times \hat{b}_i; \quad (6.8, \text{ б})$$

Алгоритм (6.8,а) используется для относительного кодирования информационного сообщения. Обратный переход от относительного кодирования к исходному потоку производится по алгоритму (6.8, б).

Если рассматривать дополнительную перекодировку исходной последовательности двоичных информационных символов $d_i = 0,1$, $d_i = 1 - b_i / 2$, то правило относительного кодирования и декодирования информационных символов $d_i = 0,1$ можно записать в виде логических операций

$$\hat{d}_i = d_i \oplus \hat{d}_{i-1} \quad (6.9, \text{ а})$$

$$\hat{d}_i = d_i \oplus \hat{d}_{i-1}; \quad (6.9, \text{ б})$$

Тогда структуру кодера и декодера информационных символов d_i можно представить в виде цифровой схемы, представленной на рис. 6.6.а) и б) соответственно.

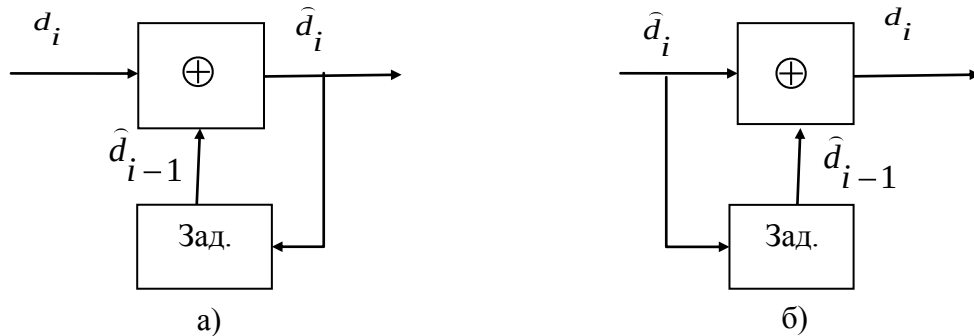


Рис. 6.6. Схема кодера и декодера информационных символов в относительном коде

Для получения радиосигнала с ОФМ достаточно умножить полученный (перекодированный) сигнал $\hat{s} \cdot t$ на несущее опорное колебание $s_{\text{оп}} \cdot t$. Схема демодулятора сигналов с ОФМ содержит фазовый демодулятор с генератором опорного колебания $s_{\text{оп}} \cdot t$.

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

6.4 Модуляция цифровой информации кодовой последовательностью

Пусть $s_k \cdot t$ - дискретный фазоманипулированный сигнал, манипулированный некоторой специфической бинарной кодовой последовательностью $b = b_i = \pm 1, (i-1)T_b < t \leq iT_b$. Положим, что на

интервале одного бита данных содержится N_s чипов дискретного сигнала. Тогда модуляция сигнала кодовой последовательностью осуществляется по правилу

$$s_k(t; \mathbf{b}_k) = s_k(t) \cdot B_k(t) \cdot \cos(2\pi f_0 t). \quad (6.10)$$

Обратимся к диаграммам, изображенным на рис. 6.7.

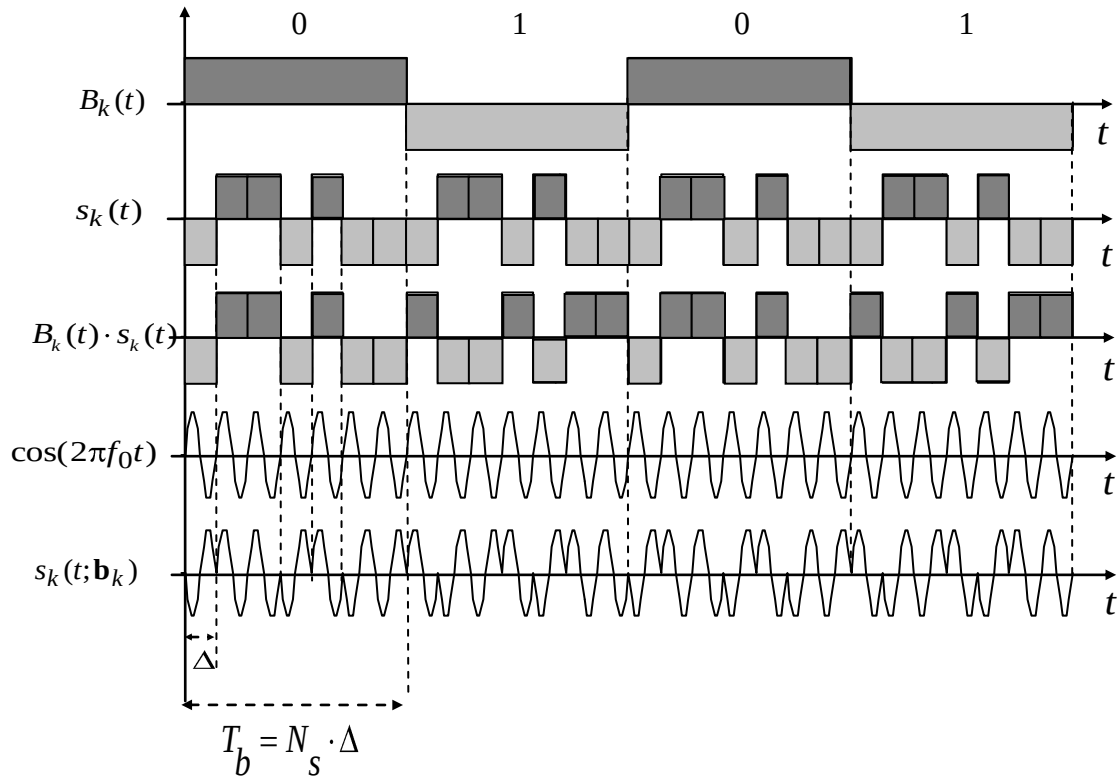


Рис. 6.7. Передача данных с помощью кодовой модуляции

Здесь битовый поток $B_k(t)$ первоначально умножается на дискретный кодовый сигнал $s_k(t)$, представляющую собой некоторую бинарную последовательность длины $N_s = 7$, $T_b = N_s \cdot \Delta$. Результат произведения $B_k(t) \cdot s_k(t)$ модулирует непрерывную несущую по фазе. Можно сказать, что в данном случае битовый поток первоначально модулирует кодовую последовательность, а результат используется для бинарной фазовой манипуляции несущей частоты сигнала.

Поскольку ширина полосы сигналов (6.8) есть величина обратная длительности чипа $\Delta = T_b / N_s$, то дополнительная модуляция прямой последовательностью $s_k(t)$ расширяет спектр в N_s раз.

После распространения по каналу, в ходе которого приобретает

задержка τ_k и начальная фаза φ_k , сигнал принимает вид

$$s_{kr}(t); \mathbf{b}_k = s_k(t - \tau_k) \cdot B_k(t - \tau_k) \cdot \cos(2\pi f_0 t + \varphi_k) \dots \quad (6.11)$$

Предполагая, как и ранее, полное знание о параметрах τ_k , φ_k , в приемнике для восстановления текущего i -го бита может быть вычислена корреляция

$$z_k = \int_{(i-1) \cdot T_b + \tau_k}^{i \cdot T_b + \tau_k} y(t) \cdot s_k(t - \tau_k) \cdot \cos(2\pi f_0 t + \varphi_k) dt \quad (6.12)$$

между наблюдаемым колебанием $y(t)$ и формируемой на месте полосной копией сигнала $s_k(t - \tau_k) \cdot \cos(2\pi f_0 t + \varphi_k)$.

Для вынесения решения принят, бит $\hat{b}_{k,i} = 0$ или $\hat{b}_{k,i} = 1$, как и ранее, используется знак корреляции: $\hat{b}_{k,i} = \text{sign}(z_k)$.

Следует отметить, что та же самая оптимальная процедура может быть реализована в два этапа: на первом снимается расширение спектра, а на втором осуществляется демодуляция данных точно так, как и в случае отсутствия расширения.

Пусть наблюдение $y(t)$ умножается на формируемую в приемнике низкочастотную копию сигнала $s_k(t - \tau_k)$, в точности синхронизированную с принимаемым сигналом. Полезная составляющая наблюдения после этой операции примет вид

$$\begin{aligned} s_{kr}(t); \mathbf{b}_k \cdot s_k(t - \tau_k) &= s_k^2(t - \tau_k) \cdot B_k(t - \tau_k) \cdot \cos(2\pi f_0 t + \varphi_k) = \\ &= B_k(t - \tau_k) \cdot \cos(2\pi f_0 t + \varphi_k), \end{aligned} \quad (6.13)$$

где учтена бинарная природа сигнала $s_k(t) = \pm 1$, вследствие чего $s_k^2(t) = 1$.

Рис. 6.8 иллюстрирует процедуру трансформации широкополосного сигнала с прямым расширением в обычный бинарный ФМ сигнал, модулированный потоком данных.

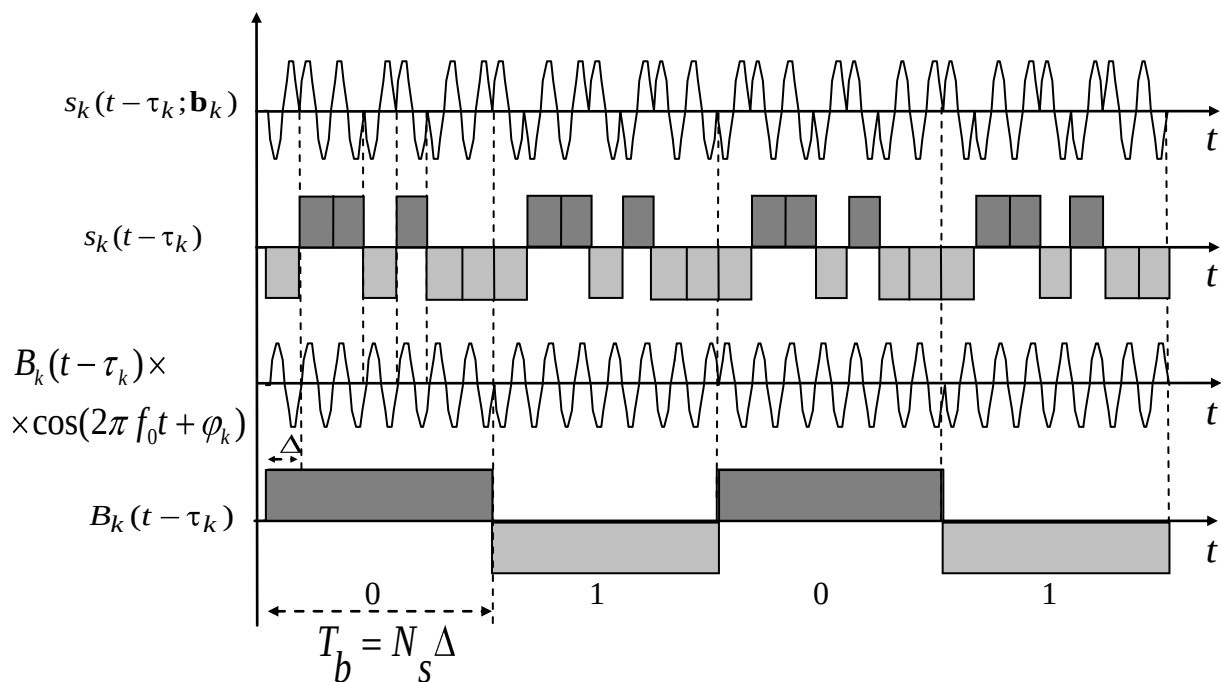


Рис. 6.8. Передача данных с помощью бинарной модуляции со снятием расширения спектра

Как видно, после данного шага принятый сигнал не обладает никакими чертами широкополосного, полностью совпадая с простым сигналом с бинарной манипуляцией, определяемой потоком данных. Вследствие этого процедура умножения наблюдения на копию сигнала называется *снятием расширения*.

На рис. 6.9 приведены схемы устройств расширения и сужения спектров сигналов при кодовой модуляции.

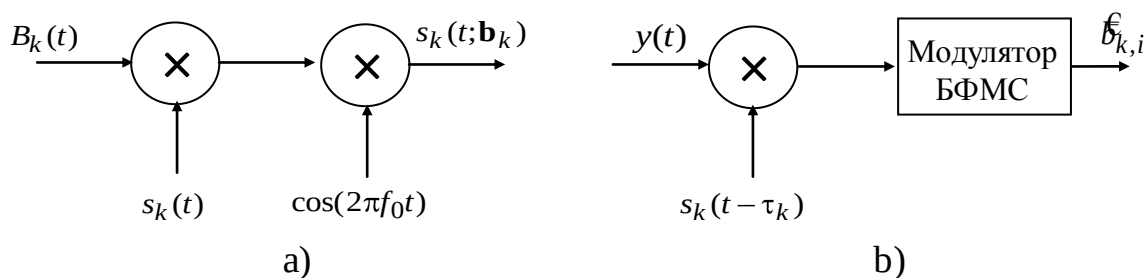


Рис. 6.9. Расширение (a) и сужение (b) сигнала с бинарной ФМ.

Поскольку сигнал со снятым расширением представляет собой непрерывную несущую с бинарной фазовой манипуляцией данными, то дальнейшее извлечение информации осуществляется привычным демодулятором бинарного ФМ сигнала

Для проведения обсуждения в терминах частотной области обратимся к рис. 6.10, на котором приведены спектральные плотности мощности $\tilde{S}_b(f)$, $\tilde{S}_{bs}(f)$ исходного потока данных $B_k t$ и его широкополосной

версии $Bs_k(t) = s_k(t) \cdot B_k(t)$ соответственно. Для последовательности $B_k(t)$ битовых импульсов длительности T_b , полярности которых случайны и независимы, спектр мощности определяется выражением

$$\tilde{S}_b(f) = T_b \cdot \text{sinc}^2(\pi \cdot f \cdot T_b) \quad (6.14)$$

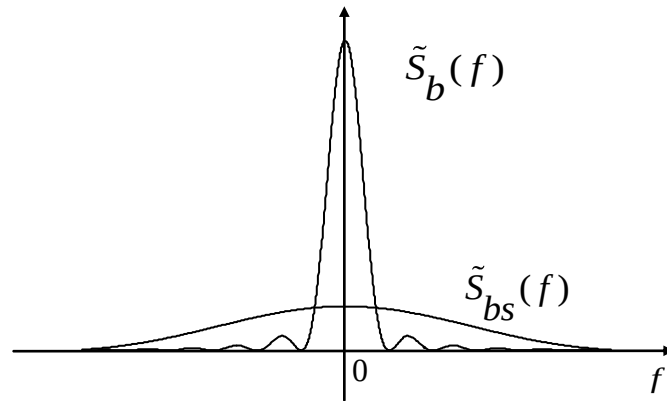


Рис. 6.10. Спектральные плотности мощности сигналов без модуляции и с модуляцией кодовой последовательностью

Рассмотрение расширенного потока данных снова как случайную последовательность импульсов с независимыми полярностями (на этот раз длительности Δ) приводит к спектру мощности той же формы, но занимающему в N_s раз большую полосу:

$$\tilde{S}_{bs}(f) = \Delta \cdot \text{sinc}^2(\pi \cdot f \cdot \Delta) = T_b / N_s \cdot \text{sinc}^2(\pi \cdot f \cdot T_b / N_s). \quad (6.15)$$

Передаваемый информационный сигнал «расплылся» по спектру кодового сигнала $s_k(t)$.

На приемной стороне снятие расширения возвращает спектр в исходную полосу, превращая сигнал в узкополосный, что позволяет применять известные методы демодуляции данных.

Следовательно, если принятый и демодулированный широкополосный сигнал перемножить с точной копией кодового сигнала $s_k(t)$, который использовался для модуляции, то получим узкополосную составляющую с очень высокой энергией на единицу частоты - переданный поток данных. Таким образом, используя различные кодовые последовательности (коды), можно организовать передачу нескольких информационных сообщений в одной и той же полосе частот.

6.5 Помехозащищённость широкополосных систем

Задача этого раздела - показать, почему и как широкополосная технология может помочь в борьбе с помехами. По этой причине мы ограничимся лишь простейшими оценками, основанными на отношении мощностей сигнала и суммы шума и помехи.

6.6.1 Узкополосная помеха

Данный тип помех обусловлен работой соседствующих РТС в перекрывающемся диапазоне частот, то есть помехами искусственного происхождения. На рис. 6.11 в виде прямоугольников представлены амплитудный спектр $|\tilde{S}(f)|$ полезного сигнала и спектральная плотность мощности помехи $\tilde{J}(f)$ на фоне равномерной спектральной плотности мощности $N_0/2$ АБГШ.

Назовем помеху *узкополосной* только по той причине, что занимаемая ей полоса W_j меньше полосы w , занимаемой сигналом, и имеются области, где спектр сигнала не подвержен искажению помехой.

Первоначально предположим, что рассматриваемая система не предпринимает никаких специальных мер для противодействия помехи за исключением только соответствующего выбора сигнала. Другими словами, всегда используется только фильтр, согласованный с АБГШ, не взирая на наличие или отсутствие помехи на входе приемного устройства.

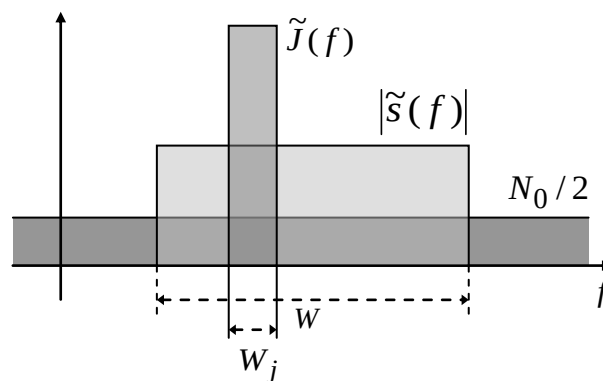


Рис. 6.11. Спектр сигнала, помехи и фонового АБГШ.

При определении отношения q_I^2 мощностей сигнала и мешающего воздействия на выходе согласованного фильтра отметим тот факт, что при прямоугольном спектре сигнала амплитудная передаточная функция фильтра равномерна в полосе сигнала w и равна нулю вне ее. Положим, что ее ненулевое значение равняется единице. Поэтому помеха проходит на выход фильтра без изменения своей мощности J , тогда как

отфильтрованное значение мощности шума составит $N_0 W$. С другой стороны, фильтр согласован с сигналом и, значит, когерентно суммирует все гармонические составляющие сигнала, обеспечивая максимальное значение амплитуды на выходе $A_{\text{out}} = \int_{-\infty}^{\infty} |\tilde{s}(f)| df = 2W\tilde{s}$, где учтена равномерность спектра в полосе сигнала w , а удвоение обусловлено вкладом «отрицательных» частот. При аналогичных обозначениях энергия сигнала вычисляется с помощью теоремы Парсеваля как $E = \int_{-\infty}^{\infty} |\tilde{s}(f)|^2 df = 2W\tilde{s}^2$. В итоге имеем

$$q_I^2 = \frac{A_{\text{out}}^2}{J + N_0 W} = \frac{4W^2 \tilde{s}^2}{J + N_0 W} = \frac{2E}{N_0 + J/W}. \quad (6.16)$$

Из последнего равенства можно увидеть, что, несмотря на конкретное значение полосы помехи W_j , отношение сигнал/ помеха+шум (SIR) на выходе согласованного фильтра ведет себя так, как если бы мощность помехи была равномерно распределена в полосе w сигнала, образуя дополнительный «АБГШ» со спектральной плотностью J/W .

Рассмотрим теперь другой вариант, когда данная система приспособливает свой приемник к действующему образцу помехи. Оптимальной процедурой обработки в этом случае следовало бы считать фильтрацию, согласованную с полным мешающим воздействием, включая и узкополосную помеху. Практически очевидно, что, если помеха значительно превосходит фоновый АБГШ, то подобная обработка эквивалентна вырезанию частотного интервала, в котором сосредоточена помеха. Рис. 6.12 иллюстрирует вид спектра на выходе режекторного фильтра. Наблюдается полное отсутствие помех, однако, вырезаются и частотные компоненты сигнала в пределах полосы помехи, точно также как и шум. Модель спектра может трактоваться, как образец, в котором исходный сигнал занимал бы только часть полосы w , свободную от помех, и обладал энергией, равной $E(1 - W_j/W)$. Тогда согласованный фильтр, очищая этот остаточный сигнал от АБГШ, обеспечивал бы выходное отношение мощностей сигнала и шума в виде (индекс «J» отвечает помехе)

$$q_J^2 = \frac{2E(1 - W_j/W)}{N_0} = q^2(1 - W_j/W), \quad (6.17)$$

где $q^2 = 2E / N_0$ - «истинное» отношение мощностей сигнала и шума (SNR) на выходе согласованного фильтра в отсутствии помехи.

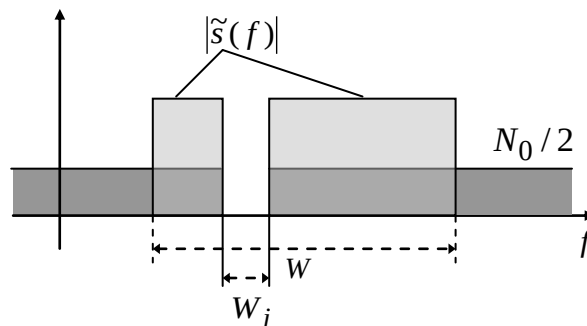


Рис. 6.12. Модель спектра на выходе режекторного фильтра.

Из анализа соотношений (6.16) и (6.17), можно заметить, что они оба явным образом свидетельствуют о достоинствах широкополосных сигналов с точки зрения противостояния воздействию помех. Действительно, чем шире полоса сигнала w в сравнении с полосой помехи w_j , тем меньше дополнительная спектральная плотность в первом случае и энергетические потери - во втором (при постоянстве мощности помехи J), и, следовательно, больше q_I^2 и q_J^2 .

В результате приходим к следующему выводу: достижение высокой помехоустойчивости по отношению к узкополосной помехе, не прибегая к увеличению энергии сигнала или пиковой мощности, возможно только расширяя спектр сигнала независимо от его длительности, т.е. *при использовании широкополосной технологии*.

6.6.2 Заградительная помеха

Во многих военных мероприятиях часто встречается ситуация, когда помеха создается преднамеренно как средство радиоэлектронного противодействия. В подобных ситуациях постановщик помехи может предполагать, что подавляемая система окажется способной зарегистрировать присутствие помехи и соответственным образом отреагировать на ее воздействие. В частности, при использовании узкополосной помехи система может прибегнуть к режекторной фильтрации или даже изменить сигнал путем переноса его спектра в зону свободную от помех. Для предотвращения любой такой возможности может быть применена т.н. заградительная шумовая помеха, спектр которой без промежутков покрывает спектр сигнала (см. рис. 6.13).

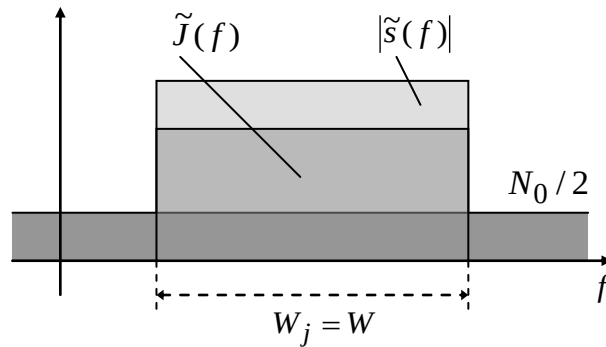


Рис 6.13. Спектр сигнала, заградительной помехи и АБГШ.

Очевидно, что заградительная помеха воздействует на сигнал как дополнительный АБГШ со спектральной плотностью мощности, равной $N_J = J / W$. Поэтому отношение SNR по мощности на выходе согласованного фильтра полезной системы будет определяться как

$$q_J^2 = \frac{2E}{N_0 + N_J} = \frac{2E}{N_0 + J / W},$$

что совпадает с отношением SIR в (6.16).

В этом случае, однако, враждебные намерения вынуждают постановщика помех обеспечить значительно больший эффект подавления по сравнению с белым шумом. Это возможно, если $J / W \gg N_0$, откуда следует, что

$$q_J^2 \approx \frac{2EW}{J} = \frac{2P(WT)}{J}. \quad (6.18)$$

Из (6.18) следует, что при наложении ограничений, как на пиковую мощность полезного сигнала, так и на мощностной ресурс постановщика помех, *единственным способом повышения помехоустойчивости системы к заградительной помехи является привлечение сигналов с большим значением частотно-временного произведения WT , т.е. широкополосных сигналов.*

Формула (6.18) объясняет другое популярное название частотно-временного произведения WT . Как следует из формулы, отношение мощностей сигнала и равномерного шума в пределах полосы сигнала увеличивается на выходе согласованного фильтра в $2WT$ раз по сравнению со входным значением P / J . Таким образом, естественно назвать величину WT *выигрышем от обработки*.

Отметим, что целью данного рассмотрения не являлось получение ответа на следующие вопросы: какого рода помеха наиболее опасна в конкретной ситуации, и что необходимо предпринять системе для оптимизации своих

характеристик в более сложных условиях, чем это было рассмотрено. Основная идея состояла только в демонстрации основных достоинств распределенности спектра в борьбе с помехами.

6.6.3 Энергетическая скрытность

При радиоэлектронном противодействии эффективная помеха может быть организована только после обнаружения сигналов и оценки таких ее параметров как частотный диапазон и занимаемая полоса. Указанные условия предполагают следующий широко распространенный сценарий конфронтации двух систем, при котором первая (назовем ее *скрытной*) старается действовать по возможности скрытно и избежать обнаружения своего сигнала, тогда как вторая (*перехват перехватчик*) постоянно находится в готовности, предпринимая все возможное для обнаружения первой, лишь иногда находящейся в активном режиме. Выступая на стороне первой системы, исследуем возможности от применения расширенного спектра по противодействию усилиям системы-перехватчика.

Как и ранее, основной целью является понимание общей идеи положительной роли расширения спектра в этом противостоянии систем. Предположим, что скрытная система использует сигнал с некоторым не тривиальным законом модуляции, параметры которого неизвестны перехватчику, лишая последнего возможности применения согласованного фильтра или коррелятора для обнаружения сигнала. В этих условиях естественно полагать, что у противостоящей системы нет иного выбора, как рассматривать перехватываемый сигнал в виде случайного и основывать его обнаружение только на факте появления или отсутствия некоторого избытка энергии в подозрительном участке частотного диапазона. Таким образом, *энергетический детектор*, называемый также *радиометром*, который является оптимальным с точки зрения обнаружения ограниченного по полосе шумового сигнала на фоне АБГШ, принимается в качестве рабочего инструмента на стороне перехвата.

В реальной практике перехватчику могут быть неизвестны заранее сведения о частотном диапазоне и интервале времени, занимаемом сигналом. Учитывая эти обстоятельства, его стратегия будет заключаться в комбинировании указанных параметров, осуществляя процедуру обнаружения либо путем сканирования частотно-временной области, либо используя набор параллельных каналов, каждый из которых ответственен за анализ ограниченного участка частотно-временной области. В любом случае качество работы приемника системы-перехватчика будет полностью определяться характеристикой энергетического детектора, настроенного на истинную для перехватываемого сигнала частотно-временную зону.

На рис. 6.14 представлена прямоугольная аппроксимация спектра сигнала вместе с равномерной спектральной плотностью мощности естественного АБГШ (a) и амплитудно-частотная передаточная функция полосового фильтра радиометра (b).

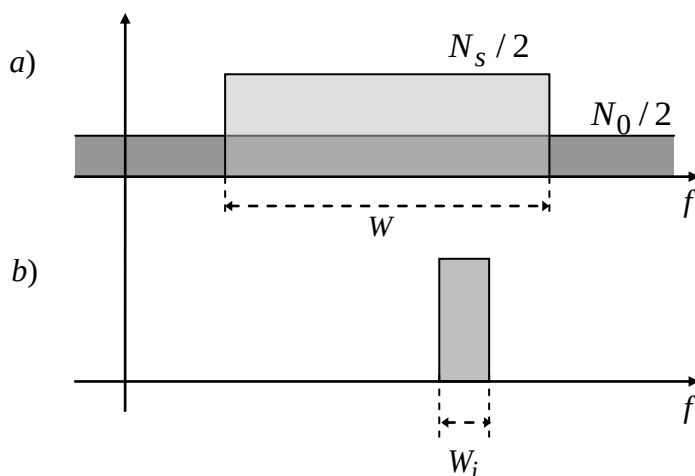


Рис 6.14. Спектр наблюдения на входе радиометра и ширина полосы полосового фильтра.

С точки зрения перехватчика указанием на присутствие сигнала служит избыточная (сигнальная) спектральная плотность мощности $N_s/2 = P/2W$, дополнительная к спектральной плотности мощности теплового шума $N_0/2$.

В отсутствии сигнала случайный процесс на выходе полосового фильтра радиометра характеризуется мощностью $\sigma_n^2 = N_0 W_i$, а при его наличии - $\sigma_n^2 + \sigma_s^2 = (N_s + N_0) W_i = (P/W + N_0) W_i$.

Эти соотношения позволяют понять, что с точки зрения перехватчика максимально возможная полоса фильтра, т.е. равная полосе сигнала ($W_i = W$), является оптимальной, обеспечивая наибольшее отношение сигнал-шум

$$q_i = \frac{P\sqrt{T}}{\sqrt{W}N_0} = \frac{q^2}{2\sqrt{WT}}, \quad (6.19)$$

где $q^2 = 2E/N_0 = 2PT/N_0$, как и ранее, отношение сигнал-шум на выходе согласованного фильтра приемника скрытной системы. Очевидно, что значение q^2 должно быть достаточно большим, поскольку в противном случае скрытная система не сможет нормально функционировать. Совершенно ясно также, что у скрытной системы имеется только единственная возможность предотвратить обнаружение своего сигнала

потенциальным перехватчиком: *использовать сигнала с распределенным спектром, обладающие максимально возможным значением выигрыша от обработки WT*.

6.6.4 Структурная скрытность сигнала

Продолжая линию предыдущего параграфа, напомним еще раз, что единственной причиной, вынуждающей перехватчик прибегнуть к такому неэффективному инструменту как энергетический приемник, является отсутствие информации о структуре обнаруживаемого сигнала, т.е. его закона модуляции. По этой причине перехватчик не может обрабатывать сигнал аналогично приемнику скрытной системы (т.е. осуществлять согласованную фильтрацию). Очевидно, что в случае недостаточной структурной сложности сигнала и осведомленности перехватчика о его возможных альтернативных вариантах, последний может попытаться их все реализовать. Соответствующим оборудованием для этого может служить набор параллельных согласованных фильтров либо единый перестраиваемый фильтр (несколько фильтров), пригодный для обработки сигналов различных по структуре последовательно во времени, если сигнал, который необходимо обнаружить, принимается достаточно долго. Поэтому другая сторона стратегии скрытной системы в борьбе с перехватчиком состоит в применении сигналов с практически не раскрываемой структурой. Подобная задача характерна для военных и коммерческих систем, которые не стремятся полностью скрыть факт своего присутствия в эфире.

Заклучим параграф еще одним заявлением о достоинствах широкополосности: *данная технология в значительной степени способствует криптографической защите структуры сигнала*.

6.6.5 Электромагнитная совместимость

Проблема электромагнитной совместимости (ЭМС) является одной из наиболее приоритетных в современной беспроводной технике. ЭМС подразумевает бесконфликтное сосуществование различных систем в эфире, несмотря на то, что каждая из них принимает не только свой собственный сигнал, но и сигналы остальных систем. Очевидно, что невозможно полностью исключить взаимное влияние, когда несколько систем одновременно функционируют в относительно малой области. Любая активная, т.е. излучающая электромагнитную волну, система неизбежно окажет воздействие на все соседние, и в задачу разработчика систем входит сделать это потенциально вредное воздействие терпимым.

В задаче ЭМС принимают участие две стороны. Первая из них, которую можно назвать «излучающей», старается минимизировать свое

вмешательство, обусловленное излучаемой мощностью, на другую соседнюю, называемую «принимающей» системой. Мотивацией подобного поведения являются строгие международные и национальные правила, за исполнением которых тщательно следят службы, уполномоченные налагать соответствующие санкции. Любая из систем второй, «принимающей» стороны также предпринимает собственные меры, нацеленные на нейтрализацию чужих сигналов, попадающих в ее приемник. К числу традиционных способов обеспечения ЭМС относятся строгое частотное планирование, контроль за которым осуществляют национальные и международные институты, использование антенн с узкой направленностью, тщательная разработка высокочастотных узлов и др.

В данном параграфе кратко рассмотрим вопрос, почему технология распределенного спектра также может быть отнесена к приведенному перечню мероприятий. Со стороны излучающей системы оправданной является следующая логика поведения. Сделать излучаемый сигнал, насколько это возможно, почти незаметным для специальных приемников радиоперехвата, путем усложнения закона модуляции, т.е. расширением спектра, в результате чего подобный сигнал почти не будет оказывать вредного влияния на обычную стороннюю систему, работающую в той же полосе частот. Задача состоит только в выборе такого выигрыша от обработки, который гарантировал бы достаточно низкий уровень спектральной плотности мощности используемого сигнала относительно спектральной интенсивности естественного шума на входе приемника сторонней системы. Как правило полагают, что «достаточно низким» считается уровень в -7 дБ, т.е. $N_s / N_0 \leq 0.2$. Подстановка

$N_s = P / W = E / WT$ в последнее соотношение приводит к критерию ЭМС $E / WTN_0 \leq 0.2$ или $q^2 / WT \leq 0.4$, где снова целевой параметр определен в

терминах отношения сигнал-шум скрытного приемника q^2 и выигрыша от обработки WT . Если, например, отношение сигнал-шум скрытной системы на входе стороннего приемника было бы 20 дБ, то величину $WT \geq 400$ можно было бы считать удовлетворительной в смысле ЭМС. На практике подобные оценки должны координироваться с дальностью, определяющей радиус окружности вокруг излучающей системы, вне которой сигнал от излучателя считается практически безвредным для других систем.

Со стороны принимающей системы любой сторонний сигнал на входе приемника может трактоваться как узкополосная или широкополосная помеха, и все доводы о пользе широкополосности в борьбе с помехами снова применимы. Следовательно, *широкополосная модуляция хорошо ложится в русло обеспечения ЭМС.*

7 ЭФФЕКТЫ РАСПРОСТРАНЕНИЯ СИГНАЛОВ В КАНАЛАХ СВЯЗИ

Исследование распространения волнового колебания достаточно сложно и с трудом поддается теоретическому анализу. Имеет место большое разнообразие факторов, вызывающих как детерминированное, так и случайное ослабление сигнала, достигающего входа приемного устройства. Принимаемый сигнал искажается не только аддитивным шумом приемника, но и *мультипликативной* помехой, которая изменяет интенсивность сигнала, т.е. перемножается с амплитудой сигнала.

7.1 Распространение в свободном пространстве

Начнем рассмотрение с модели распространения сигналов в свободном пространстве, где отсутствуют препятствия между передающей и приемной антеннами, и излученная волна распространяется по единственно возможному пути, называемому *линией прямой видимости*.

Обозначим расстояние между передатчиком и приемником как D . Если передающая антенна является всенаправленной, то излученная ее мощность P_t будет равномерно распределена по внутренней поверхности сферы радиуса D , причем на каждую единицу площади будет приходиться ее часть, равная $P_t / 4\pi D^2$. Тогда на приемную антенну с эффективной площадью раскрытия A_r попадает мощность, равная $P_r = A_r \cdot P_t / 4\pi D^2$.

Если же передающая антенна является направленной, то в сторону приемной антенны излучается мощность, которая в G_t раз превосходит всенаправленную. Величина G_t называется *коэффициентом усиления мощности* передающей антенны. В этом случае принятая мощность также увеличивается в G_t раз. Для представления принятой мощности воспользуемся зависимостью между площадью раскрытия антенны A_r и коэффициентом усиления мощности принимающей антенны $G_r = 4\pi A_r / \lambda_w^2$, где λ_w - длина волны. В результате приходим к формуле Фриза для свободного пространства

$$P_r = P_t G_t G_r \left(\frac{\lambda_w}{4\pi D} \right)^2. \quad (7.1)$$

Приведенное выражение показывает, что ослабление мощности сигнала в свободном пространстве по линии прямой видимости обратно

пропорционально квадрату расстояния.

Среда распространения наземных систем менее благоприятна и основное влияние на интенсивность сигнала оказывают два следующих компонента: *затенение* и *замирание (фединг) вследствие многолучевого распространения сигналов*.

7.2 Затенение

Затенение обусловлено деталями ландшафта, препятствующими распространению сигналов по прямой линии: возвышенностями, лесами, зданиями и т.п. Вследствие их влияния интенсивность сигнала падает.

Специалистами в области связи признана модель затенения Окумура-Хаты. Согласно этой модели зависимость средней принятой мощности $\bar{P}_r$ подчиняется соотношению

$$\bar{P}_r = k P_t / D^e, \quad (7.2)$$

в котором коэффициент k - определяется частотным диапазоном и высотой антенн, а значение показателя e зависит от типа подстилающей поверхности, изменяясь от 3 (для сельской местности) до 5 (для плотной городской застройки).

Уменьшение мощности, обусловленное затенением, носит статический характер. Флуктуации мощности P_r по дуге радиуса D с центром в месте расположения передатчика хорошо описываются логнормальным законом распределения. Когда приемник находится в движении, изменение мощности во времени, происходит сравнительно медленно.

7.3 Замирания вследствие многолучевого распространения

Обратимся теперь ко второму фактору, влияющему на интенсивность принимаемого сигнала: *многолучевое распространение*. Фактически излученный сигнал может достичь приемной антенны разнообразными путями. Другие пути образуются в результате отражения излученной волны различными объектами. Путь прямого распространения может оказаться либо одним из них, либо отсутствовать при крайне сложном распространении сигналов.

Предположим, что, распространяясь по i -му пути, излученный сигнал с комплексной огибающей вида $\dot{S}(t)$ приобретает амплитуду A_i , задержку τ_i и начальную фазу φ_i . Тогда комплексная амплитуда

принятого сигнала может быть записана в виде

$$\dot{S}_r(t) = \sum_i A_i \dot{S}(t - \tau_i) \exp(j\varphi_i). \quad (7.3)$$

В том случае, когда максимальное значение взаимной задержки τ_{ds} между сигналами различных путей, не превосходит длительности сигнала, все многолучевые сигналы перекрываются и интерферируют между собой.

Суперпозиция независимых и примерно равных по вкладу случайных слагаемых стремится к гауссовскому распределению по мере роста их числа. Следовательно, множество распространяющихся по разным путям сигналов, образуют на входе приемника полосный гауссовский процесс. Если среди входных компонентов отсутствует доминирующая детерминированная составляющая (подобная сигналу, распространяющемуся по прямой линии), результирующий гауссовский процесс будет иметь нулевое среднее, а огибающая данного процесса подчиняется релеевскому распределению. Таким образом, приходим к модели канала с релеевскими замираниями.

В результате принятая амплитуда A_r является не детерминированной, а случайной, удовлетворяющей релеевскому закону плотности распределения

$$W(A_r) = \begin{cases} 2A_r \exp(-A_r^2), & A_r \geq 0, \\ 0, & A_r < 0. \end{cases} \quad (7.4)$$

Релеевские замирания характеризуются глубокими провалами в интенсивности сигнала. В результате эффект релеевского замирания носит достаточно вредный характер.

Рассмотрим в качестве примера передачу бинарных данных по релеевскому каналу с *медленными* и *амплитудными* замираниями. Первое определение означает, что помеховый образец остается стабильным в течение многих символов. Второе определение подчеркивает, что задержка распространения многолучевых сигналов τ_{ds} достаточно мала в сравнении с длительностью T_b индивидуального бинарного ФМ символа: $\tau_{ds} \ll T_b$. В результате последовательные бинарные ФМ символы не перекрываются друг с другом, т.е. отсутствует *межсимвольная интерференция*.

Диаграммы для случая колоколообразных символьных импульсов, приведенные на рис. 7.1 демонстрируют медленные амплитудные замирания (а) в противопоставлении быстрым амплитудным замираниям

(b). Второй тип замираний характеризуется быстрым изменением во времени интерференционного образца, так что искажения последовательных символов становятся практически независимыми.

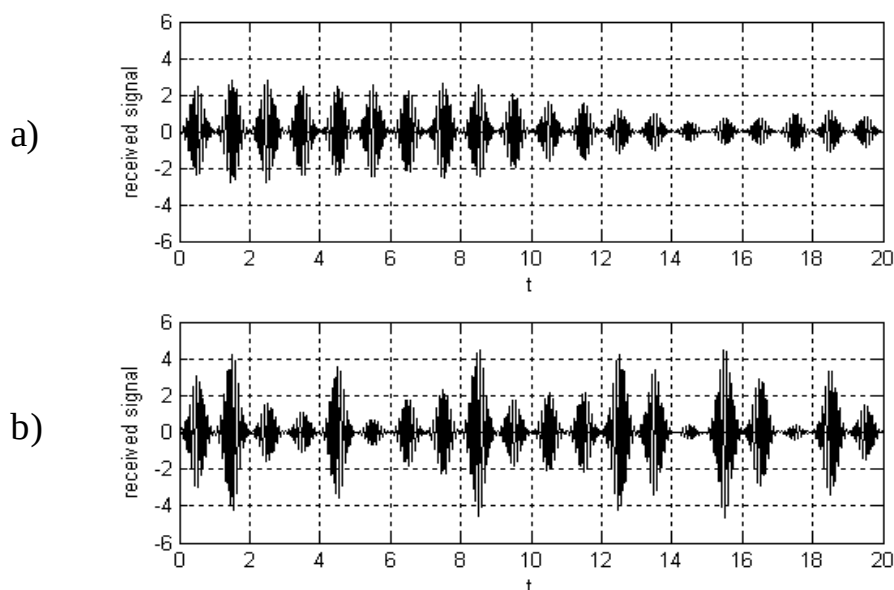


Рис. 7.1. Медленные (a) и быстрые (b) замирания.

Адекватной моделью многолучевого канала может служить линия задержки с отводами, отвечающими значениям τ_i , и весовыми коэффициентами $A_i \exp(j\varphi_i)$. Передаточная функция такой системы сильно зависит от задержек отводов и при $\tau_{ds} \ll T_b$ достаточно равномерна в полосе сигнала, так что все частотные компоненты сигнала искажаются одинаково, а форма сигнала остается неизменной.

В случае, когда замирания оказываются *частотно-селективными*, передаточная функция канала не будет равномерной в полосе сигнала. Подобная ситуация имеет место тогда, когда задержка распространения охватывает несколько излученных бит цифрового сигнала. В этом случае на выходе канала предшествующий бит перекрывается с текущим битом. Для противодействия данной межсимвольной интерференции применяются специальные фильтры (эквалайзеры), которые выравнивают неравномерную передаточную функцию канала. При соответствующем использовании частотная селективность может служить дополнительным источником борьбы с замираниями.

Далее остановимся на методах борьбы с негативными последствиями многолучевости распространения сигналов.

7.4 Разнесение

Основная идея борьбы с эффектами многолучевости заключается в использовании *разнесения*, которое означает организацию нескольких независимых каналов передачи или ветвей. Вследствие этого, несмотря на то, что каждая из ветвей остается подверженной релеевским замираниям, вероятность того, что интерференционные (помеховые) сигналы во всех из них будут одновременно неудачными, не может оказаться значительной. С увеличением числа ветвей выигрыш от разнесения становится более существенным. На приемной стороне могут быть применены различные варианты комбинирования результатов обработки сигналов, пришедших по различным ветвям.

Рассмотрим традиционные способы образования независимых ветвей разнесения.

7.4.1 Организация ветвей разнесения

Традиционные пути установления независимых ветвей разнесения могут быть классифицированы следующим образом:

- *пространственное разнесение*;
- *разнесение по частоте*;
- *разнесение во времени*;
- *многолучевое разнесение*.

Пространственное разнесение предполагает создание нескольких независимых путей распространения за счет использования многих антенн, что объясняет другое распространенное название этого метода: *антенное разнесение*. Дублирование антенн может быть применено как на приемной, так и на передающей стороне. Размещенные в пространстве на расстоянии $7 \dots 10$ длин волны или более они обеспечивают практически независимые параллельные интерференционные образцы сигнала на входе приемника. Когда несколько антенн применяются на приемной стороне (*разнесение на приеме*), антенное разнесение оказывается более эффективным, поскольку дополнительные антенны используют энергию сигнала, которая в противном случае не была бы захвачена совсем. В этом случае сигналы разнесения разделяются автоматически, поскольку принимаются различными антеннами. Пройдя индивидуально согласованную фильтрацию, в дальнейшем они могут подвергнуться процедуре комбинирования.

Разнесение с помощью передающих антенн не реализуется столь прямолинейно. Во-первых, ограниченный полный энергетический ресурс передатчика необходимо распределить между несколькими передающими антеннами. Во-вторых, на приемную антенну поступает смесь сигналов, излученных всеми передающими антеннами. Следовательно, перед проведением комбинирования необходимо осуществить несколько процедур, которые обеспечили бы возможность разделения и

индивидуальной обработки этих сигналов. Указанные факторы ставят перед данным типом разнесения сложную задачу оптимизации, решение которой является предметом специального раздела теории связи, получившей наименование *пространственно-временного кодирования*.

Очевидно, что для достижения наибольшего выигрыша может быть использовано комбинирование разнесения на передаче и приеме, когда это позволяют конструктивные ограничения.

Идея разнесения по частоте базируется на концепции канальной *полосы когерентности*. Это понятие определяет участок частотного диапазона, в пределах которого замирания рассматриваются как равные, т.е. искажения частотных компонент сигнала весьма зависимы. И наоборот, гармоники частотного пространства, лежащие вне полосы когерентности, могут трактоваться как независимо искажаемые каналом. Участок частот одинакового замирания обратно пропорционален задержке распространения, чем шире диапазон распределения сигналов во времени, тем уже полоса когерентности. Очевидно, что при одновременной передаче одного и того же сигнала n_d несущими, частоты которых сдвинуты на ширину полосы когерентности или более, образуют n_d ветвей разнесения.

Разнесение во времени использует разность времени прихода многолучевых образцов. Даже тогда, когда приемник неподвижен, профиль многолучевости может оказаться нестабильным из-за движения передатчика или окружающих отражателей. Вследствие этого возникает доплеровское смещение принимаемого сигнала, и чем шире его распределение, тем меньше *время когерентности* канала (интервал времени, в течение которого мощность принимаемого сигнала остается приблизительно стабильной). Обратим внимание на дуальность частоты и времени: диапазон коррелированности в частотной области (полоса когерентности) обратно пропорционален распределению во времени (распределению задержки), тогда как во временной области время когерентности обратно пропорционально частотному (доплеровскому) смещению. Поскольку в моменты времени, разнесенные друг относительно друга на величину времени когерентности или более, помеховые образцы могут считаться независимыми, то повторная передача n_d копий одной и той же информации в соответствующие интервалы времени образуют n_d ветвей разнесения.

Многолучевое разнесение играет особую роль и будет обсужден отдельно в следующем параграфе.

Таким образом, основной целью методов разнесения является такая совместная обработка сигналов ветвей, которая приводит к большему влиянию «лучших» (с более высоким отношением сигнал/шум) ветвей на общую характеристику по сравнению с худшими. Такая совместная обработка называется *комбинированием*.

7.4.2 Методы комбинирования

Предположим, что всего имеется n_d ветвей разнесения, и пусть A_i, φ_i и σ_i будут текущими сигнальными амплитудой и фазой и стандартным отклонением шума в i -ой ветви, где $i = 1, 2, \dots, n_d$.

Любое линейное объединение откликов ветвей $u_i, i = 1, 2, \dots, n_d$ представляется в виде взвешенной суммы $\sum_{i=1}^{n_d} w_i^* \cdot u_i$, где w_i^* - весовые коэффициенты. Тогда результирующее отношение сигнал/шум по мощности q_r^2 представляет собой отношение величины детерминированной компоненты этой суммы к дисперсии ее шумовой компоненты. Таким образом, имеем

$$q_r^2 = \frac{\left| \sum_{i=1}^{n_d} w_i^* A_i \exp(j\varphi_i) \right|^2}{\sum_{i=1}^{n_d} |w_i|^2 \sigma_i^2}. \quad (7.5)$$

Шумовая компонента есть просто взвешенная с коэффициентами $|w_i|^2$ сумма дисперсий шума ветвей, поскольку сами они являются независимыми. Сумма в числителе соотношения (7.5) может рассматриваться как скалярное произведение двух n_d -мерных векторов, компонентами которых являются $w_i \cdot \sigma_i$ и $\left(A_i / \sigma_i \right) \exp(j\varphi_i)$ соответственно.

Используя известное неравенство Шварца (не существует модуля любого скалярного произведения большего произведения длин векторов), получим

$$q_r^2 \leq \frac{\left(\sum_{i=1}^{n_d} |w_i|^2 \sigma_i^2 \right) \left(\sum_{i=1}^{n_d} \frac{A_i^2}{\sigma_i^2} \right)}{\sum_{i=1}^{n_d} |w_i|^2 \sigma_i^2} = \sum_{i=1}^{n_d} \frac{A_i^2}{\sigma_i^2} = \sum_{i=1}^{n_d} q_i^2, \quad (7.6)$$

где $q_i = A_i / \sigma_i$ - отношение сигнал-шум по напряжению в i -й ветви разнесения. При оптимальном выборе весов

$$w_i^* = \frac{A_i}{\sigma_i} \exp(-j\varphi_i), \quad (7.7)$$

неравенство (7.6) обращается в равенство, т.е. достигается максимально возможное результирующее отношение сигнал-шум. Выбор таких весов обеспечивает обобщенную согласованную фильтрацию откликов разнесенных ветвей.

Практическая реализация подобного фильтра возможна только в том случае, когда известны точные значения амплитуд и фаз всех сигналов. Тогда сигналы с соответствующими весовыми коэффициентами могут быть когерентно сложены. Подобная техника комбинирования известна как *техника с максимальным отношением*.

Для оценивания эффективности метода комбинирования обозначим максимальное по всем ветвям разнесения отношение сигнал-шум через $q_{\max}$ и введем понятие выигрыша от разнесения, как $G_d = q_r^2 / q_{\max}^2$.

Поскольку $\sum_{i=1}^{n_d} q_i^2 \leq n_d q_{\max}^2$, то ни одна из схем комбинирования не сможет обеспечить выигрыш, больший чем n_d . Причем равенство в последнем соотношении достигается только в схеме с максимальным отношением при выполнении дополнительного условия о равенстве отношения сигнал-шум во всех ветвях разнесения.

Альтернативными методами являются метод комбинирования с равными весами и метод выбора ветви с максимальным отношением сигнал/шум. Первый метод достигает эффективности способа с максимальным отношением только при одинаковом отношении сигнал/шум во всех ветвях разнесения. Выигрыш второго метода близок к величине, обеспечиваемой оптимальной схемой, при условии доминирования в отношении сигнал/шум одной из ветвей разнесения над всеми остальными. Очевидно, что данные стратегии также могут комбинироваться друг с другом, например, первоначально выбираются

несколько ветвей с наилучшим отношением сигнал/шум, выходные отклики которых затем суммируются с равными весами.

7.5 Принципы многолучевого разнесения

Схема многолучевого разнесения базируется на том факте, что сигналы, распространяющиеся различными путями, достигают приемник с различными временными запаздываниями. Предположим, что результирующий принятый сигнал с комплексной огибающей $\dot{S}(t)$ прошел фильтр, согласованный с сигналом $s(t)$. Тогда, учитывая линейность фильтрации и связь между АКФ сигнала и откликом согласованного фильтра, выражение для комплексной огибающей $\dot{S}_{rf}(t)$ сигнала на выходе фильтра имеет вид

$$\dot{S}_{rf}(t) = \sum_i A_i \dot{R}(t - \tau_i - T) \exp(j\phi_i), \quad (7.8)$$

где $\dot{R}(\tau)$ - АКФ комплексной огибающей $\dot{S}(t)$ исходного сигнала, а T - обозначает длительность сигнала $s(t)$.

Пусть время корреляции сигнала τ_c , т.е. протяженность во времени АКФ, не превышает минимальной взаимной задержки между последовательными во времени многолучевыми сигналами: $\tau_{\min} = \min_i \{\tau_i - \tau_{i-1}\} : \tau_c \leq \tau_{\min}$. Очевидно, что в такой ситуации все многолучевые сигналы после согласованного фильтра не будут перекрываться. Поскольку они полностью разрешаются во времени и не мешают друг другу, то они могут рассматриваться как сигналы независимых ветвей разнесения и обрабатываться согласно одному из алгоритмов комбинирования, рассмотренных выше. Если, например, известны временные позиции, амплитуды и начальные фазы (предварительно измеренные с использованием отдельного пилотного канала), то наилучшим является метод комбинирования с максимальным отношением.

Следует понимать, что для реализации этой схемы многолучевого разнесения сигнал должен иметь узкую АКФ. Существует прямолинейное решение данной задачи, состоящее в использовании коротких сигналов. Данный путь предполагает излучение сигналов с высокой пиковой мощностью, что может оказаться неприемлемым во многих случаях. Более привлекательным является использование специальных сигналов, для которых характерен эффект временного сжатия в согласованном фильтре, т.е. имеющих малое время корреляции в сравнении с длительностью

сигнала: $\tau_c \ll T$. Сигналы с подобными свойствами называются сложными и могут быть найдены только среди сигналов с распределенным спектром.

Метод многолучевого разнесения уникален в том смысле, что он радикально изменяет отношение к эффектам многолучевости, которые на первый взгляд представляются безоговорочно вредными. Действительно, любой отражатель, принимающий участие в распространении сигнала, направляет на приемник часть излученной мощности, которая бы, в противном случае, была бы полностью потеряна. В том случае, когда эти отраженные сигналы могут быть разделены друг относительно друга (разрешены во времени), их энергия используется для улучшения характеристик системы в сравнении со случаем их отсутствия. В этой схеме канал как будто сам создает ветви разнесения, и единственной проблемой является адекватный выбор сигнала, который позволил бы разрешить многолучевые копии.

Обратимся теперь к иллюстрации рассмотренной схемы разнесения в приложении к цифровой связи, представленной на рис. 7.2. Для большей прозрачности на диаграммах показаны только низкочастотные эквиваленты (комплексные огибающие) всех сигналов.

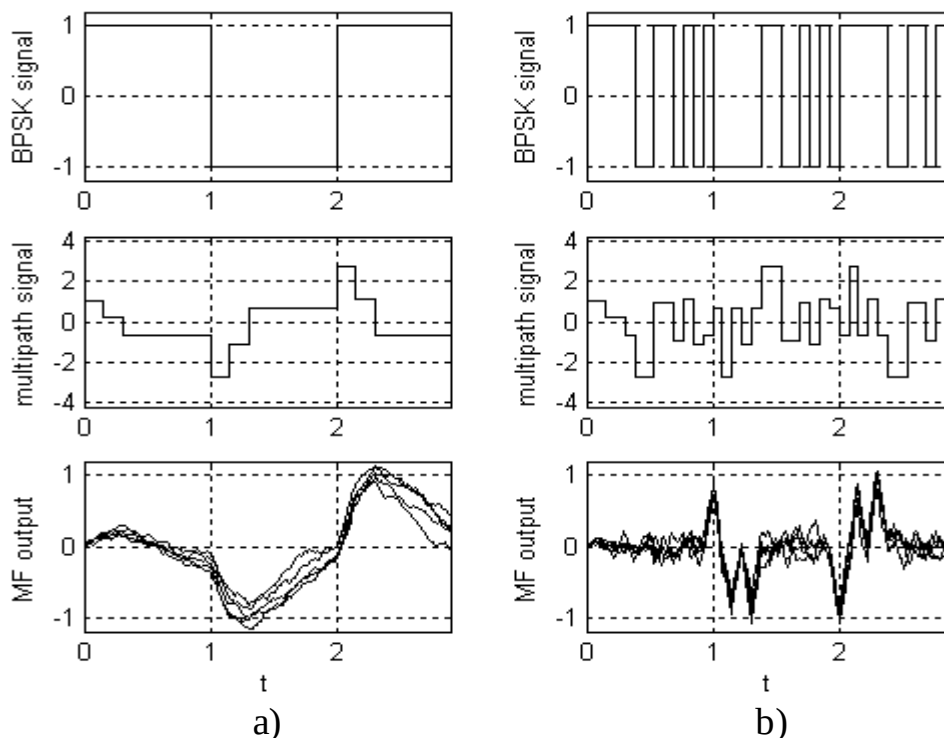


Рис. 7.2. Эффекты многолучевости при бинарной ФМ передаче:
а) простые сигналы, б) сигналы с распределенным спектром.

Левая колонка (а) отвечает передаче одного бита простым прямоугольным импульсом, причем нулевое значение бита передается

импульсом положительной полярности (верхняя диаграмма). Вторым графиком колонки соответствует результирующему сигналу на выходе канала при условии пренебрежения шумом. Хотя межсимвольные искажения проявляют себя искажением начальной части битового импульса, большая часть последнего подвергается амплитудному замиранию. Последняя диаграмма демонстрирует 5 перекрывающихся реализаций отклика согласованного фильтра на результирующий сигнал в присутствии шума. Вредный эффект многолучевого распространения очевиден: надежные решения о значениях передаваемых битов вряд ли возможен.

Диаграммы на колонке (b) отвечают методу передачи, при котором значение каждого бита определяет полярность широкополосного сигнала, в частности, кода Баркера длины $N=13$. Верхняя диаграмма демонстрирует три подобных сигнала, манипулированных таким же битовым образцом, что и ранее. На второй диаграмме представлен результирующий сигнал на выходе канала в отсутствие шума. Самый нижний рисунок отвечает отклику согласованного фильтра на результирующий сигнал, искаженный белым шумом такой же интенсивности, что и предыдущий пример. Три характерных пика, приходящихся на каждый переданный бит, пригодны для восстановления данных с высокой степенью надежности. Полагая модель канала заранее известной, можно осуществить взятие отсчетов в моменты, соответствующие максимумам каждой многолучевой компоненты сигнала на выходе согласованного фильтра в отсутствие шума.

Как видно из рисунка, правильно выбранный широкополосный сигнал для передачи бита данных обеспечивает разрешение всех многолучевых компонент на выходе фильтра без взаимного искажения. Затем три отсчета могут быть оптимально скомбинированы, т.е. взвешены пропорционально своим амплитудам и просуммированы с учетом полярности второй и третьей компонент. Устройство, реализующее описанное выше комбинирование, представлено на рис. 7.3, в котором отводы линии задержки настроены на выравнивание во времени трех многолучевых пиков на выходе фильтра. Затем производится оптимальное комбинирование отсчетов: третья компонента без задержки с весом $-A_3 = -0.9$ и вторая, задержанная на величину $\tau_3 - \tau_2$, с весом $-A_2 = -0.8$ складываются с первой, задержанной на $\tau_3 - \tau_1$. В результате получаем отношение сигнал/шум по напряжению, которое в $\sqrt{1 + A_2^2 + A_3^2} \approx 1.56$ раз больше, чем получаемое для сигнала по первому пути.

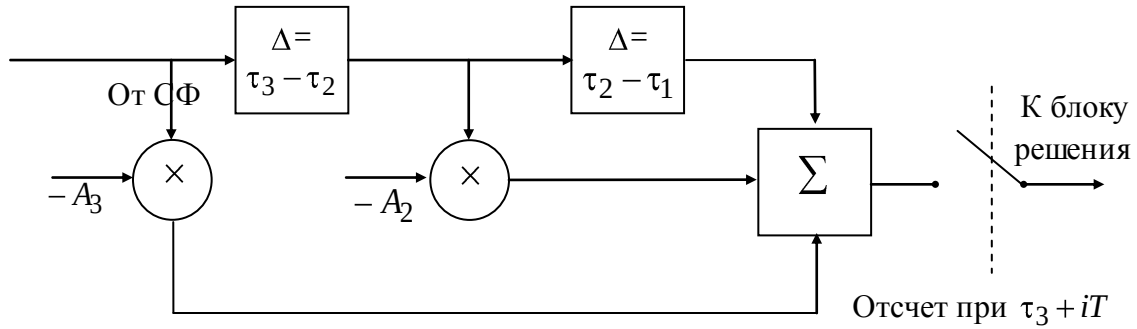


Рис. 7.3. Устройство комбинирования с максимальным отношением.

Выигрыш в отношении сигнал-шум наглядно демонстрирует рис. 7.4.

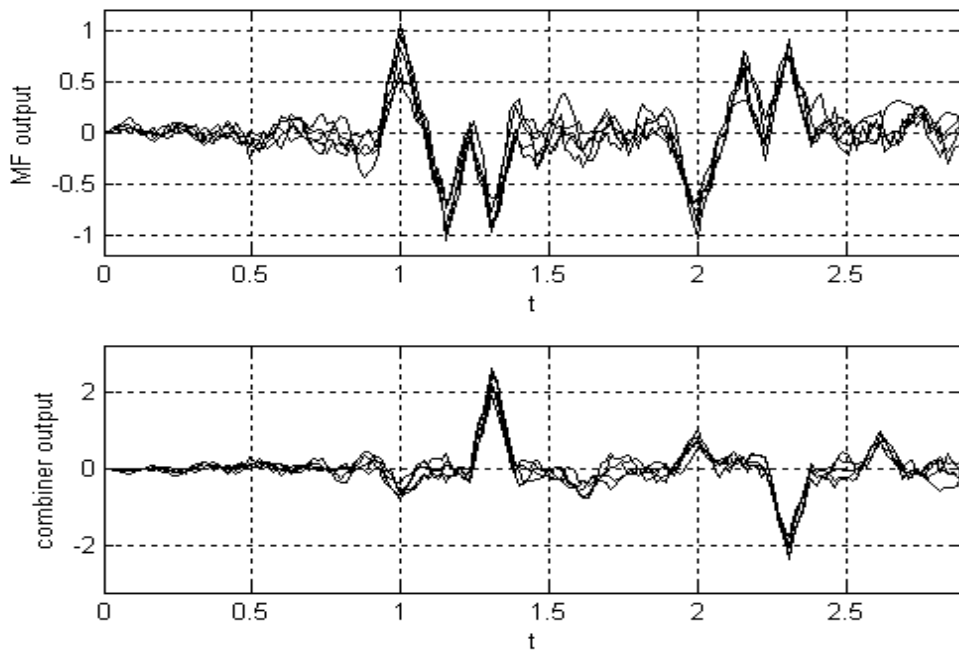


Рис. 7.4 Сигналы на выходе согласованного фильтра и сумматора.

На приведенном рисунке верхняя диаграмма повторяет сигнал на выходе согласованного фильтра рис. 7.3, б), а нижняя показывает напряжение выходного сигнала сумматора. Взятие отсчетов с выхода сумматора в моменты времени $\tau_3 + T, \tau_3 + 2T, \dots$ и фиксация их полярности обеспечивает решения о значении переданных бит данных.

8 ПРИЕМ СИГНАЛОВ В КАНАЛАХ СО СЛУЧАЙНЫМИ ПАРАМЕТРАМИ

8.1 Импульсная характеристика канала связи

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

Для иллюстрации на рис. 8.1 представлена типовая осциллограмма принятого многолучевого сигнала, когда передатчик излучал радиоимпульсы длительностью 0.1 мкс. Ее обычно называют *профилем многолучевости*.



Рис. 8.1. Профиль многолучевости в условиях городской застройки

Запаздывания сигналов отдельных лучей измеряются относительно сигнала, приходящего по линии визирования (ЛВ): передатчик - приемник (прямой луч). Его запаздывание обозначают как ЛВЗ - запаздывание вдоль линии визирования. Видно, что при этом не все лучи разрешаются. Часть из них (отраженная от протяженных объектов), сливается в импульсы длительностью более 0,1 мкс. Профиль многолучевости есть не что иное, как действительная часть импульсной характеристики канала $\text{Re } h_c$.

Применительно к широкополосным системам, которым будет уделено основное внимание, целесообразно положить, что все лучи разрешимы, т. е. выполняется условие $|t_k - t_l| > 1/W$ для всех $k \neq l$.

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

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

Прием сигналов в условиях многолучевости распространения представляет собой одну из форм разнесенного приема, где информация от передатчика к приемнику передается по нескольким естественно разнесенным лучам. Под временем многолучевого растяжения сигнала τ_p , которое выше определялось как длительность профиля многолучевости, будем понимать разницу между максимальным и минимальным запаздыванием лучей: $\tau_p = \max \tau_{зан} - \min \tau_{зан}$.

Принцип многолучевого разнесения был предложен Price и Green в 1958 г. С тех пор этот метод широко известен под названием RAKE («грабли»), поскольку пики на выходе согласованного фильтра напоминают в некоторой степени зубцы популярного садового инструмента.

Известны многочисленные версии аппаратной реализации RAKE алгоритма.

8.2 Корреляционный приемник при многолучевом распространении сигнала

Корреляционный приемник предполагает использование n_d параллельных корреляторов, где n_d определяется числом зубцов, т.е. числом используемых ветвей разнесения. При использовании комбинирования с максимальным отношением RAKE приемник на основе корреляторов представим схемой, изображенной на рис. 8.2 (версия для видеосигналов).

Эта структура наиболее практична в случае, когда задержки многолучевых сигналов точно измерены и могут считаться известными. Коррелятор с опорным сигналом, задержанным на величину τ , выдает значение, равное отсчету на выходе согласованного фильтра в момент $T + \tau$. Тогда, учитывая, что для вынесения решения необходимы отсчеты в

моменты $T + \tau_i, i = 1, 2, \dots, n_d$, в качестве опорных сигналов корреляторов можно использовать n_d задержанных копий исходного сигнала, а затем осуществить необходимое комбинирование выходных значений корреляторов.

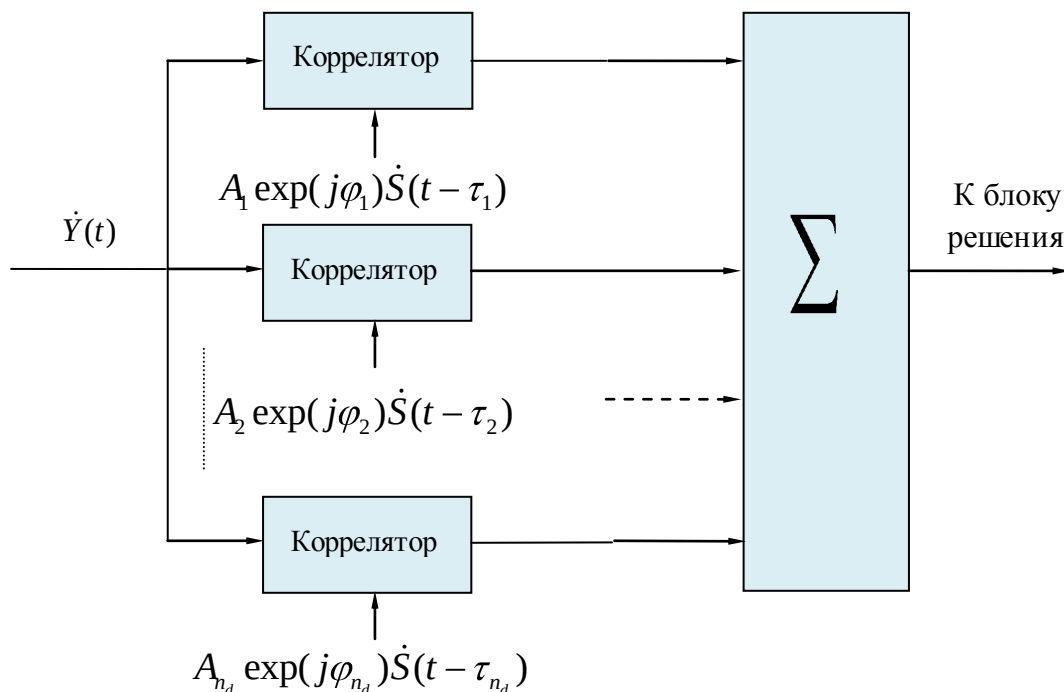


Рис. 8.2. Корреляционный RAKE приемник.

8.3 Оптимальный некогерентный приемник для однолучевого канала

Сначала рассмотрим простейший случай, когда канал связи образован только одним лучом. Предположим сначала, что интенсивность луча и его запаздывание известны (для упрощения положим $t_0 = 0$), а фаза несущей θ_0 не известна и является случайной переменной, равномерно распределенной на интервале $[0, 2\pi]$.

Поскольку отсутствие многолучевости подразумевает отсутствие межсимвольной интерференции, то основное внимание мы уделим приему отдельного символа. Оптимальный приемник, т. е. приемник, который с минимальной вероятностью ошибки принимает решение о том, какой из сигналов $s_i, i = 1 \dots M$ передан, имеет, как известно, структуру, показанную на рис. 8.3. Это некогерентный приемник: фаза любого луча полагается случайной, так как при приеме сигналов подвижной станцией она изменяется слишком быстро, для того чтобы можно было применить когерентные методы приема.

Из рисунка видно, что принятый сигнал $y(t)$ пропускается через набор из M фильтров, каждый из которых согласован с одним из M сигналов, т. е. через фильтры с импульсными характеристиками $s_i, T-t, 0 \leq t < T$. К выходам фильтров подключены детекторы огибающей, выходные напряжения которых сравниваются в момент времени $t = T$. Индекс $i = 1 \dots M$ наибольшего из них является выходом приемника. Изображенный на рис. 8.3 случай соответствует переданному сигналу с индексом “ j ”, шумовая составляющая принятого сигнала полагается пренебрежимо малой.

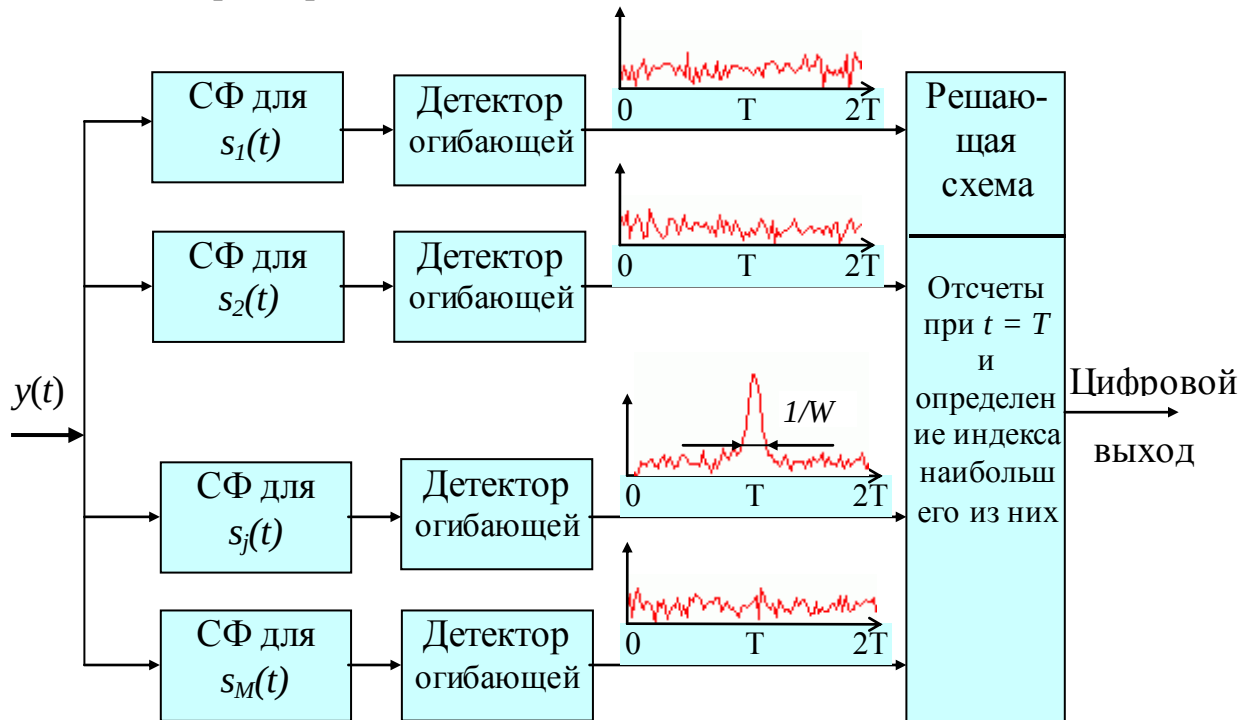


Рис. 8.3. Оптимальный некогерентный приемник для однолучевого канала (равновероятные сигналы равной энергии; гауссовский шум).

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

$$R_{ik} = \int_0^T S_i^*(\tau) \cdot S_k(\tau - t) d\tau, \quad i, k = (1, M), \quad |t| \leq T, \quad (8.1)$$

показанные на рисунке в виде шумовых дорожек.

Максимум (8.1) гарантирует, что огибающая j -го канала будет иметь острый главный пик, окруженный низкими «боковыми лепестками», а остальные выходные сигналы будут содержать только боковые лепестки низкого уровня. При тщательном выборе сигналов максимальный уровень боковых лепестков у всех сигналов будет примерно в $2\sqrt{WT}$ раз ниже уровня главного лепестка.

8.4 Оптимальный приемник многолучевого сигнала. Случай известных запаздываний

Если мы попытаемся использовать приемник, показанный на рис. 8.3, в случае, когда имеется несколько лучей ($K > 1$), то вследствие линейности среды и наличия согласованных фильтров можно ожидать, что напряжения на выходах детекторов огибающей будут выглядеть примерно так, как показано на рис. 8.4, на котором они изображены для случая $K = 4$. Здесь через Δ обозначено время многолучевого растяжения: $\Delta = \tau_p$.

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

$$e_l(t) = \left| \sum_{k=0}^{K-1} a_k \exp(j\theta_k) \int_0^T S_j^*(\tau) \cdot S_l(\tau + T + t_k - t) d\tau \right|. \quad (8.2)$$

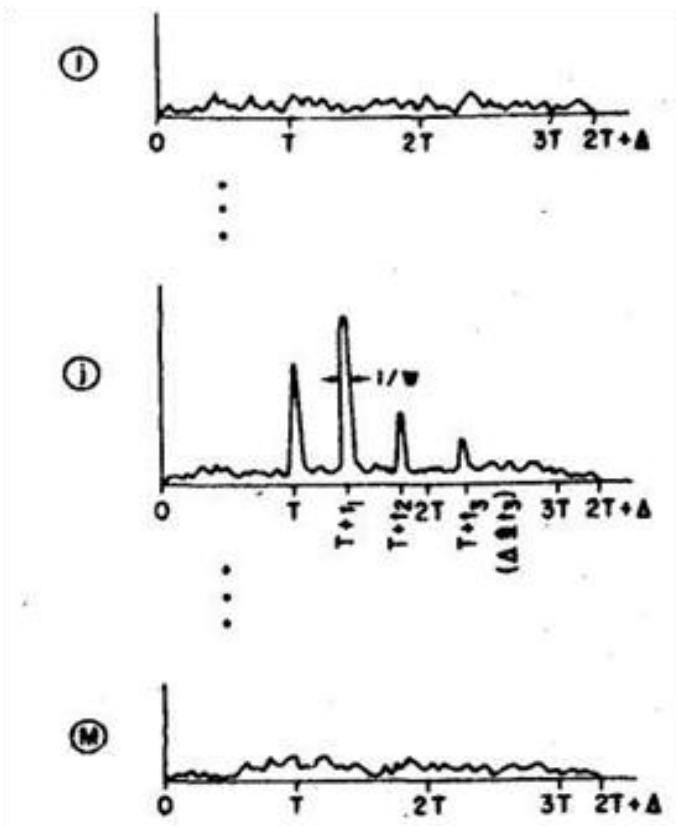


Рис. 8.4. Выходы детекторов огибающих. 4-х лучевой случай.

При выполнении условия разрешимости главные пики огибающей $e_l(t)$ различны и соответствуют моментам времени $t_0, t_1, t_2, t_3 = \Delta$. Высоты этих пиков пропорциональны интенсивностям лучей. Боковые лепестки и сигнала $e_j(t)$, и других выходных сигналов (ни один из последних не имеет главного пика) представляют собой смесь боковых лепестков сигналов, пришедших по нескольким лучам. Еще раз напомним, что рис. 8.3 соответствует передаче только одного сигнала $s_j(t)$, $0 \leq t < T$.

Сигналы, показанные на рис. 8.3 и 8.4, имеют несколько важных различий.

1) Огибающая $e_l t$ имеет несколько сильных пиков. Если решающей схеме приемника, показанного на рис. 8.2, известны времена запаздываний $t_0 \dots t_{K-1}$, то она может объединить отсчеты всех выходных сигналов, тем самым обеспечивая приемнику преимущества разнесенного приема.

2) Уровни боковых лепестков всех выходных сигналов возрастают, что, согласно выражению (8.2), обусловлено дополнительными вкладами из-за многолучевости.

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

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

Возможность разрешения лучей по сигналам, показанным на рис. 8.3, составляет суть широкополосного подхода. Если бы мы имели $WT \approx 1$, то пики на рис. 8.3 слились бы и идея разнесенного приема (т. е. объединение сигналов) в явном виде была бы больше не применима.

Рассмотрим структуру приемника, для оптимального приема отдельного символа в условиях многолучевости распространения.

Предположим сначала, что времена запаздывания сигналов t_k для всех лучей известны. Кроме того, будем снова полагать, что фазы сигналов случайны, независимы и равномерно распределены на интервале $[0, 2\pi]$, а интенсивности лучей a_k - также случайны и, возможно, имеют различные распределения.

Ясно, что при наложенных условиях оптимальный приемник все еще будет иметь форму, показанную на рис. 8.3. Однако решающая схема должна теперь брать отсчеты каждого своего входного сигнала в моменты времени $T + t_k$, $k=0, \dots, K-1$, объединять их для каждого входного сигнала и затем сравнивать полученные величины для определения индекса наибольшей из них.

Пусть x_{ik} - отсчет i -й огибающей, взятый в момент времени $T + t_k$ при отсутствии шума $x_{ik} = e_i t_k$.

Тогда, если интенсивности лучей известны, оптимальный закон объединения отсчетов будет определяться выражением

$$w_l = \sum_{k=0}^{K-1} \ln I_0 \left(\frac{2a_k x_{ik}}{N_0} \right). \quad (8.3)$$

Если интенсивность k -го луча имеет релеевское распределение со среднеквадратичным значением $\psi_k = E[a_k]^2$, а интенсивности всех лучей независимы, то оптимальный закон объединения отсчетов будет иметь следующий вид

$$w_l = \sum_{k=0}^{K-1} \frac{\psi_k x_{ik}^2}{N_0 + \psi_k E}, \quad (8.4)$$

где E – энергия сигналов.

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

Таким образом, любой оптимальный закон объединения состоит в относительном подчеркивании более надежных данных и в относительном подавлении менее надежных.

8.5 Оптимальный приемник многолучевого сигнала. Случай неизвестных запаздываний

Как отмечалось ранее, времена запаздывания лучей и число лучей часто являются случайными и заранее не известными переменными. Для того чтобы определить структуру оптимального приемника в этом случае, мы несколько упростим модель многолучевости.

Будем полагать, что времена $t_0 \dots t_{K-1}$ независимы и описываются одним общим распределением плотности вероятностей $p(t_k)$, $0 \leq t_k < \Delta$, а их индексы переупорядочены в соответствии с ростом величины t_k . Такая модель противоречит принятым допущениям о разрешимости лучей. В данном случае возможны два луча, для которых $|t_k - t_l| < 1/W$. Но вероятность этого события будет мала, если мала суммарная вероятность интервалов длительностью, равной или меньшей $1/W$. Следовательно,

распределение $p(t_k)$ должно иметь размытый («диффузный») характер, не иметь высоких пиков и удовлетворять условию $W\Delta \gg 1$.

Однако для отыскания структуры нашего квазиоптимального приемника указанные отличия от реальности несущественны. Поэтому для случая неизвестных запаздываний интенсивности любых двух лучей будут полагаться независимыми, что отличается от реальности, так как лучи с незначительно различающимися запаздываниями имеют, как правило, коррелированные интенсивности. Однако интуитивно ясно, что приемник, полученный на основе этой упрощенной модели интенсивностей лучей, будет близок к оптимальному.

Такой приемник, с учетом сделанных допущений, должен вычислять величины

$$w_i = \int_T^{T+\Delta} p(t-T) \cdot F[X_l(t)] dt, \quad (8.5)$$

где $p(\cdot)$ – определенная выше плотность распределения времени запаздывания лучей, $X_l(t)$ – огибающая выхода l -го фильтра, а $F[\cdot]$ – некоторая нелинейная функция.

Затем значения этих величин сравниваются и индекс наибольшей из них считается принятым сигналом. Таким образом, величину w_i можно реализовать, пропуская $X_l(t)$ сначала через нелинейность F , а затем через фильтр с импульсной характеристикой $h_p(t) = p(\Delta - t)$ и беря отсчет сигнала на выходе этого фильтра в момент времени $t = T + \Delta$. Заметим, что фильтр $h_p(t)$ в идеале должен быть согласован с импульсной характеристикой канала $h_c(t) : h_p(t) = h_c(\Delta - t)$.

Решающая схема, показанная на рис. 8.3, в этом случае заменяется схемой, показанной на рис. 8.6.

Следует отметить, что в отличие от выражений (8.3) и (8.4), где отсчеты x_{ik} объединяются по линейному и квадратичному законам, для объединения отсчетов в выражении (8.5) используются экстремальные нелинейности экспоненциального типа

$$F(X_i) = \exp \left[\frac{\psi x_i^2}{N_0 + \psi E} \right] - 1. \quad (8.6)$$

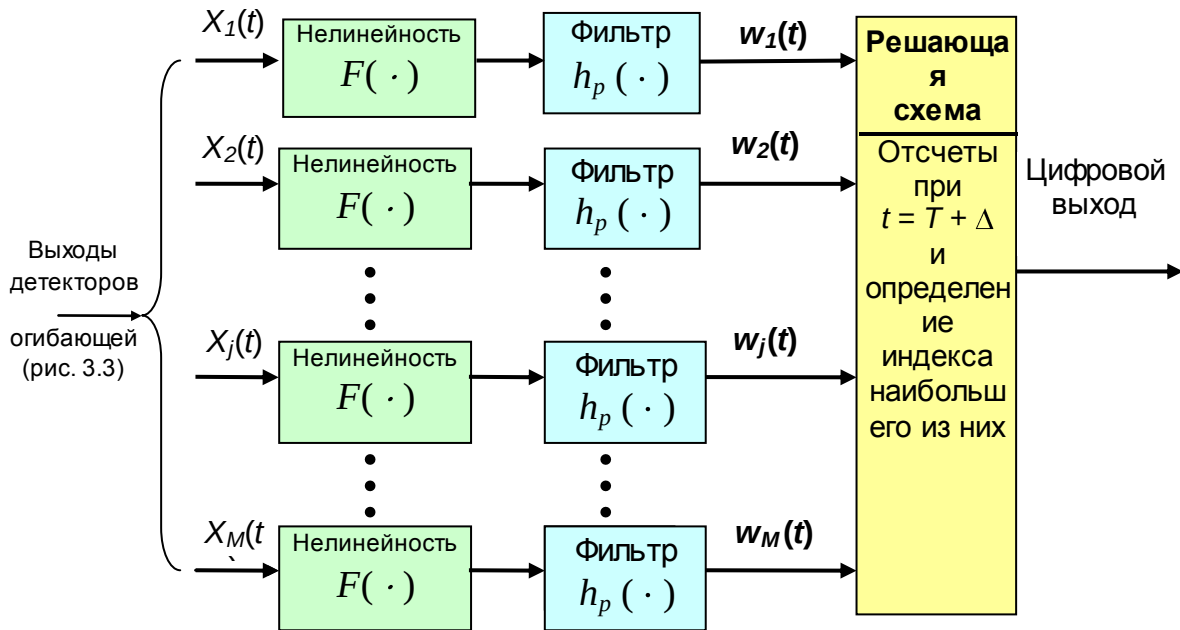


Рис. 8.6. Оптимальный некогерентный приемник для многолучевого канала с неизвестными временами запаздывания лучей.

Такая нелинейность (см. рис.8.6) подчеркивает сильные сигналы и подавляет слабые. В этом проявляется определенная «самоадаптивность» приемника.

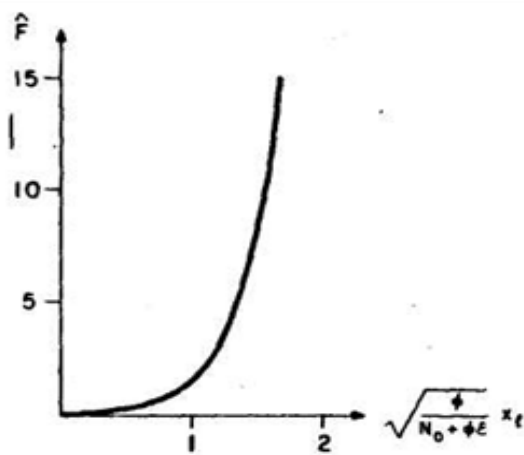


Рис. 8.6. Используемый тип нелинейности $F(\cdot)$

Следовательно, сигналы на выходе нелинейностей несут в себе информацию, непосредственно характеризующую апостериорные вероятности существования лучей.

Эти сигналы далее взвешиваются в соответствии с априорными вероятностями размещения лучей, в неявном виде входящими в функцию $p(\cdot)$.

Следует подчеркнуть, что изложенная позитивная роль нелинейности проявляется при относительно большой средней величине ОСШ, когда пики сигнала

имеют большую величину по сравнению с шумом и уровнем боковых лепестков. При малом ОСШ эффект подавления шума нелинейностями будет выражаться слабее.

Можно показать, что в данном случае приемник сигналов с известными запаздываниями, в котором отсчеты выходных сигналов детекторов огибающей берутся только в моменты времени, соответствующие пикам сигналов, а вклады шума не учитываются, будет

иметь более высокие характеристики, чем приемник сигналов с неизвестными запаздываниями.

8.6 Оптимальные приемники, зондирующие канал

До сих пор, рассматривая структуру приемника, мы исходили из априорных сведений о статистиках канала. Однако во многих случаях характеристики канала можно измерить с помощью зондирующих сигналов, что позволяет определить апостериорные статистики. В качестве зондирующих сигналов могут использоваться либо специальные сигналы, либо сами информационные сигналы; в последнем случае зондирование и передача данных осуществляются одновременно.

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

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

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

«Оптимальный» приемник M -лучевого сигнала, в котором используются оценки интенсивностей и запаздываний, можно, очевидно, реализовать на основе выражения (8.3). Такой приемник должен вычислять величины

$$w_l = \sum_{k=0}^{K-1} \ln I_0 \left(\frac{2\hat{a}_k \hat{x}_{lk}}{N_0} \right) \quad (8.7)$$

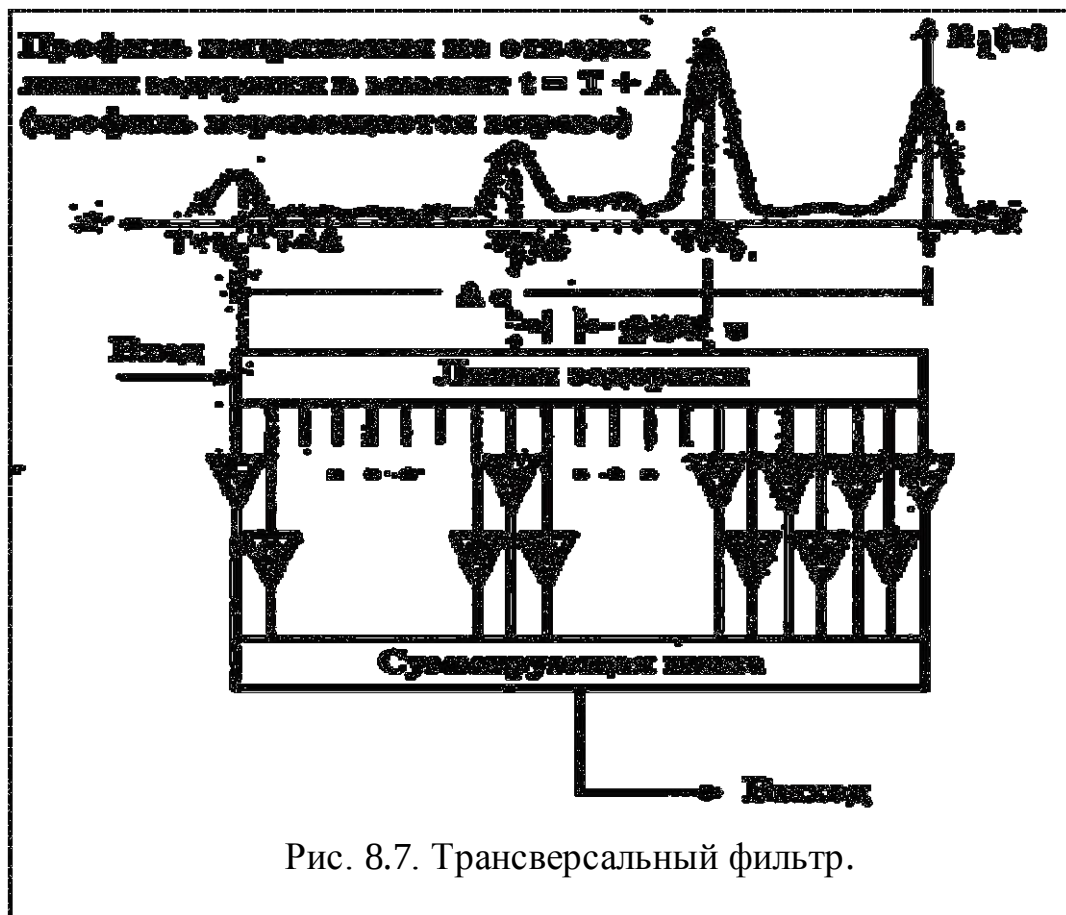
где $\hat{a}_k$ - оценка интенсивности k -го луча, $\hat{x}_{lk}$ - отсчет огибающей сигнала на выходе l -го согласованного фильтра в момент времени $t - T + t_k$, а $\hat{t}_k$ - оценка запаздывания k -го луча. Индекс $l=1, \dots, M$, для которого величина w_l максимальна, является цифровым выходом приемника.

Если ОСШ для каждого луча достаточно велико, то (8.7) можно аппроксимировать выражением

$$w_l \cong \frac{2}{N_0} \sum_{k=0}^{K-1} \hat{a}_k \hat{x}_{lk}, \quad (8.8)$$

так как при больших x $\ln I_0(x) \approx x$. Выражение (8.8) описывает оптимальный линейный сумматор Бреннана, используемый при разнесенном приеме сигналов; мы также будем его использовать, поэтому остановимся на нем несколько подробнее.

Линейный сумматор, описываемый выражением (8.8), можно реализовать с помощью трансверсального фильтра, как показано на рис. 8.7.



Этот фильтр состоит из линии задержки на время Δ , у которой сделаны отводы с интервалом, равным по крайней мере $1/W$, так что всего имеется $W\Delta$ отводов. На вход трансверсального фильтра подается напряжение с выхода детектора огибающей, например $x_j(t)$ (см. рис. 8.5). Выход трансверсального фильтра представляет собой взвешенную сумму

сигналов с определенных отводов, номера которых зависят от оценок запаздываний лучей.

Оценки t_k используются для включения усилителей, подсоединенных к отводам, задержки которых (измеряемые с правого конца линии) наиболее точно аппроксимируют оценки ϵ_k . Иными словами, активируется K усилителей, коэффициенты усиления которых устанавливаются пропорционально величинам соответствующих оценок интенсивностей ϵ_k . На рис. 8.7 указаны коэффициенты усиления усилителей при приеме четырехлучевого сигнала.

Для того чтобы лучше понять, как работает трансверсальный фильтр, на рис. 8.7 показан профиль изменения напряжения вдоль линии задержки для случая, когда на ее вход поступает сигнал $x_j(t)$ с выхода детектора огибающей, включенного после j -го согласованного фильтра (см. рис. 8.4). Этот профиль постепенно сдвигается вправо, и на рис. 8.117 его положение соответствует моменту времени $t = T + \Delta$. В этот момент пики сигнала соответствуют задержкам $t_0, t_1, \dots, t_{K-1}$, измеряемым с правого конца линии. На рис. 8.7 изображен случай, когда $t_0 = 0$ и $K=4$, причем предполагается, что $t_s = \Delta$, где Δ - максимальное относительное запаздывание (т. е. превышение над t_0 , ожидаемое в данном канале). Если ϵ_k - точные оценки времен запаздывания, то в момент времени $t = T + \Delta$, напряжение на выходе трансверсального фильтра будет приближенно равно величине

$$\sum_{k=0}^{K-1} \hat{a}_k \hat{x}_{j, T+\hat{t}_k} \cong \sum_{k=0}^{K-1} \hat{a}_k \hat{x}_{jk}, \quad (8.9)$$

т. е. оказывается пропорциональной величине w_l , определяемой выражением (8.8) при $l = j$.

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

В процессе перемещения $x_j t$ на рис. 8.7 вправо трансверсальный фильтр осуществляет свертку $x_j t$ с $h_{ij} t$, показанных на рис. 8.8, а) и 8.8, б). В результате чего формируется выходной сигнал, показанный на рис. 8.8, в). Главный пик этого сигнала соответствует моменту времени $t = T + \Delta$, т. е. тому моменту, когда входной сигнал $x_j t$ занял такое

положение, при котором его пики расположены точно над активированными отводами линии задержки. Высота главного пика пропорциональна величине (8.9). Выходной сигнал имеет также небольшие побочные пики, пропорциональные величинам $\hat{a}_k \cdot \hat{x}_{jr}$

$k=0, 1, 2, 3; r \neq k$, и боковые лепестки, накопленные в результате свертки $h_{ij}(t)$ с боковыми лепестками сигнала $x_j(t)$. Ясно, что для получения соотношения (8.9) необходимо оставить только один главный пик.

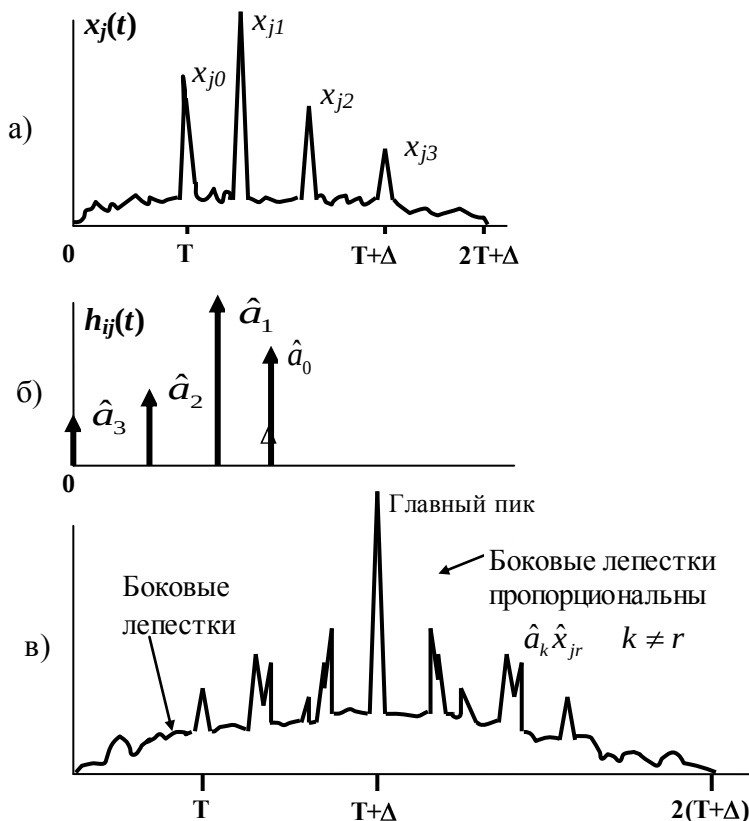


Рис. 8.8. (а) - вход фильтра; (б) - импульсная характеристика фильтра; (в) - выход фильтра (свертка входа и импульсной характеристики фильтра)

Таким образом, полный приемник M -лучевого сигнала, в котором используются оценки характеристик канала, будет иметь вид, показанный на рис. 8.9. Выход $x_l(t)$, $l=1, \dots, M$ каждого детектора огибающей подается на вход трансверсального фильтра типа показанного на рис. 8.7, параметры которого изменяются в соответствии с оценками $\{\hat{t}_k\}$ и $\{\hat{a}_k\}$, полученными от зондирующего приемника.

В момент времени $t = T + \Delta$ выходы трансверсальных фильтров считаются и

сравниваются, в результате чего определяется индекс наибольшего из них.

Зондирующий приемник, показанный на рис. 8.9, также может иметь структуру, образованную последовательным соединением согласованного фильтра, детектора огибающей и линии задержки с отводами, причем согласованный фильтр согласуется с известным колебанием, которое может иметь большую энергию и длительность, чем информационные сигналы s_j . Перед передачей данных (или во время передач, если характеристики канала изменяются достаточно быстро) зондирующий

приемник «слушает» свой сигнал. Когда сигнал принят, огибающая напряжения на выходе согласованного фильтра будет иметь множество пиков, т. е. будет иметь вид, аналогичный показанному на рис. 8.8, а). Затем это колебание будет перемещаться по линии задержки так же, как показано на рис. 8.7. Когда первый пик сигнала достигает правого конца линии, срабатывает пороговый элемент и берутся отсчеты напряжений на всех отводах линии задержки, и их значения, которые пропорциональны интенсивностям лучей, запоминаются, если они превышают некоторую пороговую величину, в противном случае они полагаются равными нулю. Далее эти отсчеты используются для установки коэффициентов усиления усилителей, подключенных к отводам трансверсальных фильтров приемника информационных сигналов.

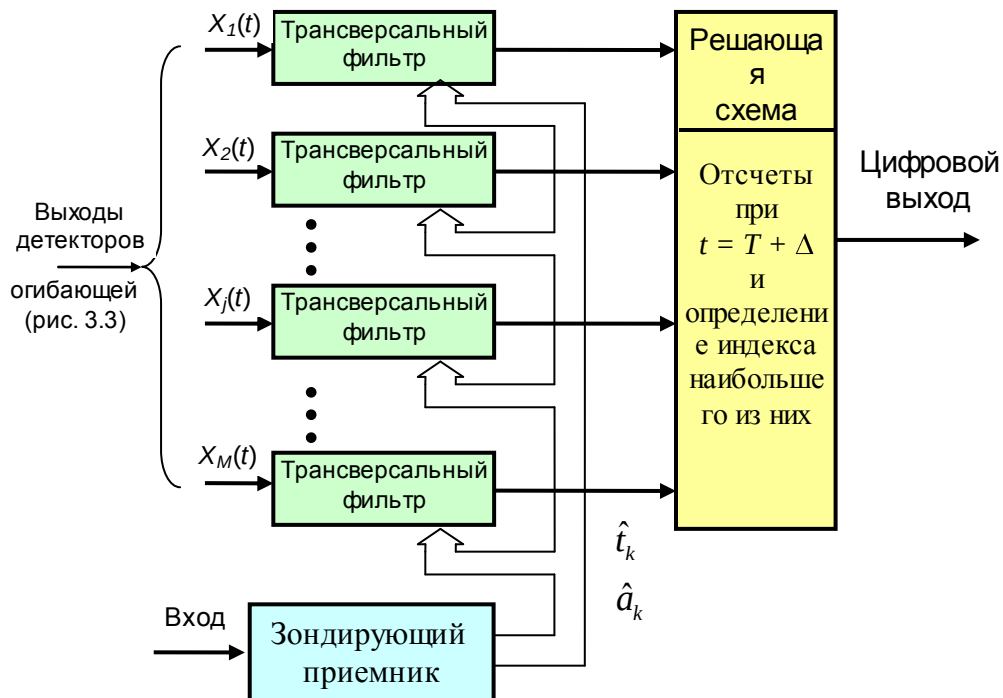


Рис. 8.9. Приемник многолучевого сигнала с зондированием канала (приемник системы «Рейк»).

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

8.7 Оптимальный прием при наличии межсимвольной интерференции

До сих пор мы пренебрегали эффектами межсимвольной интерференции, поскольку основное внимание уделялось приемнику отдельных изолированных посылок.

Предположим теперь, что каждая посылка имеет длительность T , а последовательная передача любых посылок осуществляется с интервалом T_s . В соответствии с выражением (8.2) отклики согласованных фильтров в рассмотренных выше приемниках на каждую посылку длиной T будут длиться $2T + \Delta$, что может привести к межсимвольной интерференции.

Для того чтобы наглядно проиллюстрировать возникновение межсимвольной интерференции в широкополосном случае $TW \gg 1$, на рис. 8.10 показана последовательность огибающих откликов отдельного согласованного фильтра на периодически повторяемый входной сигнал при двух значениях T_s . На рис. 8.10, а) показаны огибающие откликов согласованного фильтра на три последовательные посылки для случая, когда $T_s \geq T + \Delta$. Для наглядности показаны также линия задержки приемника «Рейк» длиной Δ с. и сумматор (см. рис. 8.7). Межсимвольная интерференция в данном случае полностью отсутствует. Указанная последовательность огибающих выходных напряжений фильтра «заморожена» в тот момент времени, когда пики центрального n -го многолучевого входного сигнала точно соответствуют отводам линии задержки и «готовы» для взятия отсчетов. Заметим, что в данный момент $n-1$ -й и $n+1$ -й отклики «находятся» вне линии задержки, и, следовательно, не могут интерферировать с n -м откликом.

На рис. 8.10, б) показан случай умеренной межсимвольной интерференции. Здесь $T_s = T$, а $\Delta = 2T$, так что $T_s = T + \Delta/3$. Огибающие откликов показаны для семи последовательных посылок, которые зафиксированы в тот момент времени, когда пики n -го многолучевого входного сигнала точно соответствуют отводам линии задержки. Однако в данном случае два предшествующих $n-2$ -й и $n-1$ -й) сигнала еще частично не вышли из линии задержки, а два последующих $n+1$ -й и $n+2$ -й уже частично вошли в нее, и это обстоятельство приводит к возникновению межсимвольной интерференции. В общем случае с каждым символом будет интерферировать $2\lceil \Delta/T \rceil$ предшествующих и последующих символов, где $\lceil x \rceil$ - наименьшее целое число, большее или равное x .

Используемый в данном случае подход основан на том факте, что, хотя в момент принятия решения о принятом символе, многолучевые сигналы, соответствующие предыдущему и последующему символам, частично находятся в линии задержки (см. рис. 8.10, b), вероятность того, что их пики будут точно соответствовать активированным отводам линии, весьма мала. Напомним, что отводы у линии задержки сделаны через каждые $1/W$ с. Так как $TW \gg 1$ и так как мы теперь полагаем $\Delta T > 1$, то $W\Delta \gg 1$ и это означает, что линия задержки должна иметь большое число отводов. Например, при $TW = 100$ и $\Delta T = 5$ линия задержки будет иметь $W\Delta = 500$ отводов.

Однако при приеме сигнала активируются только те отводы, на которых ожидаются пики этого сигнала, поэтому в рассмотренном выше примере даже при наличии, скажем, 20 лучей будет активироваться всего $20/500 = 2,5\%$ от общего числа отводов. Эти отводы будут соответствовать пикам центрального n -го отклика на рис. 8.10, b); при этом весьма маловероятно, что пики смежных с ним откликов также попадут на эти активированные отводы.

Последовательность импульсов после семи таких сверток и соответствующего их объединения будет иметь примерно такой вид, как показано на рис. 8.11.

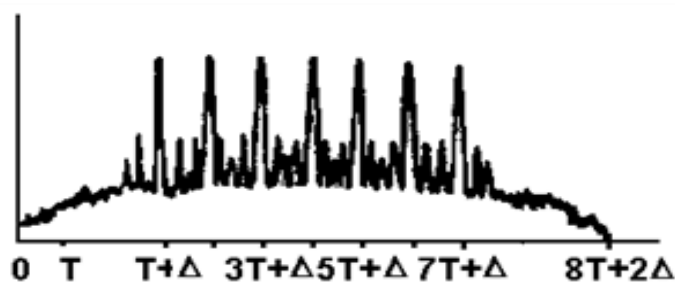


Рис. 8.11. Реакция ТФ приемника «Рейк» на 7 сигналов рис. 3.14.

Каждый из главных пиков соответствует здесь точному согласованию одной последовательности импульсов с активированными отводами линии задержки, и именно по ним берутся далее отсчеты. Пьедестал, на котором расположены эти пики, состоит из таких же небольших побочных пиков, как и на

рис. 8.8, и смеси боковых лепестков. Этот пьедестал мы будем называть «помехой, обусловленной многолучевостью». Способность приемника системы «Рейк» выделять именно ту часть отклика согласованного фильтра, которая соответствует принимаемому в данный момент символу, обусловлена структурой применяемой в нем линии задержки с отводами.

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

Рекомендуемая литература

1. Скляр Б. Цифровая связь. Теоретические основы и практическое применение. [Текст]. Изд. 2-е, испр.: Пер. с англ. – М.: Издательский дом «Вильямс», 2003. – 1104 с.
2. Ипатов В. Широкополосные системы и кодовое разделение сигналов. Принципы и приложения. [Текст]. Москва: Техносфера, 2007. – 488 с.
3. Прокис Дж. Цифровая связь. [Текст]. Пер. с англ. / Под ред. Д.Д. Кловского – М.: Радио и связь, 2000. – 800 с.
4. Пенин П.И. Системы передачи цифровой информации: Учебное пособие для вузов. [Текст]. -М.: Сов. радио, 1983. – 368с.
5. Турин Дж.Л. Введение в широкополосные методы борьбы с многолучевостью распространения радиосигналов и их применение в городских системах цифровой связи. ТИИЭР. [Текст]. Т. 68, № 2, март 1980. С. 30 – 60.