

Spread Spectrum and CDMA

Principles and Applications

Valery P. Ipatov

University of Turku
and
Saint Petersburg Electrotechnical University “LETI”

WILEY

СОДЕРЖАНИЕ

Список сокращений.	7
Список обозначений.	9
1. Широкополосные сигналы и системы.	13
1.1. Основные определения.	13
1.2. Краткий исторический обзор.	16
2. Классические задачи приема и синтез сигналов.	19
2.1. Гауссовский канал, общая задача приема, оптимальные решающие правила.	19
2.2. Передача двоичных данных (детерминированные сигналы).	23
2.3. Передача M-ичных данных (детерминированные сигналы).	29
2.4. Комплексная огибающая полосного сигнала.	35
2.5. M-ичная передача данных. Некогерентные сигналы.	39
2.6. Обмен между выигрышем от ортогонального кодирования и шириной полосы.	41
2.7. Примеры множеств ортогональных сигналов.	44
2.7.1. Кодирование путем временного сдвига.	44
2.7.2. Кодирование путем частотного сдвига.	45
2.7.3. Ортогональное кодирование с распределением спектра.	46
2.8. Оценивание параметров сигнала.	50
2.8.1. Формулировка задачи и правила оценивания.	50
2.8.2. Точность оценивания.	52
2.9. Оценивание амплитуды сигнала.	54
2.10. Оценка фазы.	55
2.11. Автокорреляционная функция и отклик согласованного фильтра.	56
2.12. Оценка временной задержки полосного сигнала.	59
2.12.1. Алгоритм оценивания.	59
2.12.2. Точность оценивания.	62
2.13. Оценка несущей частоты.	66
2.14. Одновременное оценивание времени запаздывания и частоты.	69
2.15. Разрешение сигналов.	72
2.16. Заключение.	76
Задачи.	77
3. Достоинства технологии распределенного спектра.	84
3.1. Иммуниет к воздействию помех.	84
3.1.1. Узкополосная помеха.	84
3.1.2. Заградительная помеха.	86
3.2. Низкая вероятность обнаружения.	88
3.3. Структурная скрытность сигнала.	95
3.4. Электромагнитная совместимость.	96
3.5. Эффекты распространения в беспроводных системах.	97
3.5.1. Распространение в свободном пространстве.	97
3.5.2. Затенение.	98
3.5.3. Замирания вследствие многолучевого распространения.	98
3.5.4. Анализ характеристик.	102
3.6. Разнесение.	106
3.6.1. Методы комбинирования.	106
3.6.2. Организация ветвей разнесения.	107
3.7. Многолучевое разнесение и РАКЕ-приемники.	110
Задачи.	115
4. Многопользовательская среда. Множественный доступ с кодовым разделением.	119
4.1. Многоабонентские системы и проблема множественного доступа.	119

4.2. Множественный доступ с частотным разделением.	121
4.3. Множественный доступ с временным разделением.	122
4.4. Синхронный вариант множественного доступа с кодовым разделением.	123
4.5. Асинхронный метод с CDMA.	125
4.6. Асинхронный вариант CDMA в сотовых сетях.	129
4.6.1. Проблема повторного использования ресурса и сотовые системы.	129
4.6.2. Число пользователей на соту в асинхронном варианте CDMA.	130
Задачи.	135
5. Дискретные широкополосные сигналы.	137
5.1. Широкополосная модуляция.	137
5.2. Обобщенная модель и категории дискретных сигналов.	137
5.3. Корреляционные функции АФМ сигналов.	139
5.4. Вычисление корреляционных функций кодовых последовательностей.	141
5.5. Корреляционные функции ЧМ сигналов.	143
5.6. Выигрыш от обработки дискретных сигналов.	145
Задачи.	147
6. Широкополосные сигналы для измерения времени, синхронизации и разрешения во времени.	148
6.1. Дополнительные требования, предъявляемые к АКФ.	148
6.2. Сигналы с непрерывной частотной модуляцией.	149
6.3. Критерий выбора АФМ сигналов с хорошей апериодической АКФ.	152
6.4. Об оптимизации апериодических ФМ сигналов.	153
6.5. Идеальная периодическая АКФ. Бинарные минимаксные последовательности	158
6.6. Введение в теорию конечных полей и линейных последовательностей.	161
6.6.1. Определение конечного поля.	161
6.6.2. Линейные последовательности над конечными полями.	161
6.6.3. m -последовательности.	164
6.7. Периодическая АКФ m -последовательностей.	166
6.8. Дополнительные сведения о конечных полях.	170
6.9. Последовательности Лежандра.	172
6.10. Бинарные коды с хорошей апериодической АКФ.	174
6.11. Последовательности с идеальной периодической АКФ.	178
6.11.1. Бинарные последовательности с не противоположной модуляцией.	178
6.11.2. Многофазные коды.	180
6.11.3. Троичные последовательности.	181
6.12. Подавление боковых лепестков вдоль оси задержек.	187
6.12.1. Фильтр подавления боковых лепестков.	187
6.12.2. Вычисление потерь в отношении сигнал-шум.	188
6.13. ЧМ сигналы с оптимальной апериодической АКФ.	193
Задачи.	196
7. Ансамбли широкополосных сигнатур в CDMA приложениях.	199
7.1. Широкополосная передача данных.	199
7.1.1. Прямое расширение спектра: бинарная манипуляция данных и бинарные сигнатуры.	199
7.1.2. Прямое расширение: общий случай.	203
7.1.3. Расширение спектра прыгающей частотой.	208
7.2. Синтез ансамблей сигнатур для синхронного CDMA с прямым расширением спектра.	211
7.2.1. Постановка задачи.	211
7.2.2. Оптимизация множества сигнатур по критерию минимума расстояния.	212
7.2.3. Последовательности, удовлетворяющие границе Велча.	219

7.3. Подходы к синтезу ансамблей сигнатур для асинхронного CDMA с прямым расширением спектра.	223
7.4. Сигнатуры для асинхронного варианта CDMA, получаемые временным сдвигом.	228
7.5. Примеры ансамблей минимаксных сигнатур.	231
7.5.1. Бинарные частотно-сдвинутые m -последовательности.	231
7.5.2. Множества Голда.	232
7.5.3. Множества Касами и их расширения.	235
7.5.4. Ансамбли Камалетдинова.	237
Задачи.	240
8. Поиск и слежение за сигналом с прямым расширением спектра.	243
8.1. Процедуры поиска и слежения.	243
8.2. Процедура последовательного поиска.	245
8.2.1. Модель алгоритма.	245
8.2.2. Вероятность правильного завершения поиска и среднее число шагов.	246
8.2.3. Минимизация среднего времени поиска.	250
8.3. Методы ускорения поиска.	254
8.3.1. Постановка задачи.	254
8.3.2. Последовательный просмотр ячеек.	254
8.3.3. Последовательно-параллельный поиск.	255
8.3.4. Последовательности быстрого поиска.	256
8.4. Слежение за кодом.	258
8.4.1. Оценка запаздывания в результате слежения.	258
8.4.2. Ранне-поздние дискриминаторы DLL.	259
8.4.3. Шумовые характеристики петли захвата по задержке (DLL).	262
Задачи.	266
9. Канальное кодирование в широкополосных системах.	267
9.1. Предварительные замечания и терминология.	267
9.2. Блочные коды, обнаруживающие ошибки.	269
9.2.1. Двоичные блочные коды и способность к обнаружению ошибок.	269
9.2.2. Линейные коды и их полиномиальное представление.	270
9.2.3. Вычисление синдрома и обнаружение ошибок.	273
9.2.4. Выбор порождающих полиномов для CRC.	274
9.3. Сверточные коды.	276
9.3.1. Сверточный кодер.	276
9.3.2. Решетчатая диаграмма, свободное расстояние и асимптотический выигрыш от кодирования.	279
9.3.3. Алгоритм декодирования Витерби.	282
9.3.4. Приложения.	286
9.4. Турбо-коды.	287
9.4.1. Кодер турбо-кодов.	287
9.4.2. Процедура итеративного декодирования.	289
9.4.3. Характеристики.	291
9.4.4. Приложения.	292
9.5. Канальное перемежение.	293
Задачи.	294
10. Некоторые достижения в развитии широкополосных систем.	297
10.1. Многопользовательский прием и подавление помех множественного доступа.	297
10.1.1. Оптимальное (МП) правило многопользовательского приема для синхронного варианта CDMA.	297
10.1.2. Алгоритм декорреляции.	298

10.1.3. Обнаружение по минимуму среднеквадратической ошибки.	300
10.1.4. Слепой МСКО обнаружитель.	303
10.1.5. Устранение помех.	305
10.1.6. Асинхронные многопользовательские обнаружители.	305
10.2. Модуляция со многими несущими и OFDM.	307
10.2.1. Метод CDMA с прямым расширением спектра и многими несущими.	307
10.2.2. Стандартная МС передача и OFDM.	309
10.2.3. Метод CDMA со многими несущими.	313
10.2.4. Применение.	316
10.3. Разнесение на передаче и пространственно-временное кодирование в CDMA системах.	317
10.3.1. Разнесение на передаче и задача пространственно-временного кодирования.	317
10.3.2. Эффективность разнесения на передаче.	318
10.3.3. Пространственно-временной код с переключением во времени.	320
10.3.4. Пространственно-временные коды Аламоути.	322
10.3.5. Разнесение на передаче в широкополосных приложениях.	325
Задачи.	327
11. Примеры действующих беспроводных широкополосных систем.	329
11.1. Предварительные замечания.	329
11.2. Система глобального позиционирования.	329
11.2.1. Основные принципы системы и архитектура.	329
11.2.2. Дальномерные сигналы системы GPS.	331
11.2.3. Обработка сигналов.	333
11.2.4. Точность местоопределения.	333
11.2.5. Система ГЛОНАСС и ГНСС.	334
11.2.6. Применение.	334
11.3. Эфирный интерфейс систем мобильной радиосвязи стандартов cdmaOne (IS-95) и cdma2000.	336
11.3.1. Вводные замечания.	336
11.3.2. Коды расширения в стандарте IS-95.	336
11.3.3. Каналы прямой линии связи стандарта IS-95.	337
11.3.3.1. Пилотный канал.	337
11.3.3.2. Канал синхронизации.	338
11.3.3.3. Каналы вызова.	339
11.3.3.4. Каналы трафика.	340
11.3.3.5. Модуляция в прямом канале.	342
11.3.3.6. Обработка МС сигнала прямого канала.	343
11.3.4. Обратный канал стандарта IS-95.	343
11.3.4.1. Канал связи обратного трафика.	343
11.3.4.2. Канал доступа.	346
11.3.4.3. Модуляция в обратном канале.	346
11.3.5. Эволюция эфирного интерфейса от cdmaOne к cdma2000.	347
11.4. Эфирный интерфейс мобильной радиосвязи стандарта UMTS.	349
11.4.1. Предварительные замечания.	349
11.4.2. Типы каналов стандарта UMTS.	349
11.4.3. Выделенные физические каналы «вверх».	350
11.4.4. Общие физические каналы «вверх».	351
11.4.5. Канализирующие коды линии «вверх».	352
11.4.6. Скремблирование линии «вверх».	353
11.4.7. Отображение транспортных каналов «вниз» на физические каналы.	355
11.4.8. Формат физических каналов линии «вниз».	355

11.4.9. Канализирующие коды линии «вниз».	356
11.4.10. Скремблирующие коды линии «вниз».	357
11.4.11. Канал синхронизации.	357
11.4.11.1. Общая структура.	357
11.4.11.2. Первичный код синхронизации.	358
11.4.11.3. Вторичный код синхронизации.	358
Список литературы	360

1. Широкополосные сигналы и системы

1.1. Основные определения

В настоящее время термин «*spread spectrum*» (*широкополосный, распределенный спектр*) является одним из наиболее употребляемых в радиотехническом и телекоммуникационном сообществе. В то же время, оказывается не так просто сформулировать непротиворечивое и точное определение, ясно отделяющее широкополосную философию от «не широкополосной». Конечно, любой эксперт в области проектирования систем или опытный исследователь интуитивно чувствует суть проблемы, однако это не относится ко вновь обратившимся к данной области. Перед ними не ставится задача выработки определения, а только существует потребность дать ответ на стоящую перед ним проблему. С точки зрения цели, которой служит данная книга, представляется полезным посвятить некоторое ее пространство выработке соответствующего объяснения тому, что в дальнейшем тексте понимается под широкополосной концепцией.

Начнем с напомним основы спектрального анализа. Любой сигнал $s(t)$ конечной энергии может быть представлен как сумма несчетного числа гармонических колебаний, амплитуды и фазы которых в пределах бесконечно малого диапазона частот $[f, f + df]$ определяются спектральной плотностью или спектром $\tilde{s}(f)$. Математическим отображением этого факта служит пара обратного и прямого преобразования Фурье

$$s(t) = \int_{-\infty}^{\infty} \tilde{s}(f) \exp(j2\pi ft) df, \quad \tilde{s}(f) = \int_{-\infty}^{\infty} s(t) \exp(-j2\pi ft) dt. \quad (1.1)$$

Благодаря взаимно однозначному соответствию между представлением сигнала во временной $s(t)$ и частотной $\tilde{s}(f)$ области мы вправе произвольно определять то представление, которое будет использоваться при решении конкретной задачи. Для характеристики размера зоны, в которой сосредоточена энергия сигнала во временной или частотной области, используются обозначения длительности сигнала T или полосы W соответственно. Сигнал, энергия которого концентрируется в пределах строго ограниченного участка временной оси, не может иметь ограниченный (т.е. ненулевой в конечном частотном интервале) спектр и наоборот. Вследствие этого для определения, по крайней мере, одного из параметров T , W или обоих сразу необходимо принять ряд допущений, касающихся понимания значения длительности и полосы. Этот путь, основанный на понятии эффективного, среднеквадратического и т.п. значения, определяет длительность и полосу как размер зоны, в которой сосредоточена существенная часть энергии сигнала во временной или частотной области [1].

Абсолютно ясно, что тем или иным образом слово «spread» (распределенный, расширенный) указывает на ширину спектра, т.е. большую полосу W сигнала. Однако опять возникает вопрос, что понимать под широким спектром? Что может служить эталоном для сравнения? Для демонстрации того, как определение «широкополосный» вызывает неоднозначное понимание, сошлемся на ряд широко известных книг.

Наиболее часто при объяснении широкополосности встречается положение, утверждающее, что система или сигнал является широкополосными, если занимаемая ими полоса значительно превосходит минимальную полосу, необходимую для передачи информации [1–6]. Наиболее спорным в этом определении является положение о минимальной полосе, необходимой для передачи информации или сообщения. Согласно фундаментальной границе Шеннона спектральная эффективность (т.е. отношение скорости передачи данных R к полосе сигнала W) системы связи, работающей в условиях гауссовского канала, удовлетворяет неравенству

$$\frac{R}{W} < \log_2 \left(1 + \frac{E_b}{N_0} \frac{R}{W} \right), \text{ или } \frac{E_b}{N_0} > \frac{2^{\frac{R}{W}} - 1}{\frac{R}{W}} \quad (1.2)$$

где E_b – энергия сигнала, приходящаяся на один бит информации, N_0 – односторонняя спектральная плотность мощности гауссовского шума. Графическое изображение границы (1.2) представлено на рис.1.1, который показывает, что любые комбинации R/W и

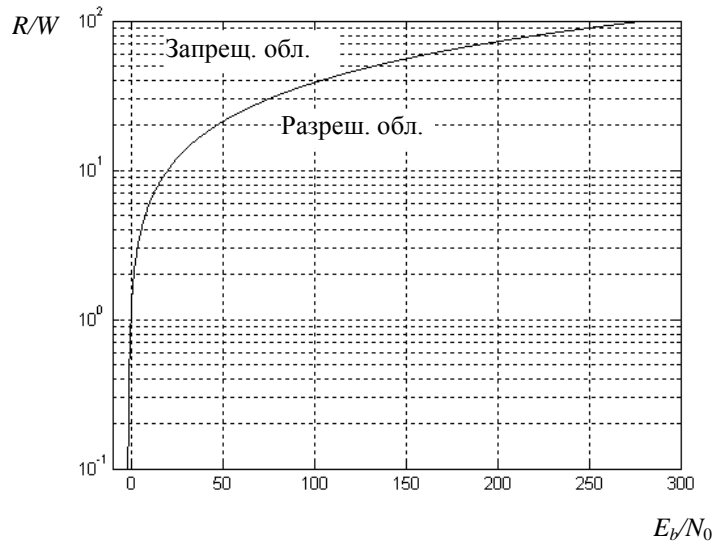


Рис 1.1. Граница Шеннона.

E_b/N_0 , лежащие ниже кривой, возможны, по крайней мере, в принципе. Последнее означает, что теоретический «минимум полосы, необходимой для передачи информации» равен нулю и, следовательно, любая реальная система, которая занимает некоторую ненулевую полосу, может трактоваться как широкополосная. Несомненно, любая попытка использования близкой к нулю полосы для передачи данных потребует значительной энергии сигнала. Так, например, работа со скоростью $R=100W$ требует отношения сигнал-шум на бит E_b/N_0 , равного 280 дБ, что представляется совершенно не реалистичным. Однако, передача данных в полосе, в десять раз меньшей скорости передачи данных, является типичной для многих цифровых линий связи (радиорелейных линий, модемной связи и т.п.). Данные примеры демонстрируют неопределенность (или бесполезность) использования понятия «минимальной полосы» и доказывают (аргументируют) возможность его использования только в качестве начального пункта при объяснении понятия распределенного спектра, или широкополосности.

В качестве попытки, исключающей указанную двусмысленность, может рассматриваться вариант использования скорости передачи данных, выраженную в битах за секунду, как замену упомянутой выше минимально необходимой полосы [7-8]. Однако, представляется не совсем логичным, что какой-то одной из многих возможных и равноправных единиц измерения скорости передачи данных присваивается особая концептуальная роль. Кроме того, определение широкополосности в терминах ширины полосы, значительно превосходящей скорость передачи данных в битах за секунду, несет определенный риск при охвате систем, которые не относятся к широкополосным. Подтверждением последнего может служить канал связи «вверх» между абонентом и базовой станцией в системе мобильной телефонии GSM. При основной скорости передачи оцифрованной речи в 9,6 кбит/сек абонентский сигнал занимает полосу порядка 200 кГц, что может при-

вести к ошибочному решению классифицировать GSM систему, как широкополосную. Однако расширение полосы обратного канала связи GSM системы не связано с особенностями широкополосности. Единственной причиной значительного превышения ширины полосы над скоростью передачи данных является использование метода множественного доступа с временным разделением, который обуславливает работу с укороченными излучаемыми символами по сравнению с временным интервалом, отводимым одному биту информации.

Существует еще одна причина для поиска альтернативных определений. Даже если игнорировать обсужденные ранее трудности, привязывание определения широкополосности к скорости передачи данных или полосе, отводимой для передачи сообщений, может быть оправдано только для систем передачи данных. Однако широкополосная философия находит значительное применение во многих других приложениях, таких как радиолокация, гидролокация, навигация (иначе удаленный контроль за измерением времени и расстояния), разрешения сигналов и пр. Фактически эти системы находились в первых рядах комплексов, реализовавших преимущества, предоставляемые обсуждаемой технологией. В этих приложениях использование таких категорий, как «информационная скорость» или «полоса для передачи данных», представляется затруднительным или, по крайней мере, никак не связанным со стремлением к использованию широкополосности. С целью придания идеям широкополосности универсальности, т.е. возможности применения не только к связным задачам, но и в других областях приложений, следующее определение распределенности спектра (или широкополосности) представляется наиболее уместным.

Обратимся первоначально к принципу неопределенности Габора, согласно которому произведение длительности сигнала на занимаемую им полосу (*частотно-временное произведение – time-frequency product*) удовлетворяет неравенству $WT \geq a$, в котором константа a зависит от способа определения длительности и ширины полосы, однако в большинстве случаев ее значение равно единице. Сигнал, для которого $WT \approx 1$ и, значит, длительность и полоса которого тесно связаны с друг другом, может быть назван *простым – plain* (не с распределенным спектром или не широкополосным). Единственным способом увеличения полосы, занимаемой простым сигналом, служит уменьшение его длительности, т.е. его укорачивание. С другой стороны, детерминированный сигнал, для которого $WT \gg 1$ и полоса которого может изменяться независимо от длительности, называется сигналом с *расширенным спектром* или *широкополосным (spread spectrum)*. Другими словами, энергия любого широкополосного сигнала распределена в прямоугольнике на плоскости время–частота, площадь которого значительно больше единицы. Данное определение автоматически определяет также и широкополосную систему: система, основанная на использовании сигналов с распределенным спектром, является широкополосной системой.

Отметим, что в данном определении независимость длительности и полосы выражена явным образом, что означает возможность увеличения ширины полосы (длительности) без уменьшения сигнала во времени (по частоте). Она находит свое отражение в различных способах модуляции (по углу, фазе или частоте) при любых технологиях распределения спектра. Действительно, как амплитудная модуляция может способствовать расширению спектра? Ответ заключается в следующем: только уменьшением зоны, в которой энергия сигнала эффективно распределена во времени, т.е. путем реального уменьшения эффективной длительности сигнала. Таким образом, только угловая модуляция способна расширить спектр сигнала, не оказывая влияния на распределение энергии сигнала во времени.

Иллюстрацией вышесказанному служит рис. 1.2, на котором представлены два прямоугольных импульса одинаковой длительности T с заполнением на несущей частоте f_0 : сигнал без внутренней модуляции (a) и сигнал с линейной частотной модуляцией с девиацией $W_d = 20/T$ (b). Нижние кривые соответственно иллюстрируют вид спектра этих сигналов. Как видно из рисунка, полоса W , занимаемая сигналом, определяется как

$W \approx 1/T$, т.е. энергия сигнала распределена в частотной области на интервале, примерно равном обратной величине длительности импульса. Следовательно, длительность и поло-

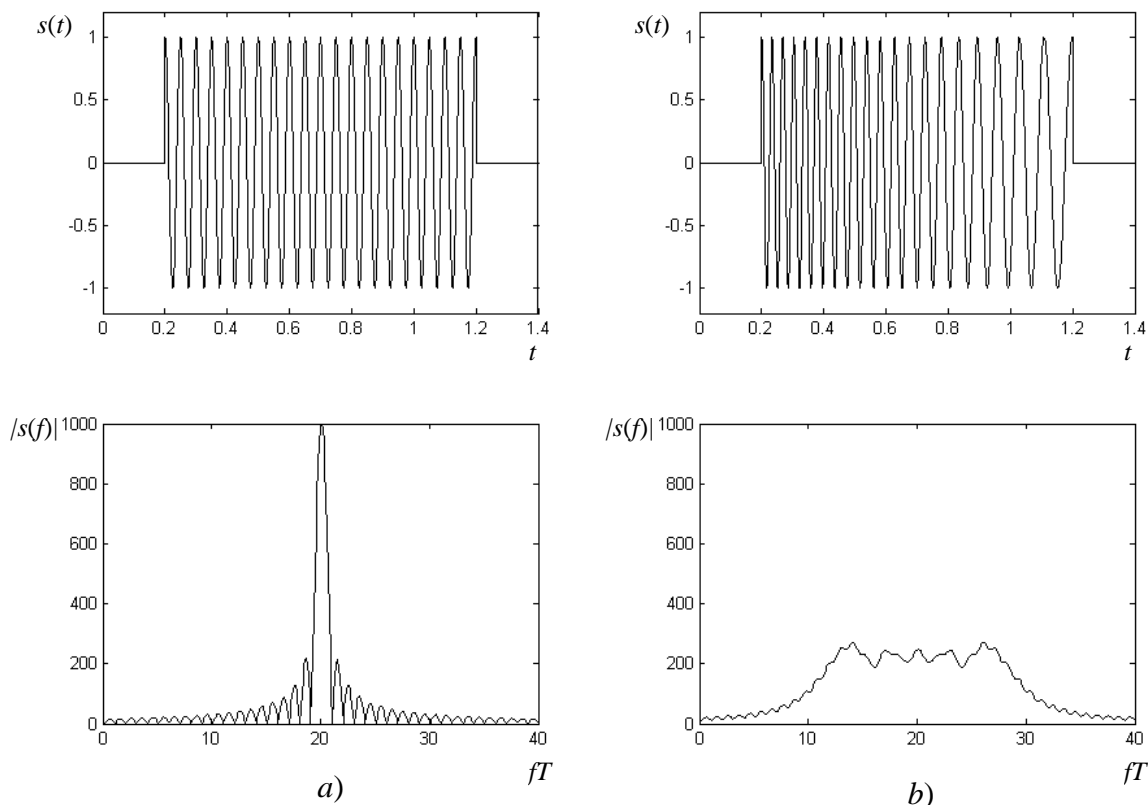


Рис. 1.2 Немодулированный (а) и частотно–модулированный (b) прямоугольные импульсы и их спектр.

са тесно связаны, а частотно–временное произведение фиксировано и, значит, расширение спектра может быть достигнуто только в обмен на укорочение импульса. В свою очередь полоса, занимаемая импульсом (b), близка к значению девиации ($W \approx W_d$) и значительно больше, чем величина, обратная длительности. В результате независимо от длительности сигнала полоса легко регулируется только изменением девиации. В соответствие с ранее введенным определением первый сигнал относится к простым, а второй – к широкополосным сигналам.

Введенное ранее определение широкополосности фактически является первым, которое широко используется при анализе радиолокационных и родственных им системам, однако оно хорошо согласуется и с задачами, связанными с передачей информации. В связи с чем, в дальнейшем изложении основной упор будет сделан именно на нем.

1.2. Краткий исторический обзор

История технологии распределенного спектра (широкополосности) охватывает шесть десятилетий и может служить предметом собственного изучения. Читатель, заинтересованный в хронологии ключевых событий может обратиться к обширным (однако почти полностью посвященным развитию данного направления в США) обзорам в [9, 10]. Здесь же ограничимся лишь кратким упоминанием основных исторических вех.

Вероятно, что первый патент на РЛС, которую в современной терминологии без сомнения можно отнести к широкополосной, был получен G. Guanella в 1938 г. Далее, в течение Второй Мировой войны и после нее интенсивные исследования в области разра-

ботки широкополосных РЛС были предприняты в Германии, США, Великобритании и СССР. Параллельно с технологическими и техническими успехами были осуществлены многочисленные и основательные теоретические исследования, касающиеся точности радиолокационных измерений и разрешения сигналов. Наиболее важные и глубокие результаты в этом направлении были опубликованы в книге Р.М. Woodward в 1953 г. Следует заметить, что многие из упомянутых результатов объяснимо базируются на фундаментальных работах С. Shannon и В.А. Котельникова за 1946–48 гг, роль которых сказывается значительно шире, чем только в их «чисто» коммуникационном приложении.

Конечно, в течение длительного времени значительная доля информации о новых практических разработках в области широкополосной радиолокации и навигации была трудно доступна, поскольку большинство этих проектов курировались военными и разведывательными органами. Тем не менее, многие идеи, как только они материализовались в общедоступных системах, стали широко известны. Ярким примером подобного рода может служить всемирно известная радионавигационная система «Лоран-С», введенная в эксплуатацию в начале 60-х годов. В данной системе наземные длинноволновые радиомаяки излучают истинно широкополосные (фазоманипулированные) сигналы с частотно-временным произведением $WT=16$. Свидетельством жизнеспособности этой системы служит тот факт, что с учетом постоянной модернизации и многочисленных улучшений ей удалось продлить свое функционирование до третьего тысячелетия.

Другой значительный шаг в практическом внедрении широкополосной концепции в области разработки дальномерно-временных измерительных систем был сделан в конце 80-х – начале 90-х годов в результате создания спутниковых радионавигационных систем второго поколения типа GPS NAVSTAR (США) и ГЛОНАСС (СССР/Россия). Сигналы с очень большим значением частотно-временного произведения, измеряемого в тысячах, являются сердцевиной упомянутых систем, которые в настоящее время превратились в составную часть человеческой цивилизации наряду со спутниковым телевидением и мобильной связью.

Ранние работы по использованию широкополосной идеологии в области передачи информации были первоначально стимулированы задачами маскирования речевых данных и защиты передаваемой информации. Начало им было положено в Германии перед Второй Мировой войной и вскоре было подхвачено США, СССР и другими заинтересованными странами. Одно из интригующих действий романа А. Солженицына «В круге первом» происходит в специальной советской тюрьме, где были собраны осужденные студенты и инженеры для разработки советской скрытной шумоподобной системы передачи речевой информации.

Среди поворотных моментов в развитии широкополосной связи следует отметить Rake-алгоритм, предложенный R. Price и P. Green (1957), который положил начало направлению, названному позднее многолучевым разнесением. Работы 60-х годов S. Golomb, N. Zierler, R. Gold, T. Kasami и других в области синтеза дискретных последовательностей со специальными корреляционными свойствами сыграли ключевую роль в формировании широкополосной технологии и многочисленных практических достижений.

Коммерческая эра широкополосной идеологии стартовала в конце 70-х годов, когда мобильная телефонная связь начала свое триумфальное завоевание всего мира. Первые предложения по сотовым сетям с CDMA в США и Европе (1978-1980) породили альтернативные проекты, которые эволюционировали позднее в стандарты GSM и DAMPS. Однако в середине 90-х годов стандарт второго поколения IS-95 полностью освоил SS-CDMA платформу. С космической скоростью сети этого стандарта (позднее названного cdmaOne) завоевали широкое признание в Америке, Азии и бывшем Советском Союзе. Огромный успех стандарта IS-95, а также тщательный анализ и дальнейшие эксперименты привели к признанию SS-CDMA философии в качестве основы для большинства проектов мобильной радиосвязи 3-го поколения: UMTS и cdma2000. В настоящее время оба

эти проекта находятся на предэксплуатационном этапе и, несомненно, станут основным инструментом мобильной связи на ближайшие десятилетия.

В заключение вступительного раздела следует добавить несколько слов о развитии широкополосной технологии в Советском Союзе, а позднее России. Обзоры, публикуемые на Западе, обычно немного сообщают только об уровне советских достижений в данной области. Это положение объясняется объективными причинами, обусловленными периодом холодной войны: самоизоляция страны, строгие ограничения на контакты советских специалистов со своим зарубежными коллегами и их публикациями, чрезмерная и часто излишняя секретность и т.п. К числу серьезных препятствий следует отнести и языковой барьер. Однако, на самом деле советские достижения в области широкополосности в 50–90 гг. были своевременными и конкурентно способными в сравнении с аналогичными результатами, достигнутыми в США и Европе. Работы Д.Е. Вакмана, Д.Д. Ширмана, М.Б. Свердлика (в области синтеза широкополосных радиолокационных сигналов и их обработки), И.А. Амиантова, Л.Е. Варакина (в области широкополосной связи) были во многих отношениях пионерскими и рекрутировали в эту привлекательную и захватывающую область исследований поколения молодых профессионалов.

2. Классические задачи приема и синтез сигналов.

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

2.1. Гауссовский канал, общая задача приема, оптимальные решающие правила.

Любая информационная система, в которой данные передаются из одной точки пространства в другую, может быть представлена следующей абстрактной моделью. Пусть имеется некоторый источник сообщений, вырабатывающий в каждый конкретный момент времени одно из M возможных сообщений. Данный источник может быть как управляемым, или созданным человеком, так и иметь независимую от человеческой деятельности природу. В любом случае каждое из M конкурирующих сообщений передается посредством специфического сигнала, так что имеется множество S из M возможных сигналов: $S = \{s_i(t) : i = 1, 2, \dots, M\}$. На мощность множества S , т.е. число сигналов M , не накладывается никаких ограничений и, если это необходимо, множество S может быть бесконечным. Источник выбирает некоторый определенный сигнал $s_i(t) \in S$ и подает его на вход канала (см. рис. 2.1). На приемной стороне (на выходе канала) наблюдается при-



Рис. 2.1. Обобщенная модель системы.

нятое колебание $y(t)$, которое является не точной копией переданного сигнала $s_i(t)$, а результатом трансформации $s_i(t)$, обусловленной искажающим воздействием шумов и помех, присущих любому реальному каналу. Для приемной стороны имеется M конкурирующих гипотез H_i , состоящих в том, какой из M возможных сигналов был в действительности передан и трансформирован каналом в принятое наблюдение $y(t)$, и требуется выбрать только одну из них как истинную. Обозначим результат этого выбора, т.е. решение, через \hat{H}_j , означающее, что «решение принято в пользу сигнала с индексом j ». Из данной классической задачи приема вытекает следующая: *какова наилучшая стратегия принятия решения о возможном переданном сообщении (или сигнале), основанная лишь на наблюдении $y(t)$?*

Для ответа на поставленный вопрос необходимо знать модель канала. Математическое описание модели канала осуществляется с помощью *переходных вероятностей* (tran-

sition probability) $p[y(t)|s(t)]$, которые задают вероятность трансформации каналом заданного входного сигнала в то или иное выходное наблюдение $y(t)$. Полное задание канала предполагает знание переходных вероятностей $p[y(t)|s(t)]$ для всех возможных пар $s(t)$ и $y(t)$.

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

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

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

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

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

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

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

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

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

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

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

т.е. решение принимается в пользу сигнала $s_j(t)$, поскольку он наиболее близок (в смысле Евклидова расстояния) к наблюдению $y(t)$ среди всех M конкурирующих сигналов (см. рис.2.2). Другим, более наглядным представлением (2.3), является следующее

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

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

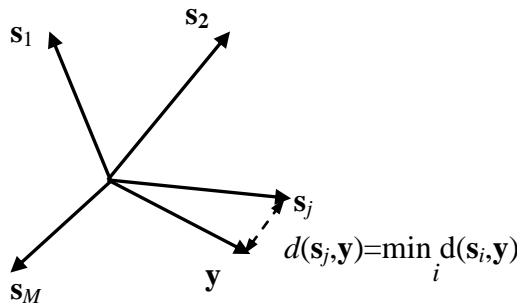


Рис. 2.2. Правило минимума расстояния.

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

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

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

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

которое снова может трактоваться как предельная форма скалярного произведения двух n -мерных векторов. Эта же характеристика может быть вычислена с помощью длины

векторов и косинуса угла α между ними: $(\mathbf{u}, \mathbf{v}) = \|\mathbf{u}\| \|\mathbf{v}\| \cos \alpha$, и, таким образом, скалярное произведение свидетельствует о близости, или *похожести* (*resemblance*) 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* (*correlation*) сигналов.

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

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

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

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

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

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

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

Интересно отметить, что даже эти предварительные рассуждения указывают уже более чем явный способ конструирования множества сигналов. Обратимся к рис.2.3, на

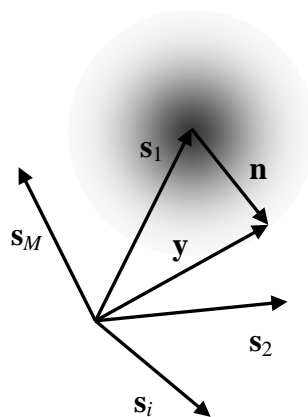


Рис. 2.3. Размытое наблюдение и задача выбора сигналов.

котором изображены сигнальные вектора. Предположим, что передавался сигнал s_1 , и что он подвергся искажению в АБГШ канале, следствием чего служит добавление к s_1 вектора шума \mathbf{n} . Гауссовский вектор \mathbf{n} характеризуется симметричным (сферическим) вероят-

ностным распределением, экспоненциально спадающим с увеличением длины вектора \mathbf{n} , что очевидным образом следует из (2.1) после удаления из него сигнала (т.е. при подстановки $s(t) = 0$). Следовательно, вектор наблюдения $\mathbf{y} = \mathbf{s}_1 + \mathbf{n}$ будет случайным образом перемещаться вокруг \mathbf{s}_1 , как это показано на рисунке, и тогда, согласно правилу минимума расстояния (2.3), как только \mathbf{y} окажется ближе к некоторому другому, чем \mathbf{s}_1 , сигналу, то будет принято ошибочное решение. Для минимизации возникновения ошибки подобного рода следует располагать другие сигналы на максимально большом расстоянии от \mathbf{s}_1 . Поскольку любой из M сигналов может передаваться равновероятно, т.е. занимать место \mathbf{s}_1 , то, очевидно, что все расстояния $d(\mathbf{s}_i, \mathbf{s}_j), 1 \leq i < j \leq M$, следует делать максимально большим. В случае достаточно большого значения M задача одновременной максимизации всех расстояний не представляется простой, поскольку они могут конфликтовать друг с другом: отодвигая некоторый вектор от другого, возможно приближение первого к некоторому третьему. Учитывая последнее, задача построения множества максимально удаленных друг от друга сигналов (входящая в обширный класс т.н. задач *упаковки* – *packing*) во многих случаях оказывается достаточно сложной, и пока что не имеет общего решения.

Отметим, что в ранее рассмотренном материале все M сигналов считались полностью *детерминированными* (*deterministic*), т.е. полагалось, что на приемной стороне а priori 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, что найдет свое отражение в 2.5.

Освежив эти основные идеи оптимального приема, теперь можно перейти к рассмотрению специфических задач, уделяя основное внимание аспектам конструирования сигналов и анализируя перспективность использования распределенного спектра в различных классических задачах приема сигналов или ее отсутствие.

2.2. Передача двоичных данных (детерминированные сигналы).

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

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

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

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

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

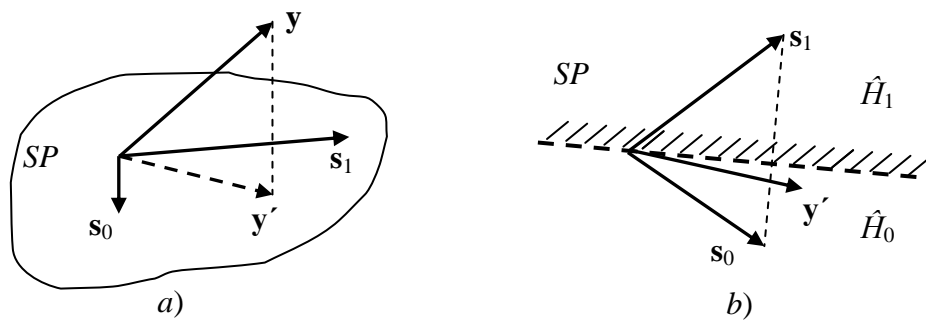


Рис. 2.4. Сигнальная плоскость и полуплоскости решения.

две полуплоскости прямолинейной границей, проходящей через начало координат и перпендикулярной к прямой линии, соединяющей сигнальные вектора. Тогда решения \hat{H}_0 и \hat{H}_1 фактически определяются как попадание проекции \mathbf{y}' на соответствующую полуплоскость (см. рис. 2.4, б). Из рис. 2.4, б также следует, что вероятность перепутывания сигналов (вероятность ошибки) зависит от расстояния между векторами \mathbf{s}_0 и \mathbf{s}_1 в сравнении с диапазоном случайных флуктуаций \mathbf{y}' , обусловленных канальным шумом. Согласно (2.10), фактически принятый сигнал $s_0(t)$ будет ошибочно засчитан за $s_1(t)$, если и только если разность корреляций будет меньше порога $(E_0 - E_1)/2$. Следовательно, вероятность p_{01} подобной ошибки может быть найдена как

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

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

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

интегрирование которого, согласно (2.11), приводит к выражению

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

где

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

– дополнительная функция ошибок (*complementary error function*).

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

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

где

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

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

При нахождении дисперсии σ^2 учтем тот факт, что на нее не оказывает влияние детерминированный компонент наблюдения $y(t)$ (в рассматриваемом случае сигнал $s_0(t)$), поскольку шум в канале аддитивен. Фактически, тогда можно удалить сигнал из выражения для $y(t)$, полагая $y(t) = n(t)$, где $n(t)$ – белый шум с двусторонней спектральной плотностью мощности $N_0/2$. После приведенных рассуждений нахождение дисперсии корреляции (2.7) наблюдения $y(t)$ и произвольного сигнала $s(t)$ сводится к вычислению

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

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

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

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

Другими словами, любые два отсчета белого шума, не совпадающие во времени, являются

некоррелированными. Используя последний результат в выражении для дисперсии, а также *фильтрующее* (*sifting*) свойство дельта-функции

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

приходим к выражению

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

где E – энергия сигнала $s(t)$.

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

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

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

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

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

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

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

Из (2.18) очевидным образом вытекает, что единственным путем достижения высокой достоверности передачи данных является увеличение расстояния $d(\mathbf{s}_0, \mathbf{s}_1)$ между сигналами до максимально возможного значения. Ясно, что увеличение $d(\mathbf{s}_0, \mathbf{s}_1)$ может быть достигнуто за счет большей энергии сигналов, или длины соответствующих векторов, как это демонстрирует соотношение (2.16). Однако какова будет *оптимальная* пара сигналов, если ресурс прямого решения лимитирован, т.е. энергия сигналов ограничена заранее? Рассмотрим первоначально типичный случай сигналов равных энергий $E_0 = E_1 = E$, когда интенсивность сигналов не используется как индикатор передаваемого сообщения. Тогда правило принятия решения (2.10) сводится к сравнению значений корреляций z_0 и z_1 или, что эквивалентно, определению знака их разности

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

Очевидно, что для максимизации расстояния между двумя векторами фиксированной длины следует выбирать их *противоположными* (*antipodal*), как это показано на

рис. 2.5, а. Тогда угол между векторами s_0 и s_1 составит величину $\alpha = \pi$, а $\cos \alpha = \rho_{01} = -1$ и, значит, $d(s_0, s_1) = 2\sqrt{E}$, что преобразует соотношение (2.18) в следующее

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

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

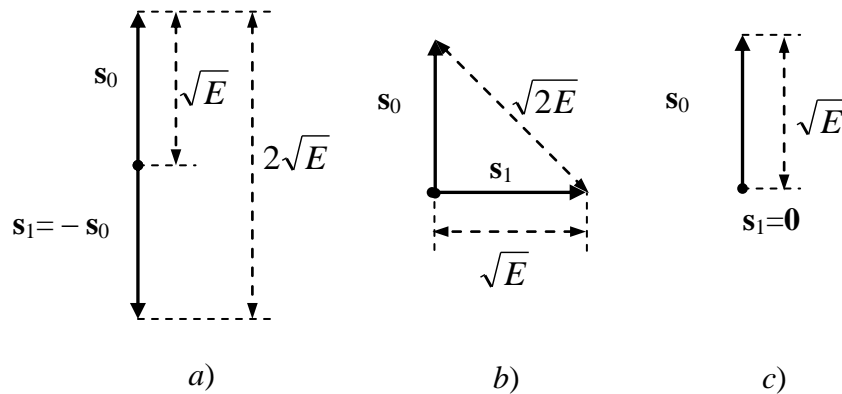


Рис. 2.5. Варианты выбора пар двоичных сигналов.

де упомянутых устройств.

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

Для определения критичности в выборе сигнальной пары сравним BPSK с другим популярным методом передачи двоичных данных. Хотя BPSK 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* (*frequency shift keying* (FSK)), при которой сообщения 0 и 1 передаются сигналами на различных частотах. Типичным является такой выбор несущих частот, при котором сигналы оказываются *ортогональными* (*orthogonal*), т.е. $\cos \alpha = \rho_{01} = 0$, $d(s_0, s_1) = \sqrt{2E}$ (см. рис. 2.5, b). Подстановка этих значений в (2.18) дает

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

Сравнение результатов (2.20) и (2.19) демонстрирует, что для ортогональной пары (FSK) обеспечение вероятности ошибочного приема, аналогичной случаю использования проти-

в противоположных сигналов (BPSK), достигается только при двукратном увеличении энергии сигналов. Иными словами, ортогональные сигналы проигрывают 3 дБ противоположным в уровне необходимой энергии.

Существует еще один, достаточно старый, способ передачи двоичной информации, до сих пор используемый на практике: *амплитудная модуляция (amplitude shift keying (ASK))*, при которой символ данных «1» передается сигналом $s_1(t) = s(t)$ с энергией $E_1 = E$, а символ «0» – паузой, т.е. $s_0(t) = 0, E_0 = 0$. В этом случае (см. рис. 2.5, с) $d(\mathbf{s}_0, \mathbf{s}_1) = \sqrt{E}$, а вероятность ошибки (2.18) становится равной

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

Сравнивая последний результат с (2.19), можно сделать вывод о том, что ASK требует в 4 раза (6 дБ) большей энергии, чем в случае BPSK, 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 ASK 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 ASK проигрывает BPSK в средней энергии только в два раза. Значит, для ASK по отношению к BPSK характерны такие же энергетические потери, как и в случае применения FSK, т.е. 3 дБ.

На основании предыдущих рассуждений можно сделать следующий вывод о задаче выбора пары сигналов для передачи двоичной информации: *отсутствуют малейшие намеки на получение каких-либо преимуществ при использовании сигналов с распределенным спектром*, поскольку расширение полосы сигнала сверх ее минимума $1/T$ не приведет к уменьшению вероятности ошибки. Действительно, для обеспечения желаемой достоверности приема достаточно только применить пару сигналов, максимально удаленных друг от друга, что автоматически предполагает использование противоположных сигналов без дополнительных требований к их форме и модуляции. Если по каким-либо причинам использование пары противоположных сигналов отвергается, то альтернативный вариант выбора сигналов – ортогональная, иначе частотно-модулированная, или амплитудно-модулированная пара сигналов также не предполагает применение технологии распределенного спектра.

2.3. Передача M-ичных данных (детерминированные сигналы).

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

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

Определение точной оценки этой вероятности состоит в интегрировании условной обобщенной плотности вероятности всех M корреляций при условии истинности принятия сигнала $s_1(t)$ по всей области выполнения неравенства $z_1 \geq z_i - \frac{E_i - E_1}{2}$, $i = 1, 2, \dots, M$.

Этот M -кратный интеграл в общем случае, т.е. в отсутствии предположений о свойствах множества сигналов, никаким образом не может быть упрощен. Однако очень продуктивная и точная верхняя граница для вероятности $p_{1,e}$ может быть получена на основании аддитивной (*union*) границы. Пусть событие A_j состоит в том, что наблюдение $y(t)$ оказывается наиболее близким не к $s_1(t)$, а к некоторому другому сигналу $s_j(t)$, $j = 2, 3, \dots, M$. Тогда событие, заключающееся в перепутывании сигнала $s_1(t)$ с некоторым другим сигналом, войдет в объединение всех A_j . Согласно аддитивной границы вероятность объединения событий никогда не может быть больше суммы их вероятностей, т.е.

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

С другой стороны, вероятность $P(A_j)$, как следует из определения события A_j , в точности совпадает с вероятностью перепутывания только между двумя сигналами $s_1(t)$ и $s_j(t)$. Значение этой вероятности определяется соотношением (2.17) после соответствующей замены номеров сигналов:

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

Подстановка последнего соотношения в ранее приведенное неравенство приводит к желаемой оценке

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

Аналогичный результат (с необходимой подстановкой новых числовых индексов) будет получен и в предположении, что вместо $s_1(t)$ фактически принятым считается сигнал $s_i(t)$, так что, с учетом а priori равной вероятности M сигналов, окончательная аддитивная верхняя граница для полной (безусловной) вероятности ошибки примет вид

$$P_e = \frac{1}{M} \sum_{i=1}^M \sum_{\substack{j=1 \\ j \neq i}}^M p_{i,e} \leq \frac{1}{M} \sum_{i=1}^M \sum_{\substack{j=1 \\ j \neq i}}^M Q\left(\sqrt{\frac{d^2(\mathbf{s}_i, \mathbf{s}_j)}{2N_0}}\right). \quad (2.22)$$

Первым заслуживающим внимание фактом, который непосредственно следует из

(2.22), является то, что при $M = 2$ данное неравенство превращается в равенство. Другое замечание связано с поведением (2.22) в асимптотике при росте отношения сигнал-шум. В самом деле, дополнительная функция ошибки $Q(x)$ уменьшается как $\exp(-x^2/2)$ при достаточно большом x и даже незначительное его увеличение может уменьшить $Q(x)$ до пренебрежимо малого уровня по сравнению с начальным значением. Вследствие этого при достаточно большом отношении сигнал-шум только близко расположенные сигнальные пары могут вносить заметный вклад в значение суммы в (2.22), и если d_{\min} есть минимальное расстояние на множестве всех возможных пар, встречающееся n_{\min} раз, то оценка (2.22) в асимптотическом случае трансформируется к виду

$$P_e \approx \frac{n_{\min}}{M} Q\left(\sqrt{\frac{d_{\min}^2}{2N_0}}\right), \quad \frac{d_{\min}^2}{2N_0} \gg 1. \quad (2.23)$$

Приближение (2.23) указывает, прежде всего, на асимптотическую сходимость аддитивной границы с истинным значением вероятности ошибки при увеличении отношения сигнал-шум. Для физического объяснения данного факта обратимся вновь к рис. 2.3 и отметим, что при малом уровне шума только те сигнальные вектора, которые расположены вблизи от истинного, могут быть ошибочно приняты за искомый. Это означает, что асимптотически только сигнальные пары с расстоянием d_{\min} определяют истинное значение вероятности ошибки P_e (а не только ее верхней границы), что влечет за собой близость величины P_e и ее аддитивной границы.

Соотношение (2.23) содержит одну из возможных и наиболее важных формулировок задачи конструирования сигналов: *максимизацию минимума расстояния между M сигналами*. Как уже упоминалось в 2.1, подобная задача в геометрической интерпретации эквивалентна упаковке M векторов таким образом, чтобы ближайшая их пара находилась на максимально достижимом расстоянии, т.е. $d_{\min} = \max$. На созвездие сигналов (векторов) могут быть наложены различные ограничения. Прежде всего, следует предусмотреть определенные энергетические ограничения, учитывая практические мощностно-энергетические ресурсы. 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 (volume packing)*. 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 (spherical packing)*.

Другим типичным ограничением при синтезе сигналов является размерность n_s сигнального пространства, внутри которого и осуществляется их упаковка. Физическая сущность этого ограничения снова обусловлена практическим ресурсом, но теперь шириной частотной полосы. Для объяснения существующей связи рассмотрим первоначально видеосигналы и предположим, что общий (двумерный) частотно-временной ресурс, в котором могут располагаться все M сигналов, ограничен параметрами W_t и T_t соответственно. Первое из этих ограничений учитывает экономию полосы, тогда как второе отражает желание передавать данные с приемлемой скоростью $R = \log M / T_t$. Тогда, согласно теореме отсчетов, имеется около $W_t T_t$ независимых отсчетов, которые могут быть использованы при синтезе M сигналов, причем каждый из сигналов трактуется как вектор в пространстве размерности $n_s = W_t T_t$. Определенная осторожность в оценке числа незави-

симых отсчетов объясняется невозможностью концентрации энергии любого сигнала в пределах конечного и во времени и по частоте интервала. Однако в первом приближении этим фактом можно пренебречь.

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

$$s(t) = S(t) \cos[2\pi f_0 t + \gamma(t)], \quad (2.24)$$

где $S(t)$ – вещественная огибающая сигнала, отражающая закон модуляции амплитуды, $\gamma(t)$ описывает закон фазовой модуляции, а f_0 – несущая частота. Используя тригонометрическое преобразование для косинуса суммы углов, (2.24) можно представить в виде

$$s(t) = S_I(t) \cos 2\pi f_0 t - S_Q(t) \sin 2\pi f_0 t, \quad (2.25)$$

где $S_I(t) = S(t) \cos \gamma(t)$, а $S_Q(t) = S(t) \sin \gamma(t)$ – *квадратурные компоненты (quadrature components)* сигнала. Поскольку и $S(t)$, и $\gamma(t)$ являются видеосигналами, то данное утверждение справедливо и в отношении $S_I(t), S_Q(t)$. Последнее означает, что при переносе на несущую частоту любой радиосигнал исчерпывающе описывается двумя независимыми низкочастотными квадратурными компонентами. Следовательно, при синтезе полосных сигналов могут использоваться в два раза больше независимых координат (отсчетов), чем в случае видеосигналов при одинаковых значениях частотно-временного произведения и, значит, $n_s = 2W_f T_f$.

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

Простейшая версия этой задачи ($n_s = 1$) отвечает случаю амплитудной (ASK) модуляции (простейший вариант которой с $M = 2$ – бинарная модуляция – был упомянут в 2.2). Альтернативным наименованием ASK служит *импульсно-кодовая модуляция (pulse amplitude modulation (PAM))*. В этом случае все сигнальные точки располагаются на одной прямой линии, и при $M > 2$ речь может идти только о задаче «объемной» упаковки. Не составляет труда убедиться, что при заданном d_{\min} оптимальное созвездие, минимизирующее среднюю энергию, образуется в результате равномерного и симметричного расположения сигнальных точек с пространственным разнесением соседних на величину, в точности равной d_{\min} (см. рис. 2.6, а).

При $n_s = 2$ задача нахождения оптимального созвездия с объемной упаковкой становится более трудной и может даже привести к асимметричным моделям, тогда как сферическая упаковка осуществляется тривиальным образом и реализуется равномерным размещением M точек по окружности радиуса \sqrt{E} . Широко используемая в современной цифровой связи M -ичная *квадратурная амплитудная модуляция (quadrature amplitude modulation (QAM))* служит примером двумерного, симметричного объемно-упакованного созвездия, которое, не являясь оптимальным теоретически, практически с точки зрения аппаратной реализации (Рис. 2.6, б). С другой стороны, обычные M -ичные *фазоманипулированные (phase shift keying (PSK))* созвездия образованы равномерно расположенными на окружностях точками и оптимальны с точки зрения сферической упаковки (Рис. 2.6, в).

Задача оптимального расположения точек в пространстве большей размерности

($n_s > 2$) является чрезвычайно сложной и не имеет до сих пор общего математического

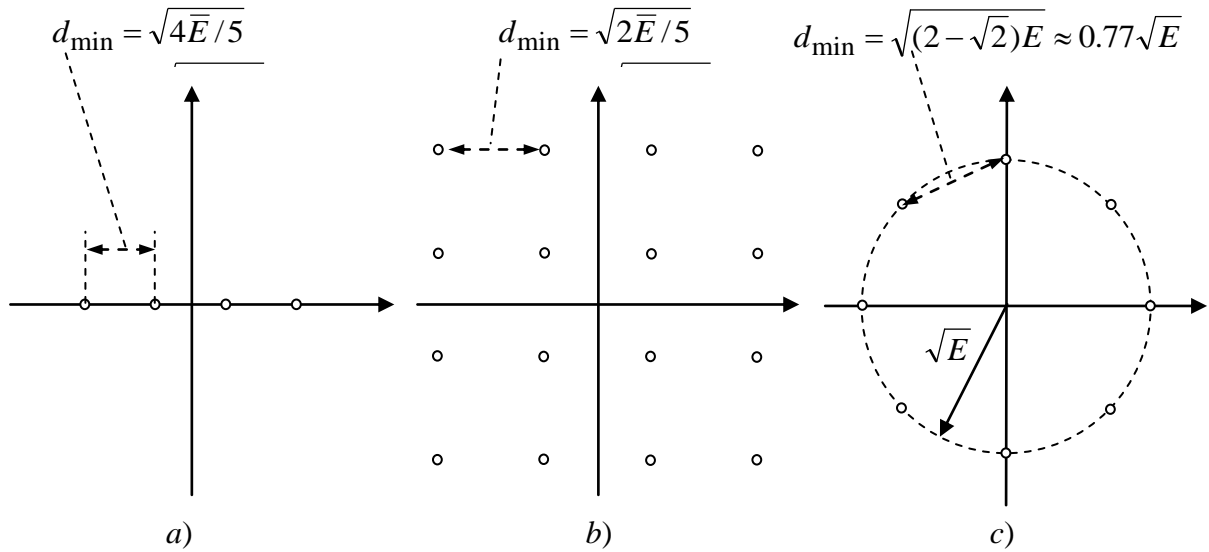


Рис. 2.6. Одно- и двумерные созвездия: 4-ASK (a), 16-QAM (b), и 8-PSK (c).

решения. Много полезных, но частных результатов содержится в целом ряде книг и статей (см., например, библиографию в [11] и веб-сайт [12]).

Попытаемся теперь определить верхний предел минимума расстояния в отсутствие предварительного ограничения на размерность сигнального пространства n_s , а также минимальное значение n_s , которое позволит обеспечить найденный предел. Ограничиваясь только случаем сферической упаковки ($E_i = E, i = 1, 2, \dots, M$), вычислим сумму всех M^2 возможных квадратов расстояний, включая и тривиальные (т.е. расстояния для любых сигналов с самим собой). Тогда на основании теоремы косинусов (2.16) получаем

$$\sum_{i, j=1}^M d^2(\mathbf{s}_i, \mathbf{s}_j) = 2M^2E - 2E \sum_{i, j=1}^M \rho_{ij}, \quad (2.26)$$

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

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

Поскольку интеграл от квадрата подынтегрального выражения никогда не будет отрицательным, из (2.26) вытекает, что

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

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

$$d_{\min}^2 \leq \frac{2M}{M-1} E. \quad (2.27)$$

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

$$(\mathbf{v}_i, \mathbf{v}_j) = (\mathbf{u}_i, \mathbf{u}_j) - \frac{1}{M}(\mathbf{u}_i, \mathbf{u}) - \frac{1}{M}(\mathbf{u}, \mathbf{u}_j) + \frac{1}{M^2}(\mathbf{u}, \mathbf{u}) = \delta_{ij} - \frac{1}{M}, \quad (2.28)$$

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

$$\mathbf{s}_i = \sqrt{\frac{ME}{M-1}} \mathbf{v}_i, i=1, 2, \dots, M. \quad (2.29)$$

Тогда, согласно теореме косинусов (2.16) и соотношению (2.29), квадрат расстояния между двумя сигналами будет

$$d^2(\mathbf{s}_i, \mathbf{s}_j) = \frac{ME}{M-1} d^2(\mathbf{v}_i, \mathbf{v}_j) = \frac{ME}{M-1} \left[\|\mathbf{v}_i\|^2 + \|\mathbf{v}_j\|^2 - 2(\mathbf{v}_i, \mathbf{v}_j) \right] = \frac{2ME}{M-1}, i \neq j, \quad (2.30)$$

что совпадает с правой частью соотношения (2.27). Следовательно, сигналы, удовлетворяющие границе (2.27) действительно существуют. Более того, расстояние между любой их парой также определяется этой величиной, поэтому приведенные сигналы попадают в категорию *эквидистантных* (*equidistant*). Данный класс сигналов широко известен под специальным наименованием *симплексных* (*simplex*) сигналов. Непосредственно из их определения вытекает, что

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

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

Свойство эквидистантности симплексных сигналов влечет за собой равенство коэффициентов корреляции ρ_{ij} любой их пары. Оценка величины ρ_{ij} с помощью (2.14), (2.29) и (2.28) приводит к следующему результату

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

демонстрирующему, что угол между любыми двумя симплексными сигналами один и тот

же и больше, чем $\pi/2$. Для простейших множеств из $M = 2, 3, 4$ симплексных сигналов, представленных на рис. 2.7, величина коэффициента корреляции принимает значения, равные -1 (противоположные сигналы), $-1/2$ и $-1/3$ соответственно, что в свою очередь отвечает углам в 180° , 120° и примерно 110° . При $M = 4$ симплексные вектора образуют простейший правильный многогранник (тетраэдр), что объясняет наименование сигналов: «симплекс» на латинском означает «простой».

В любом множестве эквидистантных сигналов $d(\mathbf{s}_i, \mathbf{s}_j) = d_{\min}$ для любой пары различных векторов, так что в (2.23) $n_{\min} = M(M - 1)$, что также совпадает с числом слагаемых в (2.22). Подстановка последнего совместно с (2.30) в соотношение (2.23) дает асимптотическую оценку вероятности ошибки, достижимую при использовании симплексных сигналов, которая в то же время, согласно (2.22), служит верхней границей вероятности ошибки

$$P_{e \min} \leq (M - 1) Q \left(\sqrt{\frac{ME}{(M - 1)N_0}} \right), \frac{2E}{N_0} \gg 1. \quad (2.31)$$

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

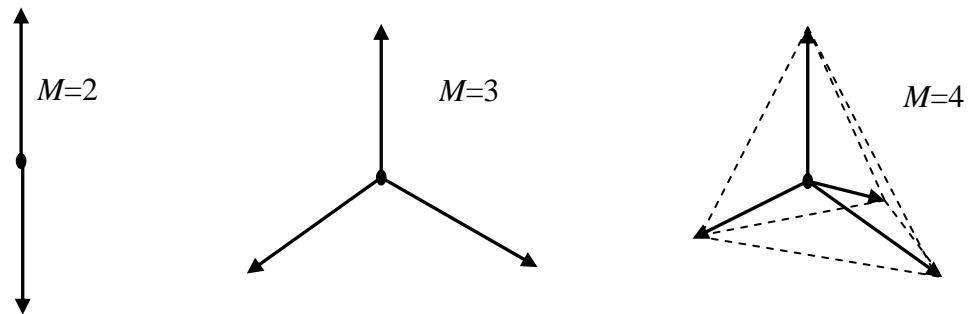


Рис. 2.7. Примеры симплексных сигналов.

Ортогональные сигналы, являющиеся еще одним примером эквидистантных, практически столь же эффективны, как и симплексные, при достаточно большом значении M . Действительно, коэффициент корреляции для ортогональных сигналов равен нулю, а расстояние между любой их парой $d(\mathbf{s}_i, \mathbf{s}_j) = d_{\min} = \sqrt{2E}$. Используя этот факт в (2.23), получим асимптотическую оценку вероятности ошибки для M ортогональных сигналов, которая ограничивает сверху точное значение вероятности ошибки:

$$P_{e \text{ ort}} \leq (M - 1) Q \left(\sqrt{\frac{E}{N_0}} \right), \frac{2E}{N_0} \gg 1. \quad (2.32)$$

Сравнение (2.32) с (2.31) демонстрирует, что для выравнивания в обоих случаях вероятностей ошибок необходимо увеличить энергию ортогональных сигналов в $M/(M - 1)$ раз по сравнению с симплексными, т.е. энергетические потери γ первых относительно вторых определяются как $\gamma = M/(M - 1)$. При $M \gg 1$ эти потери пренебрежимо малы и ортогональные сигналы могут рассматриваться как оптимальные. Например, при $M = 64$ потери $\gamma = 64/63$, что соответствует увеличению энергии ортогональных сигналов по сравнению с симплексными меньше, чем на 0.07 дБ (или 2%). Данное различие,

конечно, не имеет практического значения и, если M достаточно велико, ортогональные и симплексные сигналы могут рассматриваться взаимозаменяемыми в случае учета реализационных или других причин.

Говоря об M -ичной ортогональной сигнализации (в литературе используются также термины *ортогональная модуляция* или *ортогональное кодирование*) вспомним, что максимальное число ортогональных сигналов в точности совпадает с размерностью сигнального пространства, т.е. $M = n_s$. Поэтому в пределах общей полосы W_t и длительности T_t могут быть расположены $W_t T_t$ видео или $2W_t T_t$ полосных ортогональных сигналов. Дополнительное физическое обоснование удвоения числа полосных ортогональных сигналов в сравнении с видеосигналами непосредственно следует из соотношений (2.24) и (2.25). Построение n_s ортогональных сигналов вида (2.24) 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 использование для идентификации сообщений. В действительности, однако, подобное не осуществимо, поскольку либо сам передатчик, либо канал может нарушить когерентность сигналов, в результате чего их фазы становятся случайными и поэтому не могут использоваться для различения сообщений. Более подробно этот случай будет рассмотрен в 2.5.

2.4. Комплексная огибающая полосного сигнала.

Перед тем, как перевести обсуждение в область более сложных задач при передаче M -ичных сигналов, рационально несколько отклониться в сторону от основной линии с тем, чтобы несколько больше познакомиться с описанием полосных сигналов.

Начнем с замечания о том, что вещественная огибающая $S(t)$ в соотношении (2.24) является воображаемой, т.е. служит подходящим искусственным инструментом, тогда как только собственно сигнал $s(t)$ является наблюдаемой физической реальностью. Более того, выражение (2.24) не позволяет единственным образом определить огибающую сигнала $s(t)$. Действительно, как следует из (2.24), в качестве закона «фазовой модуляции» возможно выбрать $\gamma(t)$, а в качестве «огибающей» – $S(t) = s(t) / \cos(2\pi f_0 t + \gamma(t))$, что позволит задать данный сигнал $s(t)$. Следовательно, необходимо установить некоторые соглашения о том, как определять понятие огибающей или амплитудной модуляции $S(t)$.

Общепринятой основой для определения понятия огибающей служит преобразование Гильберта. Под физическим содержанием преобразования Гильберта понимается просто операция фильтрации, при которой фазы всех гармонических компонентов независимо от частоты сдвигаются на один и тот же угол $-\pi/2$, а амплитуды гармоник не претерпевают никаких изменений. В частотной области подобное преобразование означает просто умножение спектра сигнала на $-j\pi/2$ для положительных и на $j\pi/2$ для отрицательных частот и, следовательно, передаточная функция фильтра Гильберта имеет вид

$$\tilde{h}_g(f) = -j \frac{\pi}{2} \operatorname{sign} f, \text{ где } \operatorname{sign} x = 1, x \geq 0 \text{ и } \operatorname{sign} x = -1, x < 0.$$

Непосредственное вычисление обратного преобразования Фурье приводит к импульсной характеристике фильтра $h_g(t) = 1/\pi t$. Откуда во временной области преобразование Гильберта $s_{\perp}(t)$ сигнала $s(t)$ может быть представлено в виде интеграла свертки

$$s_{\perp}(t) = \frac{1}{\pi} \int_0^{\infty} \frac{s(\theta)}{t - \theta} d\theta. \quad (2.33)$$

Используя определение преобразования Гильберта и теорему Парсеваля, не составляет труда убедиться в справедливости следующих соотношений

$$s(t) = -\frac{1}{\pi} \int_0^{\infty} \frac{s_{\perp}(\theta)}{t-\theta} d\theta \quad (\text{обратное преобразование Гильберта}),$$

$$(\mathbf{u}, \mathbf{v}) = (\mathbf{u}_{\perp}, \mathbf{v}_{\perp}), \quad (\mathbf{u}, \mathbf{v}_{\perp}) = -(\mathbf{u}_{\perp}, \mathbf{v}). \quad (2.34)$$

Первое из выражений (2.34) демонстрирует, что преобразование Гильберта не изменяет скалярного произведения сигналов $u(t), v(t)$, тогда как второе устанавливает соотношение между скалярными произведениями одного из сигналов и преобразования Гильберта другого.

Возвращаясь теперь к определению огибающей сигнала, положим

$$S(t) = \sqrt{s^2(t) + s_{\perp}^2(t)}. \quad (2.35)$$

На первый взгляд данное определение огибающей выглядит как нечто искусственное, однако более глубокое проникновение открывает его естественность. Действительно, как можно было бы вычислить неизвестную постоянную амплитуду A немодулированного непрерывного колебания наблюдаемого сигнала $u(t) = A \cos(2\pi f_0 t + \gamma)$? Одним из возможных вариантов является использование самого сигнала и его копии, сдвинутой на угол $-\pi/2$, а затем применение к ним теоремы Пифагора: $A = \sqrt{u^2(t) + v^2(t)}$. Однако сразу же становится очевидным, что для немодулированного сигнала $u(t)$ его сдвинутая по фазе копия $v(t)$ является ничем иным, как трансформантой Гильберта: $v(t) = u_{\perp}(t)$. Таким образом, получаем результат абсолютно совпадающий с (2.35). Возьмем теперь модулированный сигнал $s(t)$. Его огибающая $S(t)$ в момент времени t представляет собой мгновенное значение амплитуды. В случае полосного сигнала $S(t)$ изменяется достаточно медленно по сравнению с несущей частотой непрерывного колебания $\cos 2\pi f_0 t$. Значит в пределах достаточно малого интервала времени, примыкающего к моменту t , $s(t)$ может трактоваться как немодулированное гармоническое колебание с амплитудой $S(t)$. Тогда каким способом может быть найдено значение этой амплитуды $S(t)$? Точно таким же, как это было сделано в случае немодулированного сигнала, т.е. с помощью фазового сдвига на угол $-\pi/2$ (или преобразованием Гильберта) и применения теоремы Пифагора (2.35), иллюстрацией чему служит рис. 2.8. Таким образом, проблема неоднозначного понимания огибающей полосного сигнала разрешена и определение (2.35) может использоваться универсально.

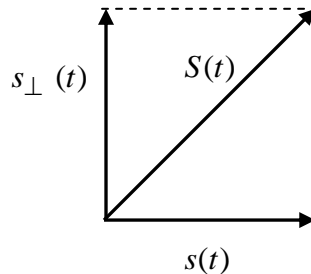


Рис. 2.8. Определение огибающей.

Анализ полосных сигналов становится значительно проще при введении в рассмотрение еще одного удобного инструмента – *комплексной огибающей (complex envelope)* $\dot{S}(t)$, которая представляет собой комплекснозначную функцию времени, непосредственно вытекающую из соотношений (2.24) или (2.25), после установления опреде-

ления для вещественной огибающей:

$$\dot{S}(t) = S_I(t) + jS_Q(t) = S(t)[\cos \gamma(t) + j \sin \gamma(t)] = S(t) \exp[j\gamma(t)], \quad (2.36)$$

где использована формула Эйлера, а $j = \sqrt{-1}$. Как следует из последнего соотношения, комплексная огибающая описывает как амплитудную, так и угловую модуляцию сигнала. Если рассматриваются несколько сигналов, обладающих одной и той несущей частотой, их отличие состоит только в законах модуляции, и, следовательно, комплексные огибающие дают их исчерпывающее описание.

Несомненно, комплексная огибающая, как и ее вещественный аналог, являются лишь удобными математическими образами, и «подлинный» сигнал (2.24) выражается с их помощью как

$$s(t) = \operatorname{Re}[\dot{S}(t) \exp(j2\pi f_0 t)], \quad (2.37)$$

где Re обозначает сохранение только вещественной части комплексной переменной, а второй сомножитель в квадратных скобках – комплексную запись несущей частоты непрерывного колебания, полученную с помощью формулы Эйлера. Обратившись снова к рис. 2.8, можно увидеть, что если согласно (2.37) трактовать $s(t)$, как вещественную часть комплексного сигнала $\dot{S}(t) \exp(j2\pi f_0 t)$, то его мнимая часть представляет сигнал, полученный преобразованием Гильберта:

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

Тогда возможен еще один вариант комплексного представления реального сигнала, называемый *аналитическим сигналом* (*analytic signal*):

$$\dot{s}(t) = \dot{S}(t) \exp(j2\pi f_0 t) = s(t) + js_{\perp}(t). \quad (2.38)$$

Формально аналитический сигнал использует комплексное представление для распространения модели (2.24) на полосный сигнал так, чтобы первый сомножитель охватывал все возможные виды модуляции (а не только ее амплитудную разновидность), а второй отвечал только за немодулированную несущую непрерывного колебания.

Используя основное правило спектрального анализа, легко можно доказать, что спектр комплексной огибающей полосного сигнала (2.37) располагается около нулевой частоты. Следовательно, поскольку при заданной несущей частоте сигнал полностью определяется своей комплексной огибающей, то последняя является низкочастотным эквивалентом полосного сигнала, упрощающим аналитическую и вычислительную работу за счет избавления от зависимости от несущей частоты.

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

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

где комплексное сопряжение используется для сохранения равенства между скалярным произведением вектора с самим собой и квадратом длины вектора (всегда вещественного и не отрицательного), тогда как совпадение скалярного произведения аналитических сигналов и комплексных огибающих следует из определения (2.38). В частности, для сигнала $s(t)$ с энергией E согласно соотношениям (2.36) и (2.35) получаем

$$(\dot{\mathbf{S}}, \dot{\mathbf{S}}) = \|\dot{\mathbf{S}}\|^2 = \int_0^T |\dot{S}(t)|^2 dt = \int_0^T S^2(t) dt = \int_0^T s^2(t) dt + \int_0^T s_{\perp}^2(t) dt = 2E, \quad (2.40)$$

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

Рассмотрим теперь два сигнала $s_i(t)$ и $s_j(t)$, и вычислим квадрат расстояния между их комплексными огибающими $\dot{S}_i(t)$ и $\dot{S}_j(t)$:

$$d^2(\dot{S}_i, \dot{S}_j) = \|\dot{S}_i - \dot{S}_j\|^2 = (\dot{S}_i - \dot{S}_j, \dot{S}_i - \dot{S}_j) = 2E_i + 2E_j - 4\text{Re}[\dot{\rho}_{ij}\sqrt{E_i E_j}], \quad (2.41)$$

где учтена линейность скалярного произведения (2.39) и соотношение (2.40); а $\dot{\rho}_{ij}$, как и в (2.16), коэффициент корреляции, но адаптированный к комплекснозначным сигналам, т.е. комплексным огибающим

$$\dot{\rho}_{ij} = \frac{(\dot{S}_i, \dot{S}_j)}{\|\dot{S}_i\|\|\dot{S}_j\|} = \frac{1}{2\sqrt{E_i E_j}} \int_0^T \dot{S}_i(t)\dot{S}_j^*(t) dt. \quad (2.42)$$

Соотношение (2.41) может рассматриваться как теорема косинусов для комплексных векторов, а $\text{Re}[\dot{\rho}_{ij}]$ как адекватная мера сходства между комплексными огибающими сигналов $s_i(t)$ и $s_j(t)$. Используя равенство скалярных произведений аналитических сигналов и комплексных огибающих (2.39), а также соотношения (2.38) и (2.34), интеграл в (2.42) может быть приведен к следующему виду

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

так что $\text{Re}(\dot{\rho}_{ij}) = \rho_{ij}$, т.е. совпадает с обычным коэффициентом корреляции сигналов $s_i(t)$ и $s_j(t)$, определяемым соотношением (2.14). Учитывая это в (2.41), получаем соотношение, связывающее расстояние между комплексными огибающими и собственно сигналами

$$d^2(\dot{S}_i, \dot{S}_j) = 2d^2(\mathbf{s}_i, \mathbf{s}_j). \quad (2.43)$$

Последний результат является одним из примеров продуктивности понятия комплексной огибающей: операции с комплексными огибающими очень часто оказываются значительно компактнее и осуществимее, чем непосредственно с полосными сигналами, предоставляя возможность избежать громоздких манипуляций с тригонометрическими функциями членов с несущей частоты.

2.5. M -ичная передача данных. Некогерентные сигналы.

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

Предположим, что закон модуляции i -го полосного сигнала $s_i(t)$ описывается полностью известной комплексной огибающей $\dot{S}_i(t)$ и не изменяющейся во времени случайной начальной фазой φ_i . Тогда, в соответствие с моделью (2.37), данный сигнал представим в виде

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

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

Вместо непосредственного вычисления расстояния между сигналами $s_i(t; \varphi_i)$ и $s_j(t; \varphi_j)$ можно воспользоваться соотношением (2.43) и оценить расстояние между комплексными огибающими $\dot{S}_i(t; \varphi_i)$ и $\dot{S}_j(t; \varphi_j)$. Следуя этим путем в соответствии с обобщенной теоремой косинусов (2.41) и предполагая равенство энергий E всех сигналов, получаем

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

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

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

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

$$\dot{\rho}_{ij} = \frac{(\dot{\mathbf{S}}_i, \dot{\mathbf{S}}_j)}{2E} = \frac{1}{2E} \int_0^T \dot{S}_i(t) \dot{S}_j^*(t) dt = |\dot{\rho}_{ij}| \exp(j\varphi_{ij}), \quad (2.44)$$

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

Теперь квадрат расстояния, приведенный выше, будет записан как

$$d^2(\dot{\mathbf{S}}_{i\varphi}, \dot{\mathbf{S}}_{j\varphi}) = 4E [1 - |\dot{\rho}_{ij}| \cos(\varphi_{ij} + \varphi_i - \varphi_j)]. \quad (2.45)$$

Единственная трудность, возникающая при его вычислении, заключается в его зависимо-

сти от неизвестных фаз φ_i, φ_j . Вследствие этого, существует множество расстояний для заданных детерминированных законов модуляции $\dot{S}_i(t)$ и $\dot{S}_j(t)$, отвечающих случайной разности фаз $\varphi_i - \varphi_j$. При значительном уровне модуля коэффициента корреляции $|\rho_{ij}|$ всегда присутствует вероятность, что при неудачной комбинации фаз (т.е. при малости $\varphi_{ij} + \varphi_i - \varphi_j$) расстояние (2.45) окажется малым. Для исключения подобной ситуации необходимо, чтобы модуль коэффициента корреляции был минимально возможным, а наилучшее множество сигналов должно удовлетворять условию

$$\dot{\rho}_{ij} = \frac{(\dot{S}_i, \dot{S}_j)}{2E} = 0, i \neq j, i, j = 1, 2, \dots, M. \quad (2.46)$$

Вновь, как и в 2.3, приходим к необходимости использования ортогональных сигналов. Однако теперь условие ортогональности является более обязывающим, обуславливая ортогональность комплексных огибающих, или, другими словами, законов модуляции, сигналов, а не только самих сигналов. Вследствие этого полосные сигналы сохраняют ортогональность при любых сочетаниях фаз, поскольку $\dot{\rho}_{ij} = 0$ влечет и $\dot{\rho}_{ij}(\varphi) = 0$. С другой стороны, условие (2.46) исключает возможность обеспечения ортогональности за счет квадратурного сдвига фазы несущей, что было приемлемым в случае когерентных полосных сигналов. Прямым следствием этого является сокращение вдвое размерности сигнального пространства n_s с величины $2W_t T_t$, как это было в случае видеосигналов, до значения $W_t T_t$ при заданной полосе W_t . Данный факт может быть объяснен и несколько иным образом: ортогональность должна обеспечиваться для комплексных сигналов, т.е. в комплексном векторном пространстве. Каждый комплексный вектор представим в виде двух, отвечающих вещественной и мнимой частям, и, следовательно, требует для своего представления двумерного векторного пространства, т.е. плоскости. Все эти плоскости должны быть ортогональны друг другу. Поскольку максимально достижимая размерность пространства составляет величину $2W_t T_t$, то в нем можно расположить только $W_t T_t$ ортогональных плоскостей, а, значит, и $W_t T_t$ ортогональных комплексных сигналов.

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

Еще один момент, о котором следует упомянуть, касается выяснения оптимальной стратегии решения при некогерентном приеме. Наблюдение $y(t)$, рассматриваемое как полосный сигнал, может быть выражено в терминах комплексной огибающей или закона модуляции $\dot{Y}(t)$: $y(t) = \text{Re}[\dot{Y}(t) \exp(j2\pi f_0 t)]$. Очевидно, что $\dot{Y}(t)$ является случайным процессом. При решении какой из M возможных сигналов был принят необходимо сравнить расстояния между наблюдением и сигналами, что согласно, соотношению (2.43), эквивалентно сравнению квадратов расстояний между соответствующими комплексными огибающими: $d^2(\dot{S}_i, \dot{Y}) = \|\dot{Y}\|^2 + 2E - 2\text{Re}[\dot{z}_i(\varphi_i)]$, где $\dot{z}_i(\varphi_i)$ – значение корреляции (скалярного произведения) комплексных огибающих наблюдения $y(t)$ и i -го сигнала со случайной фазой $s_i(t; \varphi_i)$. Для исключения зависимости от случайной фазы логично выбрать для сравнения только минимальное по всем возможным значениям φ_i расстояние для каждого i . Опуская детали, воспроизводящие почти дословно приведенные выше рассуждения, окончательное решающее правило выражается в терминах модуля корреляции

$$Z_i = \left| \int_0^T \dot{Y}(t) \dot{S}_i^*(t) dt \right|, i = 1, 2, \dots, M, \quad (2.47)$$

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

2.6. Обмен между выигрышем от ортогонального кодирования и шириной полосы.

Предыдущее рассмотрение очевидным образом высветило особую роль, которую играют ортогональные сигналы: они практически (при $M \gg 1$ для когерентных сигналов) и даже теоретически (для некогерентного случая) оптимальны в случае M -ичной передачи данных. Оценим теперь выигрыш, который сопровождает применение ортогональных сигналов со случаем отсутствия кодирования, т.е. непосредственной передачей потока информационных бит источника. Предположим, что энергетический ресурс позволяет передавать каждый бит данных с энергией E_b , отсутствует специальное кодирование потока бит данных и каждый информационный бит передается оптимальным образом с помощью пары противоположных сигналов, другими словами используется BPSK. Среди блоков из m последовательных бит возможны любые комбинации, в том числе и такие, которые отличаются друг от друга значением только одного бита. Поэтому минимальная величина квадрата расстояния между сигналами, отвечающими не идентичным m -битовым блокам, совпадает со значением квадрата расстояния между однобитовыми противоположными сигналами, т.е. согласно рис. 2.5, а $d_{\min, u}^2 = 4E_b$, где второй индекс обозначает отсутствие кодирования (uncoded).

Рассмотрим теперь другую систему, в которой все различные m -битовые блоки передаются ортогональными сигналами. Ясно, что каждый такой сигнал характеризуется энергией, равной mE_b , сохраняя неизменной ее значение на бит E_b . Тогда квадрат расстояния (одинаковый для любой пары сигналов, поскольку ортогональные сигналы эквидистантны) между сигналами снова легко может быть определен на основании рис.2.5, б: $d_{\min, ort}^2 = 2mE_b$. Очевидно, что выигрыш G_a ортогональных сигналов в величине минимума квадрата расстояния у варианта отсутствия кодирования составляет $m/2$. При альтернативном подходе для обеспечения одинакового минимума расстояния передача данных без кодирования требует энергетических затрат в $G_a = m/2$ больше, чем в случае ортогонального кодирования. Принимая во внимание соотношение (2.23), минимум расстояния асимптотически определяет вероятность ошибки для любых сигналов, используемых в M -ичной передаче. Следовательно, требование высокой достоверности приема, автоматически влекущее за собой необходимость высокого значения отношения сигнал-шум, означает, что одинаковая надежность двух рассмотренных систем возможна только в случае G_a раз больших энергетических затратах для варианта без кодирования. Таким образом, *асимптотический* выигрыш G_a от кодирования ортогональными сигналами является адекватным показателем преимущества ортогональной сигнализации в пределе, т.е. при отношении сигнал-шум, стремящемся к бесконечности.

Для определения величины выигрыша от ортогонального кодирования при конечном значении отношения сигнал-шум, также как скорости его сходимости к асимптотическому случаю на рис.2.9 представлены семейства кривых для двух значений длины m -

битового блока: $m = 6$ и $m = 20$. Первая кривая (сплошная линия) демонстрирует вероятность ошибочного приема некодированного блока, а две других вычислены в предположении кодирования m -битовых блоков $M = 2^m$ ортогональными сигналами, которые обрабатываются когерентно. Пунктирная кривая построена на основании аддитивной границы (2.32), тогда как кривая, обозначенная через точку и тире, отвечает точной формуле вычисления вероятности ошибки когерентного приема M ортогональных сигналов, вывод которой можно найти во многих популярных книгах по основам теории связи [например 5,7]:

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

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

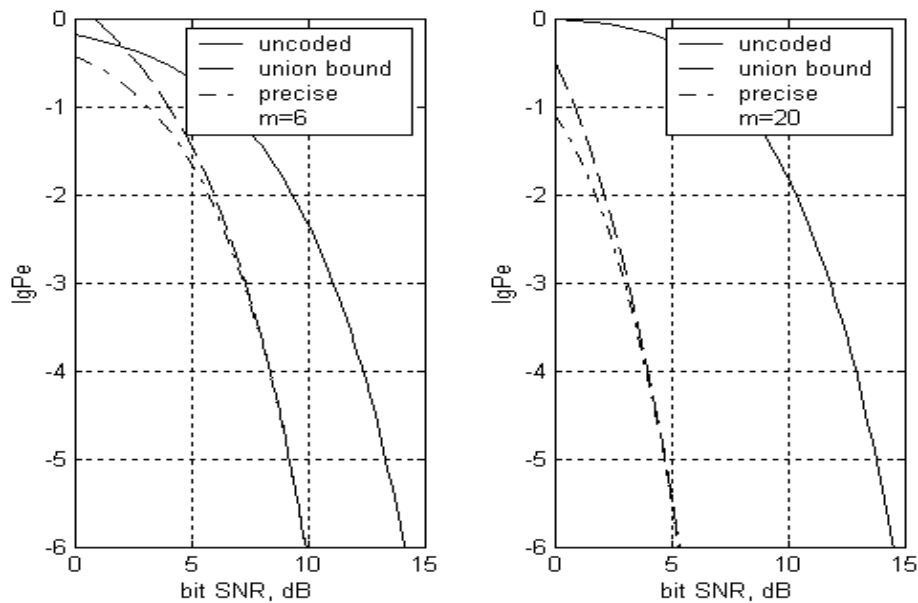


Рис. 2.9 Сравнение вероятности ошибки для некодированной передачи и ортогональной сигнализации

Основной вывод, который следует из графиков на рис. 2.9, состоит в высокой достоверности аддитивной границы. В области типичных для практики требований к вероятности ошибки просвет между величиной отношения сигнал-шум на бит, оцененного из границы и вычисленного точно, составляет менее 0.5 дБ ($P_e \approx 10^{-2}$) и быстро уменьшается с ростом требований к достоверности передачи (менее 0.2 дБ при $P_e \leq 10^{-3}$). В диапазоне значений P_e от 10^{-2} до 10^{-6} действительный выигрыш G от ортогонального кодирования возрастает с 3.5 до 4.2 дБ ($m = 6$) и от 8.5 до 8.9 дБ ($m = 20$). Сравнивая эти цифры с асимптотическими значениями (4.8 и 10 дБ соответственно), нетрудно увидеть более чем хорошее совпадение между ними, оправдывающее использование асимптотического выигрыша от кодирования в качестве первого приближения оценки эффективности ортогонального кодирования.

Очень оптимистическое мнение об ортогональной сигнализации, которое может возникнуть свете приведенных выше результатов, значительно снижается реальной стои-

мостью выигрыша от кодирования. Последний приобретается за счет расширения полосы, поскольку, как это было установлено в 2.3, размерность сигнального пространства n_s , т.е. число ортогональных сигналов $M = n_s$, непосредственно определяется общим частотно-временным ресурсом $W_t T_t$ системы. Интересуясь в основном установлением количественного порядка и игнорируя тривиальный путь удвоения числа когерентных ортогональных сигналов, имеем $M = W_t T_t$, или $W_t = M / T_t$. Пусть необходимая скорость передачи в системе составляет R бит/с, что соответствует передаче $m = R T_t$ бит за выделенный интервал времени T_t . Ясно, что ортогональное кодирование битовых блоков такой длины может быть осуществлено с помощью $M = 2^m = 2^{R T_t}$ сигналов, обеспечивая асимптотический выигрыш от кодирования $G_a = m/2 = R T_t / 2$. Тогда спектральная эффективность R/W_t системы, т.е. скорость, приходящаяся на 1 Гц полосы (см. 1.1), использующей ортогональные сигналы, будет

$$\frac{R}{W_t} = \frac{R T_t}{2^{R T_t}} = \frac{2 G_a}{2^{2 G_a}}, \quad (2.48)$$

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

Обратимся к характеристикам численного примера.

Пример 2.6.1. Для цифровых систем передачи речи очень типичной является скорость $R = 9.6$ кбит/с (мобильная телефония, системы мультимедиа и др.). Предположим, что необходимо уменьшить в три раза (на 4.8 дБ) излучаемую мощность без потери в величине скорости. Если достижение поставленной цели осуществляется путем использования ортогональных сигналов, то это становится возможным лишь за счет снижения спектральной эффективности с 1 до 6/64, как следует из соотношения (2.48). Другими словами, для сохранения неизменной скорости $R = 9.6$ кбит/с необходимо использовать полосу не менее чем 100 кГц. Эта цифра не является запредельной для многих приложений, так, например, на данном принципе строится обратный канал сотовой системы мобильной связи стандарта cdmaOne (для более конкретного ознакомления см. 11.2).

Представим теперь ситуацию, когда, основываясь на результатах успешно решенной предыдущей задаче, планируется достичь десятикратного (10 дБ) уменьшения излучаемой мощности. Реализация поставленной цели требует применения $M = 2^{20} > 10^6$ ортогональных сигналов для кодирования блоков длиной $m = 20$ бит. Это приведет к спектральной эффективности, меньшей $2 \cdot 10^{-5}$, или полосе, шире 480 МГц, что представляет совершенно неприемлемым при скорости в 9.6 кбит/с.¹ □

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.

Возвратимся к соотношению $M = W_t T_t$ и зададимся вопросом: когда число ортогональных сигналов, а значит, и произведение $W_t T_t$, измеряется десятками и более, означает ли это применение технологии распределенного спектра? Другими словами, является ли система, использующая значительное число ортогональных сигналов, системой с распределенным спектром? Как покажет содержание следующего параграфа, в общем случае от-

¹ С настоящего места символ □ обозначает окончание примера или доказательства.

вет на этот вопрос является негативным.

2.7. Примеры множеств ортогональных сигналов.

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

2.7.1. Кодирование путем временного сдвига

Очевидно, что скалярное произведение любых двух не перекрывающихся во времени сигналов равно нулю. Рассмотрим M сигналов, приведенных на рис. 2.10, а, которые совместно занимают весь временной интервал T_t . При длительности сигнала не более, чем $T = T_t / M$, и временном сдвиге между соседними сигналами не менее, чем длительность сигнала, подобное кодирование путем временного сдвига (*time-shift coding*) образует семейство ортогональных сигналов. Оценка полосы W , занимаемая каждым из сигналов, равна обратной величине от их длительности, и все сигналы, не нарушая условия ортогональности, занимают одну и ту же полосу $W = W_t$. Следовательно, максимальное число подобных ортогональных сигналов, которые могут располагаться в пределах заданного частотно-временного ресурса W_t, T_t , составляет величину $M = T_t / T = W_t T_t$, т.е., как можно было и предвидеть, равняется размерности сигнального пространства $n_s = W_t T_t$. Необходимость большего числа сигналов $M \gg 1$ требует большой величины произведения $W_t T_t = M$, которое, по-видимому, может быть достигнуто путем распре-

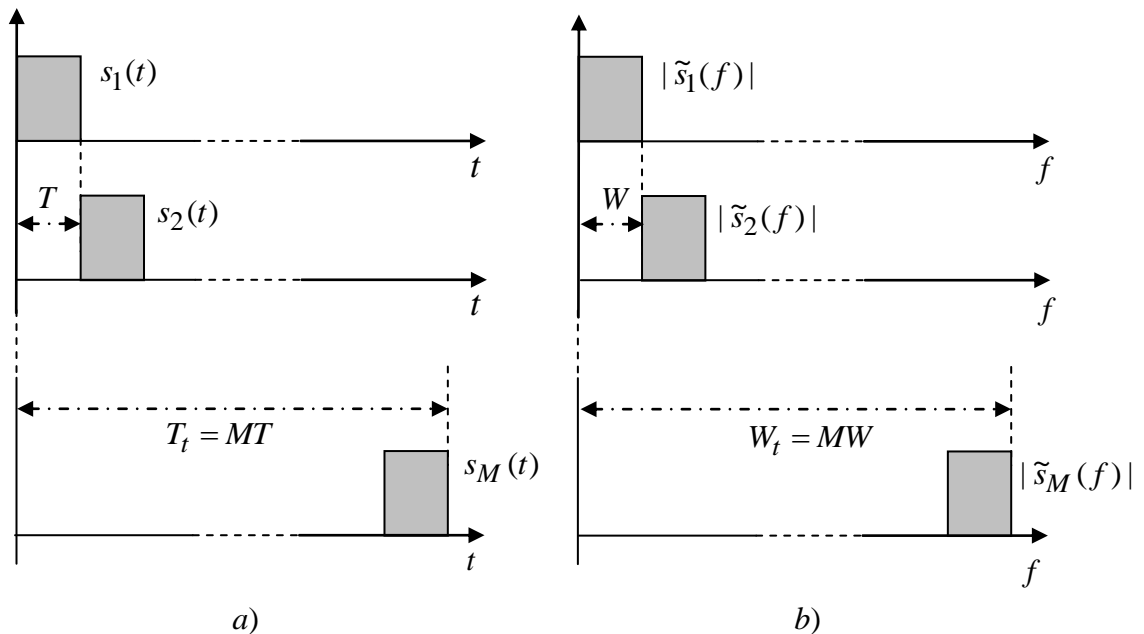


Рис. 2.10. Ортогональные по времени (а) и частоте (б) сигналы.

ления спектра. Однако для любого конкретного сигнала частотно-временное произведение $WT = W_t T = W_t T_t / M = 1$, так что данные сигналы не относятся к типу с распределенным

спектром. Поскольку согласно с ранее принятой договоренностью система называется широкополосной, если только она использует сигналы с подобной технологией (см. 1.1), то ортогональное кодирование временным сдвигом не имеет ничего общего с распределением спектра.

Пусть общий частотно-временной ресурс представлен в виде прямоугольника на координатной плоскости время–частота со сторонами, равными соответственно T_t и W_t . Тогда кодирование путем временного сдвига означает лишь разрезание этого ресурса на M вертикальных полосок, каждая из которых приписывается некоторому конкретному сигналу (см. Рис. 2.11, а). Ортогональность при передаче обеспечивается строгим распределением временного ресурса между сигналами, каждый из которых занимает весь отведенный частотный ресурс.

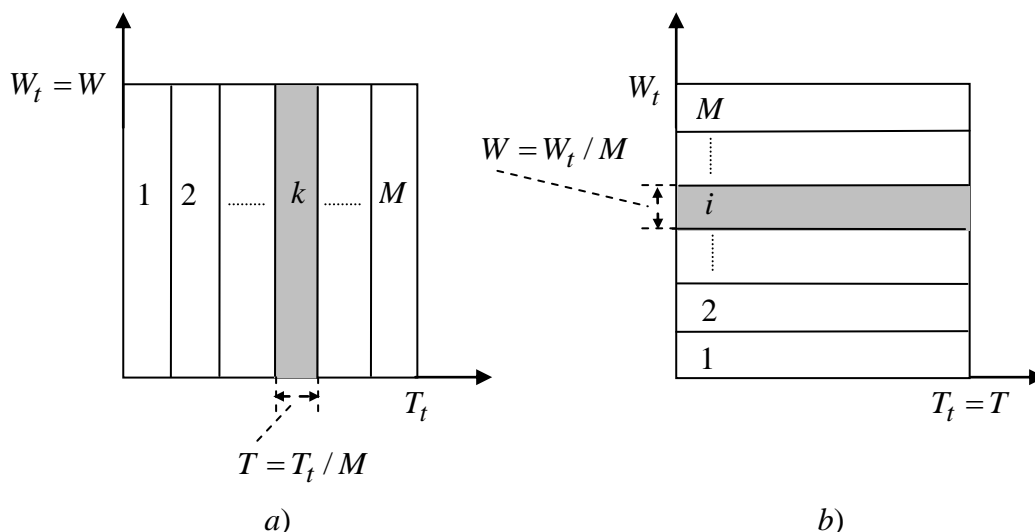


Рис. 2.11. Распределение ресурса при временном (а) и частотном (б) ортогональном кодировании.

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.

2.7.2. Кодирование путем частотного сдвига

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

$$(\mathbf{u}, \mathbf{v}) = \int_{-\infty}^{\infty} u(t) v(t) dt = \int_{-\infty}^{\infty} \tilde{u}(f) \tilde{v}^*(f) df = (\tilde{\mathbf{u}}, \tilde{\mathbf{v}}), \quad (2.49)$$

что позволяет перевести ранее представленный метод в частотную область (см. Рис. 2.10, *b*). При полном перекрытии сигналов во времени ($T = T_t$) каждый из них занимает полосу не менее чем $W = 1/T_t$. Тогда максимальное число ортогональных сигналов, образованных сдвигом спектра, снова будет $M = W_t/W = W_t T_t = n_s$. Как и в предыдущем случае, общий ресурс снова нарезается на 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. 2.11, *b*). Очевидно, что каждый индивидуальный сигнал снова не является сигналом с распределенным спектром, поскольку отвечающее ему частотно-временное произведение $WT = (W_t/M)T = 1$, и значит, любая система с неограниченно большим числом ортогональных сигналов подобного сорта, конечно, не является системой с распределенным спектром.

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.

Рассмотренные примеры объясняют, почему даже использование большого числа ортогональных сигналов и, следовательно, необходимость общего ресурса $W_t T_t \gg 1$ не означают автоматического привлечения технологии распределенного спектра.

2.7.3. 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 (BPSK) сигналов. Образуюем 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

$$s_k(t) = \sum_{i=0}^{N-1} a_{k,i} s_0(t - i\Delta), \quad (2.50)$$

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

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

$$(\mathbf{s}_k, \mathbf{s}_l) = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} a_{k,i} a_{l,j} \int_0^T s_0(t-i\Delta) s_0(t-j\Delta) dt. \quad (2.51)$$

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

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

где E_0 – энергия элементарного импульса. Использование полученного выражения в (2.51) дает

$$(\mathbf{s}_k, \mathbf{s}_l) = E_0 \sum_{i=0}^{N-1} a_{k,i} a_{l,i} = E_0 (\mathbf{a}_k, \mathbf{a}_l). \quad (2.52)$$

Последнее соотношение связывает скалярное произведение сигналов (2.50) со скалярным произведением N -мерных векторов, отвечающих кодовым последовательностям $\mathbf{a}_k = (a_{k,0}, a_{k,1}, \dots, a_{k,N-1})$. Очевидно, что M ортогональных кодовых последовательностей автоматически формируют M ортогональных сигналов типа (2.50). 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 это 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:

$$\mathbf{H}_{2M} = \begin{bmatrix} \mathbf{H}_M & \mathbf{H}_M \\ \mathbf{H}_M & -\mathbf{H}_M \end{bmatrix} = \begin{bmatrix} \mathbf{1} & \mathbf{1} \\ \mathbf{1} & -\mathbf{1} \end{bmatrix} \otimes \mathbf{H}_M, \quad (2.53)$$

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

Применение алгоритма Сильвестра начинается с матрицы $\mathbf{H}_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$, которая является простейшей матрицей Адамара. Следующий шаг состоит в построении матрицы

\mathbf{H}_4 , в которой для краткости обозначим символами «+» и «-» значения элементов «+1» и «-1». Затем от \mathbf{H}_4 перейдем к \mathbf{H}_8 и т.д.:

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

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

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

На рис. 2.12 представлены низкочастотные ортогональные бинарные фазоманипулированные сигналы (2.50) – функции Уолша, построенные на основе матрицы Адамара \mathbf{H}_8 .

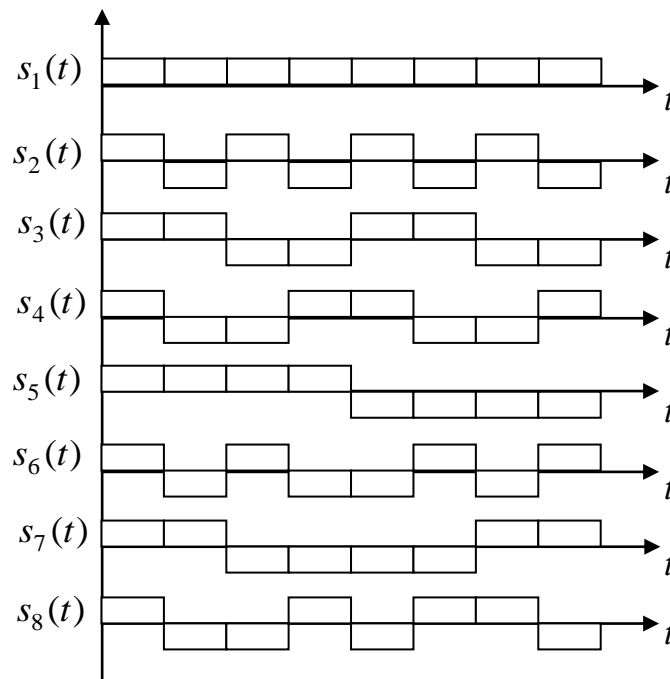


Рис. 2.12. Функции Уолша на основе \mathbf{H}_8 .

Рис. 2.13 иллюстрирует тот факт, что в рамках данного метода построения ортогональных сигналов, не происходит деления общего ресурса: все сигналы располагаются в нем, полностью перекрываясь как во временной, так и в частотной области. Действительно, полоса, занимаемая каждым сигналом, может быть оценена как $W = 1/\Delta$, тогда как длительность составляет величину $T = M\Delta$, откуда $WT = M = W_t T_t$. Ортогональность же в данном методе достигается не путем деления временного интервала или полосы, а за счет соответствующего выбора закона модуляции сигнала.

Анализируя достоинства ортогональности за счет распределения спектра, можно

отметить, что методы формирования и обработки сигналов (2.50) 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 (см. главу 3). Это объясняет значительное распространение ортогональной сигнализации данного типа в современных телекоммуникационных системах (cdmaOne, UMTS, cdma2000, см. главу 11).

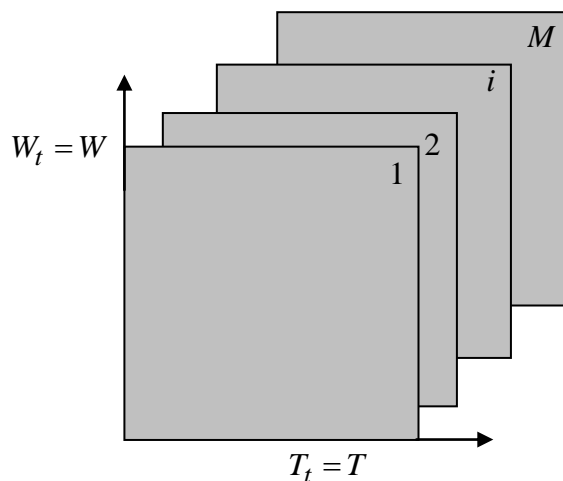


Рис. 2.13. Распределение ресурса при ортогонализации.

В заключении подведем итог материалу, изложенному в 2.5-2.7. Как можно видеть, теоретически классическая задача M -ичной передачи информации не ориентирована на безусловное применение технологии распределенного спектра и, в принципе, оптимальный выбор сигналов может быть осуществлен в классе простых сигналов. С другой стороны, существуют причины реализационного плана вместе с желанием приобрести преимущества, присущие распределенному спектру вне рамок классического приема. Поскольку упомянутые возможности полностью связаны с необходимостью использования сигналов со значительным частотно-временным ресурсом $W_t T_t \gg 1$, то в последнее время разработчики систем склоняются к предпочтению сигналов с распределенным спектром, а не простых.

2.8. Оценивание параметров сигнала.

2.8.1. Формулировка задачи и правила оценивания.

Во всех радиотехнических системах типичной является задача *измерения* (или *оценки*) *параметров* сигнала. Она касается любой ситуации, когда информация, интересующая наблюдателя, содержится в текущем значении некоторого сигнального параметра (например, амплитуде, частоте, начальной фазе, времени задержки и др.). В связи с чем, для извлечения необходимой информации наблюдателю необходимо измерить или оценить соответствующий параметр.

Обратимся к некоторым наиболее прозрачным примерам. В обычном АМ или ФМ радиовещании зависимость амплитуды (или частоты) во времени определяет передаваемую аудиоинформацию: громкость и тональность. Для восстановления звукового сообщения и передачи его слушателю необходимо первоначально измерить мгновенные значения амплитуды (частоты) и воспроизвести их как непрерывное во времени колебание. Другим подобным примером может служить обычная телевизионная система, в которой как амплитуда, так и частота используются для передачи видеоинформации. Для восстановления цвета движущегося образа необходимо выполнить измерения амплитуды, поскольку освещенность и компоненты цвета передаются с помощью амплитудной модуляции, тогда как передача аудиоинформации осуществляется с помощью фазовой модуляции и, следовательно, в любой телевизионной системе имеет место измерение частоты.

Еще одну разновидность задачи оценки параметров можно обнаружить в процедуре синхронизации, где частотно-временное рассогласование между принятым сигналом и локальным опорным генератором необходимо измерить для синхронизации последнего с первым. Эта процедура характерна для большого числа дальнедействующих систем: от каналов горизонтальной и вертикальной синхронизации ТВ до пилотного канала систем мобильной связи второго и третьего поколений.

Для локации и навигации также типичны проблемы, связанные с получением оценок: измерение времени задержки и направления прихода сигнала необходимо для определения взаимного положения и угловых координат между приемником и целью. Если же требуется также знание скорости цели и параметры ее маневрирования, то необходимым становится измерение доплеровского сдвига частоты и т.д.

Перечень примеров легко может быть продолжен, поскольку неотъемлемой частью функционирования любой системы, в которой информация передается, восстанавливается или обрабатывается, является измерение параметров.

С использованием удобных для рассмотрения обозначений задача оценки параметров может быть сформулирована следующим образом. Пусть наблюдение $y(t)$ наряду с шумом содержит детерминированный сигнал $s(t; \lambda)$, в котором единственным неизвестным является точное значение постоянного параметра λ . В зависимости от особенностей ситуации данное неизвестное может представлять собой как вектор, так и скаляр. Наблюдатель, основываясь на анализе $y(t)$, должен принять *решение* о том, какое значение из диапазона возможных принял параметр сигнала. Совместно с самой задачей это решение принято называть *оценкой* и обозначать, как $\hat{\lambda}$. Поскольку в принятом наблюдении $y(t)$ всегда присутствует шум, то при каждом сеансе приема оценка $\hat{\lambda}$ отличается от неизвестного истинного значения параметра λ . В связи с этим возникает вопрос: как принять оптимальное решение, которое гарантировало бы наименьший ущерб, обусловленный этими различиями.

Простейший подход к этой проблеме может быть найден при понимании того, что в принципе задача оценивания ничем не отличается от задачи различения M сигналов, рассмотренной в параграфе 2.3. Действительно, первоначально предположим, что измеряемый параметр λ является дискретным и принимает одно из M конкурирующих значе-

ний $\lambda_1, \lambda_2, \dots, \lambda_M$. Тогда решение о том, какое из возможных значений принял параметр сигнала в данном конкретном наблюдении, есть ничто иное, как проверка M гипотез о том, какой из M конкурирующих сигналов $s_1(t), s_2(t), \dots, s_M(t)$ был фактически принят. Причем рассматриваемое множество составляют копии сигнала $s(t; \lambda)$, отличающиеся друг от друга только значением параметра λ , т.е. $s_i(t) = s(t; \lambda_i)$. Для того, чтобы охватить в такой постановке и случай непрерывного параметра λ , необходимо предположить о бесконечном (даже бессчетном) числе возможных значений M параметра, а, следовательно, и различаемых сигналов.

Из приведенных рассуждений следует, что хорошо известная оптимальная стратегия решений – правило максимального правдоподобия – применима и для оценки параметров. Это означает, что среди всех конкурирующих значений λ в качестве оценки $\hat{\lambda}$ следует выбирать то, которое максимизирует вероятность трансформации каналом сигнала $s(t; \lambda)$ в наблюдаемое колебание $y(t)$. Для канала с аддитивным белым гауссовским шумом (АБГШ) это правило эквивалентно правилу минимума расстояния, которое с использованием введенных обозначений представимо в следующем виде

$$d(\mathbf{s}_{\hat{\lambda}}, \mathbf{y}) = \min_{\lambda} d(\mathbf{s}_{\lambda}, \mathbf{y}) \Rightarrow \hat{\lambda}, \quad (2.54)$$

где \mathbf{s}_{λ} – векторное обозначение сигнала $s(t; \lambda)$. Применение данного правила обеспечивает получение *максимально правдоподобной* (МП) оценки $\hat{\lambda}$ в результате нахождения такого значения λ , при котором сигнал $s(t; \lambda)$ наиболее близок по расстоянию Евклида к наблюдению $y(t)$. Иллюстрацией последнего утверждения служит рис. 2.14. Сигнал

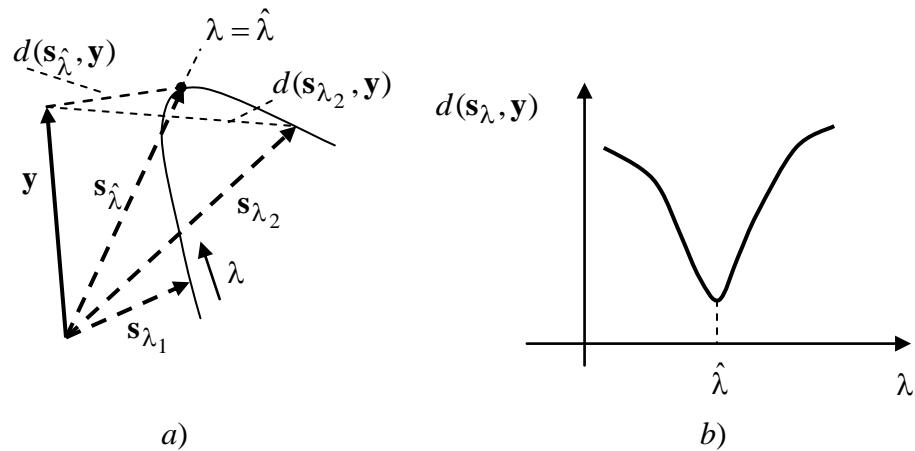


Рис. 2.14 Иллюстрация МП оценки.

$s(t; \lambda)$ может трактоваться как вектор \mathbf{s}_{λ} , перемещение которого отвечает изменению параметра λ . Конец вектора описывает некоторую траекторию, точки которой соответствуют различным значениям параметра λ (рис.2.14, a). Точка траектории, наиболее близкая к вектору наблюдения \mathbf{y} , определяется согласно правилу (2.54), а соответствующее ей значение λ принимается в качестве оценки, которое также представлено на рис. 2.14, b, демонстрирующем примерную зависимость расстояния между наблюдением и копией сигнала от величины λ . МП оценка $\hat{\lambda}$ отвечает такому значению параметра, который минимизирует это расстояние.

Все параметры сигнала могут быть разделены на *энергетические* и *неэнергетические* на основании факта влияния параметра на энергию сигнала. Если параметр λ относится ко второму типу, то энергия сигнала $s(t; \lambda)$ не зависит от конкретного значения λ ,

т.е. $E(\lambda) = \int_{-\infty}^{\infty} s^2(t; \lambda) dt = E$. Например, амплитуда и длительность сигнала являются энер-

гетическими, тогда как запаздывание, частота или начальная фаза – неэнергетическими параметрами. Отсюда становится очевидным, что оценивание неэнергетического параметра представляет собой частный случай задачи различения конкурирующих сигналов равной энергии, для которого правило максимума корреляции (2.8) в новых обозначениях представимо как

$$z(\hat{\lambda}) = \max z(\lambda) \Rightarrow \hat{\lambda}, \quad (2.55)$$

или в виде

$$\hat{\lambda} = \arg \max_{\lambda} z(\lambda),$$

где согласно (2.7)

$$z(\lambda) = (\mathbf{y}, \mathbf{s}_{\lambda}) = \int_0^T y(t) s(t; \lambda) dt \quad (2.56)$$

корреляционный интеграл между наблюдаемым колебанием $y(t)$ и сигналом $s(t; \lambda)$, зависящий от значения измеряемого параметра λ .

В свете физической трактовки корреляции правило оценивания (2.55) имеет совершенно прозрачную интерпретацию: МП оценка $\hat{\lambda}$ отвечает такому значению λ , при котором сигнал $s(t; \lambda)$ в наибольшей степени подобен наблюдаемому колебанию $y(t)$.

2.8.2. Точность оценивания.

Обратимся теперь к факту, который достаточно глубоко обсуждался в 2.2-2.3, о том, что достоверность различения сигналов критически зависит от коэффициента корреляции (2.14). В задаче оценивания параметров различаемые сигналы есть просто копии $s(t; \lambda)$ с отличными значениями параметра λ . Во многих практических ситуациях корреляция двух любых таких копий $s(t; \lambda_1)$ и $s(t; \lambda_2)$ определяется лишь их рассогласованием в величине λ , т.е. разностью $\lambda_2 - \lambda_1$, а не собственно значениями λ_1, λ_2 , так что, положив $\lambda_1 = 0, \lambda_2 = \lambda$, приходим к следующему выражению для коэффициента корреляции (2.14), отвечающему неэнергетическому параметру λ

$$\rho(\lambda) = \frac{(\mathbf{s}_0, \mathbf{s}_{\lambda})}{E} = \frac{1}{E} \int_{-\infty}^{\infty} s(t; 0) s(t; \lambda) dt. \quad (2.57)$$

Как и ранее, коэффициент корреляции характеризует степень похожести двух сигнальных копий в зависимости от их рассогласованности по параметру λ . Очевидно, что $\rho(\lambda) \leq \rho(0) = 1$, что имеет объективный характер: сигнальные копии, отличающиеся по величине λ , не могут быть более похожими друг на друга, чем при полной идентичности, которой отвечает единичная корреляция. Другим свойством величины (2.57), вытекающим из определения, является ее четность, т.е. $\rho(\lambda) = \rho(-\lambda)$.

Рис. 2.15 демонстрирует характерные примеры зависимости $\rho(\lambda)$ от λ для двух гипотетических сигналов. Сплошная кривая носит более гладкий характер, чем пунктирная, что свидетельствует о меньшей чувствительности первого сигнала к изменению параметра λ : сходство его копий, обусловленных рассогласованностью по параметру λ , выше, чем у копий второго сигнала при том же значении рассогласования.

Теперь необходимо подчеркнуть одну деталь, которая выше намеренно была опущена. В самом деле, то или иное правило решения является оптимальным только в некотором ограниченном смысле, определяемом критерием оптимальности. МП правило, используемое ранее, является оптимальным по критерию *минимума вероятности ошибки*,

применение которого вполне естественно и оправдано в случае различения дискретных сигналов или измерения дискретного параметра. Однако оно не выглядит адекватным в случае измерения непрерывного параметра. Представляется более рациональным характеризовать достоверность оценивания с точки зрения точности, т.е. величиной отклонения $\varepsilon = \hat{\lambda} - \lambda$ оценки параметра $\hat{\lambda}$ от его истинного значения λ . Прежде всего, разумно потребовать, чтобы математическое ожидание ошибки ε , усредненное по всем возможным наблюдениям $y(t)$ при фиксированном истинном значении λ , равнялось бы нулю, т.е. оценка $\hat{\lambda}$ в среднем совпадала бы с истинным значением λ :

$$\overline{\varepsilon} = \overline{\hat{\lambda} - \lambda} = 0 \Leftrightarrow \overline{\hat{\lambda}} = \lambda, \forall \lambda. \quad (2.58)$$

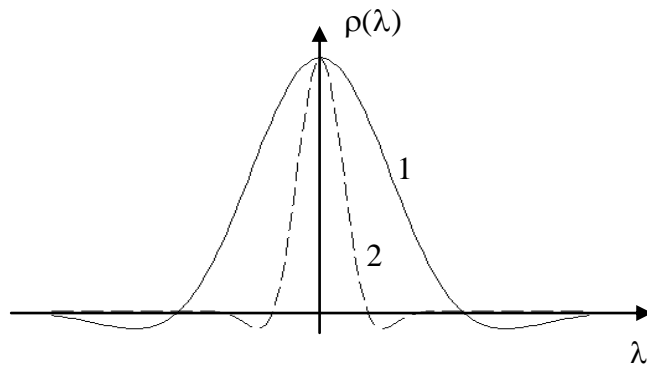


Рис. 2.15. Типичные кривые $\rho(\lambda)$ в зависимости от λ .

Оценка, для которой выполняется данное условие, называется *несмещенной (unbiased)*. Однако даже выполнение (2.58) не позволяет еще считать оценку хорошей, поскольку существенное значение имеет также величина разброса оценки относительно истинного значения. Традиционной и адекватной мерой разброса служит дисперсия ошибки $D\{\varepsilon\} = \overline{(\hat{\lambda} - \lambda)^2}$, поэтому стараются найти такое правило, которое обеспечивало бы получение несмещенной оценки с минимальным значением дисперсии $D\{\varepsilon\}$ для всех истинных значений λ :

$$D\{\varepsilon\} = \overline{(\hat{\lambda} - \lambda)^2} = \min, \forall \lambda.$$

Следовательно, минимизация дисперсии несмещенной оценки является естественным путем достижения наибольшей точности измерений.

В теории оценивания фундаментальная граница Крамера–Рао устанавливает нижний предел величины дисперсии любой несмещенной оценки. Оценка, дисперсия которой удовлетворяет этой границе, называется *эффективной*. Хотя «совершенно» эффективные оценки достаточно редки, однако это не является столь уж большой проблемой с практической точки зрения. Дело в том, что, как следует из классической теории оценивания, МП оценка оказывается *асимптотически* несмещенной и эффективной. Термин «асимптотически» фактически означает «в ситуациях, когда необходима высокая точность измерений» или, перейдя на язык практики, когда обеспечивается достаточно большое отношение сигнал-шум, либо время наблюдения. Поэтому в любой задаче, где желательно достижение высокой точности, правило максимального правдоподобия является оптимальным не только по критерию вероятности ошибки, но также и по точностному критерию. Конечно, в реальных условиях требование высокой точности измерений является типичным и именно поэтому МП оценки нашли широкое использование.

Для неэнергетического параметра λ граница Крамера–Рао имеет особенно простое выражение и дает практический инструмент для вычисления дисперсии МП оценки

$$D\{\hat{\lambda}\} = D\{\varepsilon\} \approx -\frac{1}{\rho''(0)q^2}, \quad q \gg 1. \quad (2.59)$$

Присутствие отношения сигнал–шум $q^2 = 2E/N_0$ в знаменателе правой части соотношения (2.59) не вызывает удивления: действительно, для любого рационального правила оценивания справедливо утверждение – чем выше отношение сигнал–шум, тем меньше ошибка и тем выше точность измерений. В то же время, зависимость от второй производной коэффициента корреляции заслуживает более обширного комментария. Из курса математического анализа хорошо известно, что вторая производная говорит о кривизне или остроте функции в рассматриваемой точке и для выпуклой кривой является отрицательной. В свою очередь, острота $\rho(\lambda)$ в нулевой точке показывает чувствительность сигнала по отношению к рассогласованности в величине λ : чем острее $\rho(\lambda)$, тем быстрее копия сигнала с рассогласованием по величине λ теряет свое подобие с исходной копией. Напомним еще раз, что оценивание есть частный случай различения сигналов, а последнее выполняется тем надежнее, чем меньше похожесть сигналов. Это полностью объясняет зависимость точности измерения λ от $\rho''(0)$: если копии $s(t; \lambda)$ имеют малое сходство даже при незначительном отличии в величине λ , то они различаются наиболее просто по сравнению со случаем их большего подобия.

Последний факт указывает общее направление синтеза сигналов в задачах оценивания неэнергетического параметра λ . Для достижения желаемого результата не ценой только «грубой силы», т.е. увеличения энергии, следует искать сигналы с сильной зависимостью коэффициента корреляции $\rho(\lambda)$ от параметра λ .

В следующих параграфах обратимся к конкретным задачам оценивания, среди которых рассмотрим примеры измерения как энергетических, так и неэнергетических параметров. Основная идея остается прежней: необходимо определить в каких задачах оценивания следует привлекать технологию распределенного спектра.

2.9. Оценивание амплитуды сигнала.

Задача измерения интенсивности (уровня, мощности) сигнала имеет место во множестве приложений: от систем телевизионного вещания до цифровых систем передачи данных и мобильных систем связи. Рассмотрим данную проблему как задачу измерения неизвестной амплитуды A , не изменяющейся на интервале наблюдения $[0, T]$. В этих условиях модель сигнала может быть представлена в следующем виде

$$s(t; A) = As(t),$$

где $s(t)$ – некоторый детерминированный сигнал (образец), амплитуда которого по определению считается равной единице. Тогда сигнал $s(t; A)$ рассматривается как результат произведения опорного сигнала-образца на неизвестный множитель. Пусть энергия опорного сигнала равна E . Тогда энергия сигнала с амплитудой A и его корреляция с наблюдением $y(t)$ определяются как

$$E(A) = \int_0^T s^2(t; A) dt = A^2 \int_0^T s^2(t) dt = A^2 E, \quad z(A) = \int_0^T y(t) s(t; A) dt = Az,$$

где

$$z = \int_0^T y(t) s(t) dt \quad (2.60)$$

корреляция наблюдения и опорного сигнала.

Обратившись к корреляционной версии правила минимума расстояния (2.8) и отмечая, что роль E_j и z_j в нынешних условиях выполняют $E(A)$ и $z(A)$ соответственно, процедура определения МП оценки может трактоваться, как максимизация разности $z(A) - E(A)/2$ по всем значениям A . Используя ранее приведенные соотношения, данная разность принимает вид квадратичного бинома $Az - A^2E/2$ относительно A с известными коэффициентами. Легко определяемый максимум этого соотношения дает МП оценку амплитуды

$$\hat{A} = \frac{z}{E}.$$

Таким образом, вычисляя значение корреляции наблюдаемого колебания с опорным сигналом и умножая результат на постоянный множитель $1/E$, получаем желаемую оптимальную оценку амплитуды. После нахождения математического ожидания z из соотношения (2.60)

$$\bar{z} = \int_0^T \overline{y(t) s(t)} dt = \int_0^T s(t; A) s(t) dt = A \int_0^T s^2(t) dt = AE,$$

легко заметить, что среднее значение \hat{A} точно совпадает с истинным значением амплитуды: $\overline{\hat{A}} = \bar{z}/E = A$, означающее, что МП оценка амплитуды является *строго* (а не только асимптотически) *несмещенной*. Не составляет труда оценить и дисперсию \hat{A} :

$$D\{\hat{A}\} = \frac{D\{z\}}{E^2} = \frac{N_0}{2E} = \frac{1}{q^2}, \quad (2.61)$$

где учтен результат (2.15), а q^2 – отношение сигнал–шум для опорного сигнала (т.е. сигнала, обладающего единичной амплитудой). Можно показать, что (2.61) точно воспроизводит границу Крамера–Рао, обеспечивая строгую (а не только асимптотическую) эффективность МП оценки амплитуды сигнала. Этот редкий случай строгой оптимальности оценки связан с энергетической природой амплитуды и никогда не встретится далее при рассмотрении неэнергетических параметров.

И в заключении, какие требования к синтезу сигналов предъявляет процедура измерения амплитуды? Как демонстрирует (2.61), точность оценивания непосредственно определяется только энергией сигнала. Никакое усложнение закона модуляции сигнала не способно повысить точность измерения, если только это не связано с энергией сигнала. Следовательно, нет никаких предпосылок для привлечения технологии распределенного спектра при решении данной классической задачи приема.

2.10. Оценка фазы.

Рассмотрим теперь ситуацию, когда параметром, носителем полезной информации, выступает начальная фаза сигнала. Данный случай типичен для когерентных локационных и навигационных приемников, устройств восстановления опорной частоты систем передачи данных с ФМ, демодуляторов приемника систем мобильной радиосвязи 2-го и 3-го поколений, каналов четкости систем телевидения и многих других приложений.

Преобразуем модель полосного сигнала (2.24), выделив постоянную (на интервале наблюдения) начальную фазу φ , которую является неизвестной и подлежащей измерению:

$$s(t; \varphi) = S(t) \cos(2\pi f_0 t + \gamma(t) + \varphi).$$

Поскольку φ является неэнергетическим параметром, то $E(\varphi) = E$ и нахождение

МП оценки фазы заключается в максимизации $z(\varphi) = \int_0^T y(t) s(t; \varphi) dt$ при всех возможных значения $\varphi \in [-\pi, \pi]$. При использовании соотношения (2.59) учтем, что под определением $\rho(\varphi)$ понимается косинус угла между двумя сдвинутыми по фазе на угол φ копиями сигнала, т.е. двумя векторами, разнесенными на угол φ . Тогда $\rho(\varphi) = \cos \varphi$, $\rho''(0) = -1$ и дисперсия МП оценки $\hat{\varphi}$ оказывается равной

$$D\{\hat{\varphi}\} \approx \frac{1}{q^2}, \quad q \gg 1.$$

Снова, как и в случае точного измерения амплитуды, оценивание фазы полностью определяется отношением сигнал-шум. Таким образом, указанная классическая задача приема также нейтральна к закону модуляции при условии постоянства энергии сигнала и, следовательно, *отсутствуют стимулы* к распределению спектра сигнала.

2.11. Автокорреляционная функция и отклик согласованного фильтра.

Теория распределенного спектра в значительной степени базируется на понятии *автокорреляционной функции* (АКФ) сигнала, которая определена как скалярное произведение двух копий одного и того же сигнала, сдвинутых во времени друг относительно друга на τ секунд:

$$R(\tau) = (\mathbf{s}_0, \mathbf{s}_\tau) = \int_{-\infty}^{\infty} s(t) s(t - \tau) dt. \quad (2.62)$$

Задержка сигнала во времени τ является неэнергетическим параметром ($E(\tau) = E$) и, умножая (2.62) на величину E^{-1} , получаем нормированную АКФ, которая есть просто коэффициент корреляции сдвинутых во времени копий сигнала

$$\rho(\tau) = \frac{(\mathbf{s}_0, \mathbf{s}_\tau)}{\|\mathbf{s}\|^2} = \frac{(\mathbf{s}_0, \mathbf{s}_\tau)}{E} = \frac{1}{E} \int_{-\infty}^{\infty} s(t) s(t - \tau) dt. \quad (2.63)$$

Очевидно, что последнее соотношение характеризует скорость уменьшения сходства сдвинутых во времени копий сигнала от рассогласования по задержке τ . В соответствии с общими свойствами коэффициента корреляции $\rho(\lambda)$, приведенными в 2.8, АКФ является четной функцией τ , достигающей максимума в нулевой точке

$$R(\tau) \leq R(0) = E, \quad R(\tau) = R(-\tau) \Leftrightarrow \rho(\tau) \leq \rho(0) = 1, \quad \rho(\tau) = \rho(-\tau). \quad (2.64)$$

С учетом (2.39) и (2.34), не составляет труда убедиться, что для любого полосного сигнала (2.37) АКФ имеет вид

$$R(\tau) = \text{Re} \left[\frac{\dot{R}(\tau)}{2} \exp(j2\pi f_0 \tau) \right], \quad \rho(\tau) = \text{Re}[\dot{\rho}(\tau) \exp(j2\pi f_0 \tau)], \quad (2.65)$$

где

$$\dot{R}(\tau) = (\dot{\mathbf{S}}_0, \dot{\mathbf{S}}_\tau) = \int_{-\infty}^{\infty} \dot{S}(t) \dot{S}^*(t - \tau) dt - \quad (2.66)$$

АКФ комплексной огибающей $\dot{S}(t)$ или, другими словами, закона модуляции. Нормированная версия АКФ (2.66)

$$\dot{\rho}(\tau) = \frac{(\dot{\mathbf{S}}_0, \dot{\mathbf{S}}_\tau)}{\|\dot{\mathbf{S}}\|^2} = \frac{(\dot{\mathbf{S}}_0, \dot{\mathbf{S}}_\tau)}{2E} = \frac{1}{2E} \int_{-\infty}^{\infty} \dot{S}(t) \dot{S}^*(t - \tau) dt, \quad (2.67)$$

является коэффициентом корреляции двух сдвинутых во времени копий комплексной

огибающей $\dot{S}(t)$ и служит (после осуществления операция взятия по модулю) показателем скорости уменьшения сходства между сдвинутым во времени и исходным законом модуляции с ростом рассогласования τ . Как следует из (2.65), АКФ $\rho(\tau)$ полосного сигнала $s(t)$ также может считаться полосным сигналом, закон модуляции которого определяется АКФ $\dot{\rho}(\tau)$ комплексной огибающей $\dot{S}(t)$. В частности, вещественная огибающая $\rho_0(\tau)$ АКФ $s(t)$ является модулем от $\dot{\rho}(\tau)$: $\rho_0(\tau) = |\dot{\rho}(\tau)|$, что иллюстрирует рис. 2.16.

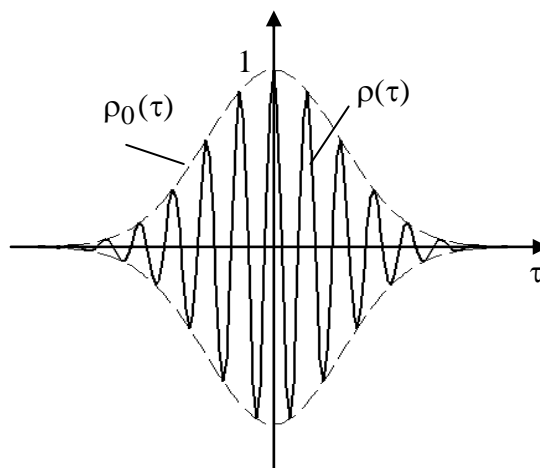


Рис. 2.16. АКФ полосного сигнала.

Реально любая АКФ может быть получена в виде выходного сигнала *коррелятора*, устройства непосредственно выполняющего операции, устанавливаемые соотношениями (2.62) или (2.66). В этом случае вычисления во всем диапазоне значений τ выполняются раздельно по точкам, т.е. последовательно во времени. Альтернативным вариантом служит применение *согласованного фильтра*, линейного устройства с импульсной характеристикой, воспроизводящей зеркальное отражение сигнала: $h(t) = s(T - t)$, где T – длительность сигнала, а несущественный коэффициент пропорциональности положен равным единице. Данный фильтр является неотъемлемой частью оптимального приемника в условиях АБГШ канала, однако его оптимальность превосходит рамки этой специфической модели канала. В частности, для заданного сигнала согласованный фильтр максимизирует выходное отношение сигнал-шум среди всех линейных систем. В рассматриваемом же контексте согласованный фильтр важен своей способностью вычислять и воспроизводить АКФ, как выходной сигнал в реальном времени. Для пояснения этого предположим, что сигнал $s(t)$ подан на вход согласованного с ним фильтра. Тогда отклик фильтра $r(t)$ может быть вычислен как интеграл свертки

$$r(t) = \int_{-\infty}^{\infty} s(\theta) h(t - \theta) d\theta = \int_{-\infty}^{\infty} s(\theta) s(T - t + \theta) d\theta = \int_{-\infty}^{\infty} s(\theta) s[\theta - (t - T)] d\theta = R(t - T) \quad (2.68)$$

и представляет собой АКФ в реальном времени с предсказуемой задержкой, равной длительности сигнала.

Для большей ясности сказанного обратимся к рис. 2.17. Прямоугольный видео импульс $s(t)$ длительности T (см. рис. 2.17, *a*) обладает треугольной АКФ $R(\tau)$ длительностью $2T$ с максимумом в нулевой точке (рис. 2.17, *b*, пунктирная линия). Согласно соотношению (2.68) отклик согласованного фильтра представляет собой копию данной АКФ, сдвинутой на величину длительности сигнала T , так что максимальное напряжение на выходе фильтра отвечает моменту времени, когда входной сигнал уже закончился (рис. 2.17, *b*, сплошная линия). Если бы импульс являлся полосным сигналом с огибающей $s(t)$,

то его АКФ представляла бы собой треугольный полосный импульс (рис. 2.17, *c*, пунктирная линия), а сигнал на выходе согласованного фильтра – сдвинутую на T копию (рис. 2.17, *c*, сплошная линия). Максимум отклика фильтра, согласованного с сигналом, всегда приходится на момент окончания сигнала (по крайней мере, не ранее), поскольку этот фильтр обрабатывает весь сигнал. Следует также отметить, что для полосного сигнала момент достижения максимума огибающей и максимального значения косинуса несущей на выходе согласованного фильтра всегда совпадают, поскольку АКФ всегда максимальна в нулевой точке (см рис. 2.16).

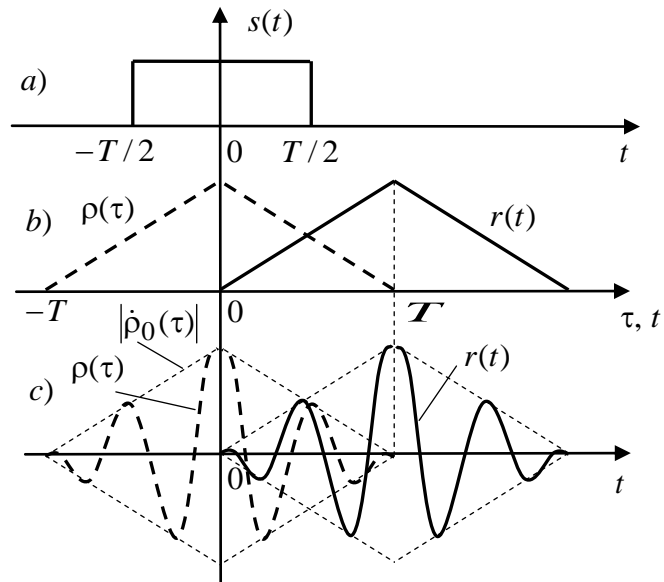


Рис. 2.17. Иллюстрация к определению АКФ и ее формированию СФ.

2.12. Оценка временной задержки полосного сигнала.

2.12.1. Алгоритм оценивания.

Задача, рассматриваемая в этом параграфе, является одной из наиболее часто встречающихся на практике. Она типична для телевизионных систем (каналы синхронизации), цифровых систем мобильной радиосвязи (пилотные каналы, схемы слежения по времени), систем локации (измерение дальности до цели), систем навигации космического и наземного базирования (измерение расстояния до маяков) и т.п. Фактически для адекватной работы любой современной системы обработки информации необходимо восстановить временную шкалу, содержащуюся в принятом колебании, и это именно то, что известно как *оценка задержки по времени*.

Предположим, что полосный сигнал (2.37) $s(t) = \text{Re}[\dot{S}(t)\exp(j2\pi f_0 t)]$, пройдя по каналу связи, приобретает неизвестные задержку во времени τ и начальную фазу φ_0 , т.е. принимает вид

$$s(t; \tau; \varphi_0) = s(t - \tau; \varphi_0) = \text{Re}\{\dot{S}(t - \tau)\exp[j2\pi f_0(t - \tau) + j\varphi_0]\}.$$

Во многих ситуациях фаза φ_0 является случайной и независимой от τ , т.е. не содержит информации об объекте интереса – временной задержке τ . Включим фазовый компонент, обусловленный задержкой τ в общую начальную фазу $\varphi = -2\pi f_0 \tau + \varphi_0$. Последняя величина остается случайной и, будучи распределенной на интервале $[-\pi, \pi]$, вновь оказывается независимой от τ , т.е. не содержащей информации о временном сдвиге, вследствие деструктивного вклада φ_0 . Тогда принятый сигнал может быть представлен как

$$s(t - \tau; \varphi) = \text{Re}[\dot{S}(t - \tau)\exp(j2\pi f_0 t)\exp(j\varphi)], \quad (2.69)$$

где задержка во времени τ является неизвестным полезным параметром, подлежащим измерению, а φ – бесполезная начальная фаза, неопределенность которой может только усложнить измерение τ .

Как было установлено в 2.8, любая процедура оценивания является частным случаем различения сигналов. В рассматриваемой ситуации следует сделать выбор между многочисленными копиям сигнала (2.69), которые отличаются друг от друга значением временного сдвига τ , тогда как *мешающий (nuisance)* параметр – начальная фаза φ – находится вне зоны внимания. К счастью, существует прямой способ преодоления неопределенности сигнала, обусловленную случайностью φ : в параграфе 2.5 было показано, что оптимальный выбор между некогерентными сигналами может быть заменен выбором между их детерминированными законами модуляции, т.е. комплексными огибающими. Следовательно, при оценивании задержки необходимо сравнивать сдвинутые во времени копии комплексной огибающей сигнала $\dot{S}(t)$, одна из которых и считается истинной. Временной сдвиг выбранной копии и принимается в качестве МП оценкой $\hat{\tau}$ времени задержки. Очевидно, что предпочтение, отдаваемое данной копии сигнала перед остальными, основывается на ее минимальном расстоянии от принятой комплексной огибающей $\dot{Y}(t)$ или, учитывая неэнергетический характер времени задержки, на максимуме корреляции с $\dot{Y}(t)$. Данное значение корреляции оценивается как абсолютная величина (2.47), которая может быть переписана с учетом того, что роль сигнала с номером i теперь принадлежит некоторому значению τ :

$$Z(\tau) = \left| \int_0^T \dot{Y}(t) \dot{S}^*(t - \tau) dt \right|. \quad (2.70)$$

Основанное на данной характеристике правило оценивания $Z(\hat{\tau}) = \max_{\tau} Z(\tau)$ совершенно прозрачно физически: МП оценка $\hat{\tau}$ есть просто значение времени запаздывания, при котором закон модуляции сигнала наиболее подобен с законом модуляции наблюдения.

Одним из возможных вариантов реализации указанного правила оценивания является использование набора корреляторов, представленных на рис. 2.18. Комплексная огибающая наблюдаемого колебания обрабатывается в M параллельных корреляторах,

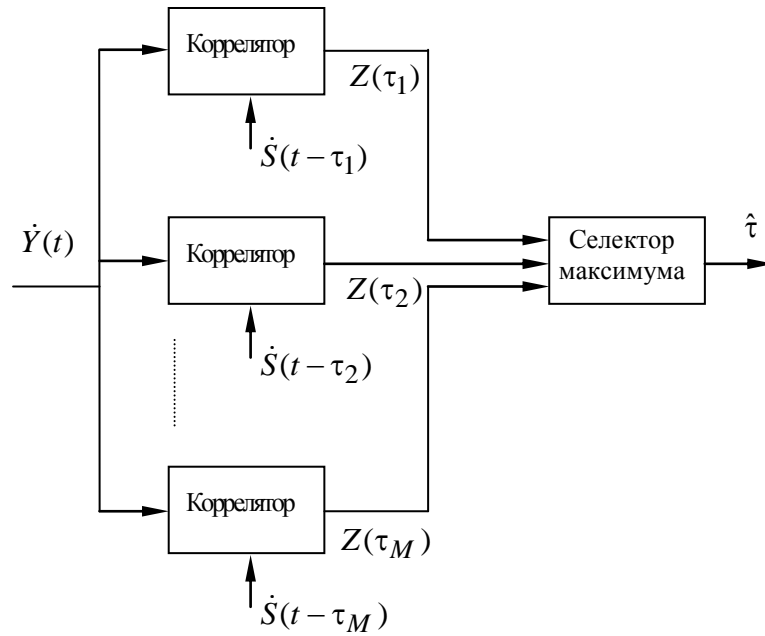


Рис. 2.18. Реализация МП оценки запаздывания набором корреляторов.

опорными сигналами которых служат сдвинутые во времени копии комплексной огибающей сигнала. С выходов корреляторов снимаются значения $Z(\tau_i), i = 1, 2, \dots, M$, которые поступают в оконечный блок, где производится их сравнение и выбор наибольшего из них. В качестве МП оценке выбирается задержка того опорного сигнала коррелятора, на выходе которого достигается максимальное значение.

Очевидно, что данная структура применима в ситуации, когда время запаздывания принимает только дискретные значения. В противном случае необходимо осуществлять квантование непрерывной величины τ и использовать такое количество корреляторов (равное числу дискретных значений τ_i), при котором ошибка квантования была бы достаточно малой.

Альтернативный вариант построения измерителя времени запаздывания предполагает использование согласованного фильтра. Пусть наблюдение $y(t)$ поступает на вход фильтра, согласованного с сигналом $s(t)$. Определим выходное колебание $r(t)$ с помощью интеграла свертки и импульсной характеристики фильтра $h(t) = s(T - t)$ в виде

$$r(t) = \int_{-\infty}^{\infty} y(\theta) h(t - \theta) d\theta = \int_{-\infty}^{\infty} y(\theta) s(T - t + \theta) d\theta.$$

Данный интеграл, представляющий собой скалярное произведение $y(t)$ и $s(T - t + \theta)$, может быть вычислен с помощью (2.39) и (2.34) как

$$r(t) = \int_{-\infty}^{\infty} y(\theta) s(T-t+\theta) d\theta = \frac{1}{2} \operatorname{Re} \left[\int_0^T \dot{y}(\theta) \dot{s}^*(T-t+\theta) d\theta \right] =$$

$$= \operatorname{Re} \left\{ \left[\frac{1}{2} \int_0^T \dot{Y}(\theta) \dot{S}^*(T-t+\theta) d\theta \cdot \exp(-j2\pi f_0 T) \right] \exp(j2\pi f_0 t) \right\}.$$

Сравнивая полученный результат с обобщенной моделью полосного сигнала (2.37), можно увидеть, что выражение в квадратных скобках представляет собой ничто иное, как комплексную огибающую выходного сигнала фильтра. Следовательно, вещественная огибающая выходного сигнала (закон амплитудной модуляции), определенная как абсолютная величина выражения в квадратных скобках

$$\left| \frac{1}{2} \int_{-\infty}^{\infty} \dot{Y}(\theta) \dot{S}^*[\theta - (t-T)] d\theta \right| = \frac{1}{2} Z(t-T), \quad (2.71)$$

т.е. повторяет в реальном времени (с несущественным множителем в 1/2) копию модуля корреляции (2.70), сдвинутую во времени на известную длительность сигнала. Данный факт непосредственно указывает на возможную структуру построения МП измерителя временной задержки τ , представленную на рис. 2.19, *a*. Первоначально наблюдение $y(t)$ обрабатывается согласованным фильтром, выходной сигнал с которого подается на детектор огибающей. Последний блок предложенной структуры фиксирует момент времени t_m , когда сигнал $r_d(t)$ на выходе детектора достигает максимального значения. В итоге МП оценка $\hat{\tau}$ получается как результат вычитания из t_m известной константы T (см. рис. 2.19, *b*).

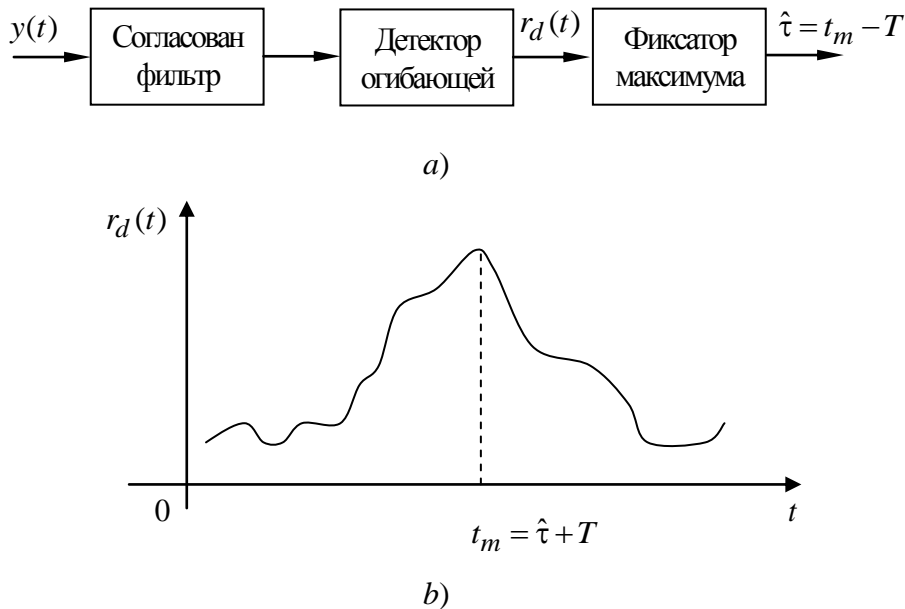


Рис. 2.19. Реализация МП оценки временной задержки на основе СФ.

Схема, представленная на рис. 2.19, представляется более прозрачной для пояснения идеи построения измерителя, однако на практике многие программно–реализуемые устройства оценки могут оказаться не столь прямо отождествленными с одной из двух рассмотренных структур.

2.12.2. Точность оценивания.

Согласно соотношению (2.59), дисперсия оценки τ зависит от скорости, с которой сдвинутая во времени копия сигнала теряет свое сходство с исходным образцом. Однако в случае сигналов со случайной начальной фазой получение МП оценки связано со сравнением детерминированных комплексных огибающих. Степень похожести рассогласованных во времени копий комплексной огибающей характеризуется огибающей АКФ сигнала (2.67)

$$\rho_0(\tau) = |\dot{\rho}(\tau)| = \left| \frac{1}{2E} \int_{-\infty}^{\infty} \dot{S}(t) \dot{S}^*(t - \tau) dt \right|, \quad (2.72)$$

и поэтому ее крутизна влияет на дисперсию ошибки измеряемого времени запаздывания, что может быть строго обосновано фактически. Действительно, как следует из соотношений (2.70)–(2.71), модуль АКФ (2.72) представляет собой (с точностью до пренебрежимо малого множителя) свободную от шума огибающую на выходе согласованного фильтра, принимающую максимальное значение в некоторый «истинный» момент t_{m0} (см. рис. 2.20, *a*, сплошная линия). В присутствии шума временное положение максимума t_m флюктуирует относительно истинного значения (см. рис. 2.20, *a*, пунктирная линия), охватывая диапазон, зависящий от крутизны сигнала на выходе фильтра, т.е. модуля АКФ (2.72).

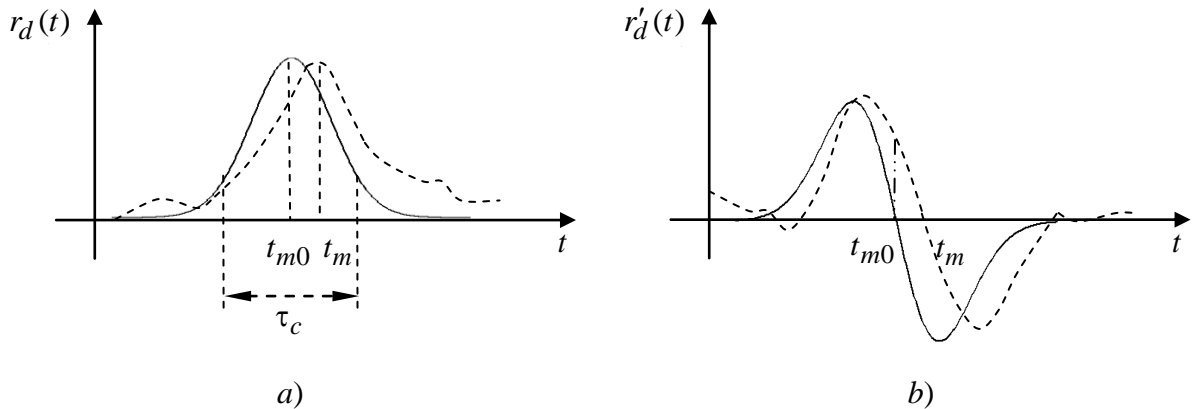


Рис. 2.20. Иллюстрация точности измерения времени.

Для более наглядной демонстрации последнего факта примем во внимание, что фиксация максимума $r_d(t)$ на выходе детектора равносильна регистрации временной точки, в которой производная $r'_d(t)$ пересекает нулевой уровень (принимая соответствующие меры, исключая выбор ошибочных «нулевых точек», которые обусловлены наличием возможных локальных максимумов). Данную ситуацию иллюстрирует рис. 2.20, *b*. При достаточно большом отношении сигнал-шум отклонение $\varepsilon = t_m - t_{m0}$ мало и можно полагать, что в присутствии шума кривая $r'_d(t)$ (пунктирная линия) линейна на интервале $[t_{m0}, t_m]$, имея такой же наклон, как и кривая (сплошная линия) в точке t_{m0} в отсутствие шума. Тогда из треугольника на рис. 2.20, *b*, значение ε может быть определено, как результат деления длины штрих-пунктирной линии на крутизну кривой $r'_d(t)$ в отсутствие шума в точке t_{m0} , т.е. на вторую производную $r''_d(t_{m0})$. В свою очередь последнее в точности есть $\rho''(0)$, так что $\varepsilon \approx r'(t_{m0})/\rho''(0)$. С другой стороны, дисперсия отклонения $r'_d(t_{m0})$ относительно нулевой точки в отсутствие шума тем больше, чем меньше отно-

шение сигнал-шум и чем выше скорость случайного изменения сигнала $r_d(t)$ на выходе детектора в присутствии шума (штриховая линия на рис.2.20, а). Именно острота АКФ случайного процесса свидетельствует о скорости его изменения и, значит, АКФ случайного процесса на выходе детектора при большом отношении сигнал-шум повторяет огибающую АКФ на входе детектора. В случае, когда фильтр согласован с сигналом, упомянутая огибающая есть ничто иное, как огибающая $\rho_0(\tau)$ АКФ сигнала. Поскольку острота любой АКФ определяется своей второй производной в нулевой точке со знаком минус, то дисперсия $r'_d(t_{m0})$ пропорциональна $-\rho''_0(0)$. Тогда

$$D\{\varepsilon\} = \frac{D\{r'_d(t_{m0})\}}{|\rho''(0)|^2} = \frac{1}{-\rho''_0(0)}$$

обратно пропорциональна остроте АКФ сигнала $-\rho''_0(0)$, как это и было предсказано ранее и как следует из границы Крамера-Рао (2.59).

В итоге приходим к следующему важному заключению: точность временного измерения полностью определяется *остротой* (*sharpness*) основного пика АКФ сигнала и чем острее АКФ, тем меньше дисперсия МП оценки времени запаздывания τ .

Наряду с $-\rho''_0(0)$ в качестве индикатора остроты АКФ сигнала может использоваться характеристика, называемая *протяженностью корреляции*, или *временем корреляции* (*correlation spread*) и обозначаемая через τ_c . Указанный параметр характеризует ширину АКФ сигнала (см. рис. 2.20, а) и, точно также как длительность и полоса сигнала, данный параметр необходимо определять по некоторому согласованию, поскольку АКФ может иметь достаточно сложную форму и спадать до нуля только асимптотически. В свете определения АКФ будем полагать, что копии сигнала (или копии комплексной огибающей), сдвинутые во времени на величину $\tau < \tau_c$, имеют значительное сходство, тогда как при $\tau > \tau_c$ их похожесть пренебрежимо мала. Очевидно, что приведенное здесь заключение может быть сформулировано с использованием новой характеристики: сигналы и узкой АКФ, т.е. *малым* временем корреляции, являются предпочтительными для осуществления точного оценивания временной задержки.

Продолжая, можно вспомнить одно из основных положений спектрального анализа, непосредственно вытекающее из (2.72) после применения теоремы Парсеваля: АКФ и энергетический спектр сигнала связаны друг с другом преобразованием Фурье. В терминах комплексной огибающей, имеющего вид

$$\rho_0(\tau) = \frac{1}{2E} \left| \int_{-\infty}^{\infty} \tilde{S}(f) \exp(j2\pi f\tau) df \right|^2,$$

где спектр комплексной огибающей $\tilde{S}(f)$ фактически (пренебрегая коэффициентом пропорциональности) совпадает со спектром полосного сигнала, перенесенного в низкочастотную область. Тогда, согласно общему свойству преобразования Фурье, $\tau_c \approx 1/W$, или чем уже АКФ, тем шире спектр сигнала и наоборот. Непосредственным следствием этого служит следующее утверждение: возможный путь повышения точности временного измерения состоит в использовании сигналов с широким спектром. Формально прийти к этому же заключению можно путем дифференцирования последнего выражения АКФ огибающей и подстановкой полученного результата в границу Крамера-Рао (2.59). После нескольких утомительных вычислений можно прийти к формуле, на которую часто ссылаются как формулу Вудворда

$$D\{\hat{\tau}\} \approx \frac{1}{(2\pi W_{rms})^2 q^2}, \quad q \gg 1,$$

где использовано понятие *среднеквадратической* (*root-mean-square* (rms)) ширины полосы

сигнала, определяемое как

$$W_{rms} = \frac{1}{2E} \int_{-\infty}^{\infty} f^2 |\tilde{S}(f)|^2 df .$$

Может оказаться не так просто понять идею, почему подобное измерение связано с шириной спектра. В данном случае благоразумно будет провести аналогию на основе более привычного вероятностно распределенного параметра. Дисперсия случайной величины x , обладающей нулевым средним и плотностью распределения $W(x)$, определяется как $D\{x\} = \int_{-\infty}^{\infty} x^2 W(x) dx$, характеризуя степень разброса x относительно ее математического ожидания, или, что эквивалентно, ширину плотности распределения $W(x)$. Однако нормированный энергетический спектр $|\tilde{S}(f)|^2 / 2E$ неотрицателен и удовлетворяет условию нормировки $\int_{-\infty}^{\infty} \left[|\tilde{S}(f)|^2 / 2E \right] df = 1$, т.е. может трактоваться как плотность распределения некоторой подходящей случайной переменной. Тогда W_{rms} является мерой ее разброса и, следовательно, шириной энергетического спектра.

Следовательно, очень грубое, промежуточное требование к сигналам, пригодным для использования в задачах временного измерения, может быть сформулировано следующим образом. Сигналы с узкой АКФ (малым временем корреляции τ_c) или, что эквивалентно, широкой полосой W представляют основной интерес. В сопоставлении двум ранее рассмотренным измерительным задачам ситуация выглядит достаточно по-новому: существует значительный ресурс для совершенствования достоверности оценивания вне прямого пути, т.е. не за счет увеличения энергии.

Следует подчеркнуть теперь, что термины «широкополосный» и «распределенный спектр» не являются синонимами. Действительно, если соглашения, принятые относительно определений времени корреляции τ_c и длительности сигнала T , достаточно справедливы, сигнальные копии, сдвинутые во времени более, чем на T , практически не перекрываются, т.е. их скалярное произведение, а значит, и сходство незначительно. Следовательно, условие $\tau_c \leq T$ показывает, что сужение АКФ сигнала и, значит, расширение его полосы может быть достигнуто тривиальным укорочением самого сигнала.

Однако, двигаясь указанным путем, не следует забывать, что отношение сигнал-шум зависит от энергии сигнала $E = k_p PT$, где P – мгновенная мощность сигнала, а k_p – коэффициент, определяемый формой сигнала. Очевидно, что сохранение постоянства отношения сигнал-шум при укорочении сигнала требует пропорционального увеличения пиковой мощности. Следовательно, добиваясь все более и более высокой точности, можно прийти к ситуации, когда необходимая мгновенная мощность станет чрезмерно высока. Как правило, требование очень высокой излучаемой мощности влечет за собой необходимость увеличения массо-габаритных характеристик передающей аппаратуры и источника энергии. В дополнении к этому, мощные короткие импульсы могут нанести существенный вред соседствующим системам, а также экологии окружающей среды.

Более элегантный способ повышения точности временных измерений обещает путь, основанный на том факте, что укорочение сигнала не является единственным методом расширения спектра или, что эквивалентно, уменьшения времени корреляции. Рассмотрим сигнал, длительность T которого достаточно велика для обеспечения необходимой энергии, т.е. отношения сигнал-шум, в сочетании с приемлемым значением мгновенной мощности P . Предположим, что определен закон внутренней угловой модуляции, обеспечивающий время корреляции сигнала значительно меньшее его длительности, т.е. $\tau_c \ll T$. Тогда АКФ сигнала имеет ярко выраженный остроконечный характер, обеспечи-

вая высокую точность оценивания временной задержки, несмотря на большую длительность самого сигнала. Но в свете существующей зависимости между временем корреляции и полосой $\tau_c \approx 1/W$ неравенство $\tau_c \ll T$ означает, сигнал характеризуется большой величиной частотно-временного произведения $WT \gg 1$, т.е. является сигналом с распределенным спектром. Вставая на упомянутый путь, видно, что привлечение технологии распределенного спектра позволяет снять противоречие между величиной мгновенной мощности и точностью оценивания: необходимая энергия вкладывается в сигнал за счет его длительности, а не мощности, тогда как высокая точность измерения достигается посредством синтеза соответствующего закона модуляции.

Фактически выполнение условия $\tau_c \ll T$ означает, что длинный во времени сигнал становится коротким после обработки согласованным фильтром в устройстве оценивания, представленном на рис. 2.19, а. Должно быть ясно, что эффект *временного сжатия* достигается в этом согласованном фильтре только при использовании сигналов с распределенным спектром. В принципе любой сигнал может быть укорочен во времени в некотором специальном образом построенном (обобщенном согласованном) фильтре, например эквалайзере. Однако для простых сигналов подобное укорочение достигается ценой потерь в отношении сигнал-шум, и только сигналы с распределенным спектром обеспечивают наилучшую «очистку от шумов» одновременно с временной компрессией. В то же время следует понимать, что условие распределенности спектра $WT \gg 1$ является только необходимым и нахождение сигналов, сочетающих большую собственную длительность с острой АКФ, является более чем сложной задачей. Подробному обсуждению данной проблемы будет посвящена глава 5.

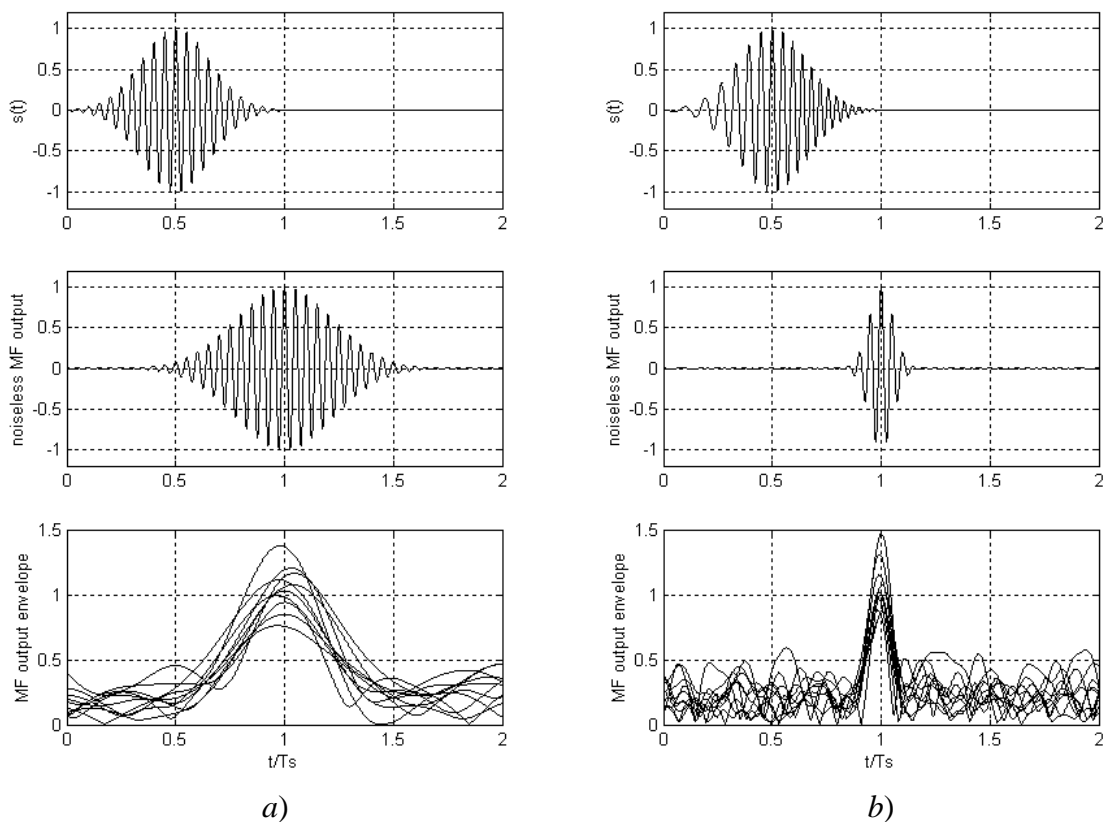


Рис. 2.21. Иллюстрация временного оценивания и эффекта сжатия СФ.

Обратимся теперь к рис. 2.21, на котором представлены диаграммы колебаний,

смоделированные в среде Matlab. Рис.2.21, *a* содержит диаграммы, которые отвечают соответственно простому полосному сигналу с гладкой огибающей, отклику согласованного с ним фильтра (АКФ с временным сдвигом) и 10 различным реализациям на выходе детектора в условиях действия шума. Диаграммы, представленные на рис. 2.21, *b*, демонстрируют аналогичные колебания, но отвечающие сигналу с распределенным спектром, который представляет собой импульс с линейной частотной модуляцией, имеющий форму, длительность и энергию идентичные сигналу на рис. 2.21, *a*. Из представленных диаграмм следует, что в варианте *b*) явным образом проявляется эффект временной компрессии, который приводит к значительно меньшему диапазону флюктуаций временной позиции максимума сигнала на выходе детектора по сравнению со случаем *a*). Данный пример убедительно свидетельствует о способности сигналов с распределенным спектром повысить точность временных измерений, не вызывая противоречий с пиковой мощностью.

Теперь сформулируем следующее заключение. В случае отсутствия ограничений на величину пиковой мощности классическая задача измерения времени запаздывания *не требует обязательного привлечения* технологии распределенного спектра. Однако при наложении жестких требований к пиковой мощности указанная технология *является крайне необходимой*. Отметим кратко, что последняя ситуация совершенно типична для импульсной радиолокации, что объясняет причины, по которым эта область применения в течение многих десятилетий стимулировала развитие технологии распределенного спектра.

2.13. Оценка несущей частоты.

Рассмотрим теперь ситуацию, при которой неизвестным информативным параметром служит несущая частота сигнала. Данная задача имеет такое же широкое распространение, как и предыдущая. Она характерна для таких приложений, как радиолокация, где измерение скорости цели основано на эффекте доплеровского смещения частоты (например, полицейский радар, осуществляющий мониторинг дорожного движения); устройства восстановления опорной частоты в приемниках систем мобильной радиосвязи второго и третьего поколений; системы автоматического контроля частоты в ФМ и ТВ приемниках и т.д.

На практике типична ситуация, когда необходимо измерить лишь смещение F от номинального значения несущей частоты f_0 , тогда модель сигнала может быть записана как

$$s(t; F; \varphi) = \operatorname{Re}\left\{\dot{S}(t) \exp[j2\pi(f_0 + F)t + j\varphi]\right\} = \operatorname{Re}\left[\dot{S}(t; F) \exp(j2\pi f_0 t) \exp(j\varphi)\right],$$

где $\dot{S}(t; F) = \dot{S}(t) \exp(j2\pi Ft)$ – комплексная огибающая сигнала, включающая линейный фазовый дрейф, обусловленный частотным сдвигом F , а φ – как и ранее, мешающий параметр – случайная начальная фаза, не содержащая информации о F .

Вследствие неэнергетической природы F правило максимума корреляции уместно записать в виде $Z(F) = \max_F Z(F)$, где модуль корреляция $Z(F)$ снова образуется в результате незначительной адаптации (2.47), учитывающей замену i некоторым значением F :

$$Z(F) = \left| \int_0^T \dot{Y}(t) \dot{S}^*(t; F) dt \right| = \left| \int_0^T \dot{Y}(t) \dot{S}^*(t) \exp(-j2\pi Ft) dt \right|.$$

Устройство, изображенное на рис.2.18 может быть применено для реализации данной оптимальной процедуры оценивания, если в качестве опорных сигналов коррелятора использовать частотно-сдвинутые копии комплексной огибающей $\dot{S}(t; F_i) = \dot{S}(t) \exp(j2\pi F_i t)$, $i = 1, 2, \dots, M$, вместо сдвинутых во времени копий. С другой

стороны, из соотношения (2.71) видно, что амплитуда сигнала на выходе согласованного фильтра, настроенного на частоту $f_0 + F_i$, принимает значение $Z(F_i)$ (пренебрегая коэффициентом) в момент окончания сигнала T . Таким образом, может быть использована альтернативная структура (см. рис. 2.22), содержащая набор из M согласованных фильтров, причем i -й фильтр должен быть настроен на частоту, отличающуюся от номинальной f_0 , на величину F_i . После амплитудного детектирования и взятия отсчета в момент времени T на выходах детекторов имеется множество значений $Z(F_i)$, наибольшее из которых выбирается для получения МП оценки \hat{F} как значение частотного рассогласования фильтра, отвечающего максимуму $Z(F_i)$. Ясно, что в случае непрерывного параметра F необходима операция квантования, а число фильтров M определяется допустимой ошибкой квантования.

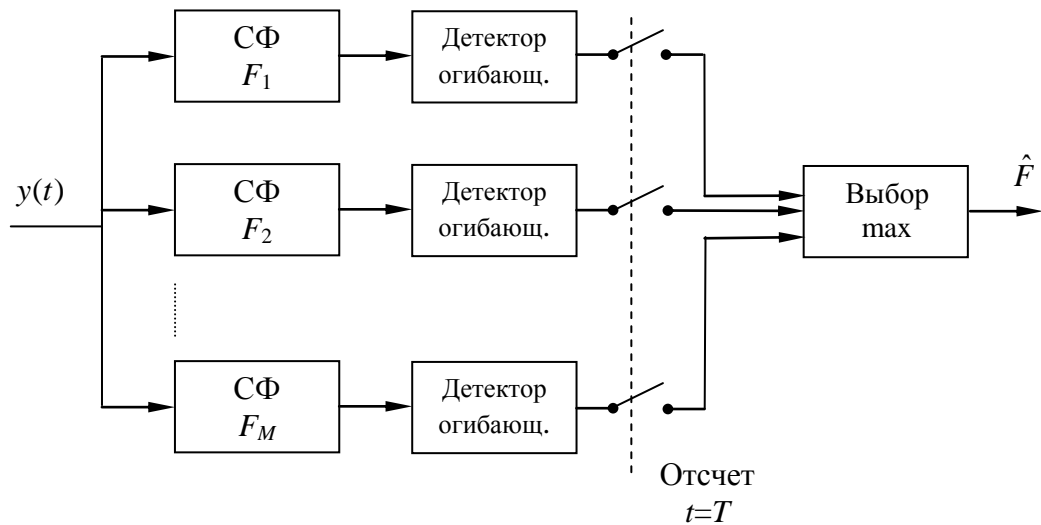


Рис. 2.22. Реализация МП оценки частоты на основе СФ.

Согласно соотношению (2.59) точность измерения частоты определяется не только отношением сигнал-шум, но также и остротой корреляции между сдвинутыми по частоте копиями сигнала в зависимости от их взаимного частотного рассогласования. В связи со случайностью фазы рассмотрению подлежат комплексные огибающие, и сходство между их частотно-сдвинутыми копиями $\dot{S}(t; 0)$ и $\dot{S}(t; F)$ характеризуется модулем коэффициента корреляции (2.44)

$$\begin{aligned} \rho_0(F) = |\dot{\rho}(F)| &= \left| \frac{1}{2E} \int_{-\infty}^{\infty} \dot{S}(t; 0) \dot{S}^*(t; F) dt \right| = \left| \frac{1}{2E} \int_{-\infty}^{\infty} |\dot{S}(t)|^2 \exp(-j2\pi Ft) dt \right| = \\ &= \left| \frac{1}{2E} \int_{-\infty}^{\infty} S^2(t) \exp(-j2\pi Ft) dt \right|. \end{aligned} \quad (2.73)$$

Как видно, зависимость коэффициента корреляции как функции F повторяет по форме амплитудный спектр квадрата вещественной огибающей сигнала. Поскольку вещественная огибающая, т.е. закон амплитудной модуляции, всегда является низкочастотным сигналом, то из свойств преобразования Фурье следует, что, чем длиннее сигнал, тем острее $\rho_0(F)$. Подходящим названием протяженности $\rho_0(F)$ вдоль оси частот может служить термин «ширина огибающей по частоте» (*envelope frequency spread*). Обозначив данный параметр через F_e , из только что упомянутого факта следует, что $F_e \approx 1/T$. Таким образом, помимо «прямого» способа, т.е. увеличением энергии, можно повысить точность

измерения частоты за счет применения сигнала, обладающего достаточно компактным спектром огибающей (малым значением F_e), т.е. достаточно большой длительностью T . Формально этот вывод снова может быть получен на основе оценки второй производной $\rho(F)$ и подстановки результата в границу Крамера-Рао (2.59)

$$D\{\hat{F}\} = \frac{1}{(2\pi T_{rms})^2 q^2},$$

где среднеквадратическая длительность T_{rms} характеризует протяженность сигнала во времени точно так же, как полоса W_{rms} описывает протяженность спектра сигнала.

В заключение можно упомянуть о двух физических аспектах приведенной выше точки зрения. Во-первых, повышение точности за счет увеличения длительности T легко объяснимо: частота может рассматриваться как скорость изменения угла полной фазы, и, как и любая скорость, измеряется более надежно тогда, когда приращение угла наблюдается в течение большого интервала. Во-вторых, в сравнении частотного и временного измерений явным образом проявляется частотно-временная дуальность. Действительно, в то время как точность временного измерения определяется протяженностью сигнала в частотной области (шириной полосы W), точность частотного измерения контролируется протяженностью сигнала во временной области (длительностью T).

Приведенный выше результат приводит нас к выводу, что в случае, когда единственным информативным параметром является частота сигнала, *отсутствуют импульсы* к привлечению технологии распределенного спектра, поскольку, помимо увеличения энергии, только длительность сигнала оказывает влияние на точность оценивания. Данный факт объясняет частое применение в доплеровских радарх измерения скорости простейших немодулированных непрерывных гармонических сигналов.

2.14. Одновременное оценивание времени запаздывания и частоты.

В рамках следующей задачи имеет место ситуация, когда как временной τ , так и частотный F сдвиги принятого сигнала являются неизвестными и информативными параметрами, т.е. подлежащими измерению. Данная постановка охватывает многие практические ситуации. В цифровой связи, например, системах мобильной связи 2-го и 3-го поколений, процедура приема, как правило, начинается с синхронизации собственного опорного сигнала с принятым. Данная операция заключается в измерении частотно-временного рассогласования собственного генератора с принятым сигналом и последующей частотно-временной подстройки первого со вторым до полной их синхронизации. В радиолокации одновременное определение расстояния и скорости цели относительно приемника требует оценивания временного запаздывания и доплеровского смещения частоты сигнала. В навигации, например системе GPS, во многом аналогичные измерения служат исходными данными для определения местоположения и скорости пользователей. Перечень примеров может быть продолжен и далее.

В отличие от материала, рассмотренного в 2.9-2.13, параметром, который необходимо оценить, является двумерный вектор $\lambda = (\tau, F)$, а не скаляр. Соответственно и модель принятого сигнала сочетает в себе особенности обоих его представлений, описанных в 2.12 и 2.13:

$$s(t; \tau, F; \varphi) = \operatorname{Re}[\dot{S}(t; \tau, F) \exp(j2\pi f_0 t) \exp(j\varphi)],$$

где $\dot{S}(t; \tau, F) = \dot{S}(t - \tau) \exp(j2\pi Ft)$ – комплексная огибающая, включающая в себя как время запаздывания, так и частотный сдвиг; а φ – как и ранее, не информативная начальная фаза. Очевидно, что теперь правило максимума корреляции для оценки τ, F формулируется с использованием модуля корреляции вида

$$Z(\tau, F) = \left| \int_0^T \dot{Y}(t) \dot{S}^*(t; \tau, F) dt \right| = \left| \int_0^T \dot{Y}(t) \dot{S}^*(t - \tau) \exp(-j2\pi Ft) dt \right|, \quad (2.74)$$

которое показывает, насколько близко комплексная огибающая сигнала $\dot{S}(t; \tau, F)$, сдвинутая по времени и частоте, подобна наблюдаемой комплексной огибающей $\dot{Y}(t)$. Определение τ и F , максимизирующих указанное подобие, обеспечивает получение пары МП оценок $\hat{\tau}, \hat{F}$, а именно $Z(\hat{\tau}, \hat{F}) = \max_{\tau, F} Z(\tau, F)$.

Как и ранее, построение устройства оценивания возможно на основе набора корреляторов, опорными сигналами которых служат копии $\dot{S}(t; \tau, F) = \dot{S}(t - \tau) \exp(j2\pi Ft)$ с различными значениями τ и F . Однако более практичной является схема с использованием согласованных фильтров, которая приведена на рис. 2.23 и объединяет структуры на рис. 2.22 и 2.19, а. Вид ее непосредственно определяется результатом сравнения соотношений (2.74) и (2.71), которое показывает, что $Z(\tau, F)$ (без учета постоянного множителя) повторяет вещественную огибающую на выходе согласованного фильтра, имеющего расстройку по частоте, равную F . В схеме используются M ветвей, содержащих фильтр и детектор. Каждая из ветвей настроена на свою конкретную частоту, обеспечивая тем самым оценивание F , тогда как оценка τ получается фиксированием точки максимума сигнала с выхода детектора. Блок «Выбор max» одновременно выполняет обе указанные операции, фиксируя момент достижения глобального максимума среди выходных значений всех ветвей. Затем это значение времени (после вычитания длительности сигнала) принимается в качестве оценки $\hat{\tau}$, тогда как частота настройки ветви, на выходе которой регистрируется глобальный максимум, дает оценку \hat{F} .

Точность оценивания зависит от скорости, с которой падает похожесть между рассогласованными по времени и частоте копиями $\dot{S}(t; 0, 0)$ и $\dot{S}(t; \tau, F)$ комплексной оги-

бающей сигнала с ростом τ, F . Иными словами, точность определяется остротой пика модуля корреляционного коэффициента, задаваемого (2.44)

$$\rho_0(\tau, F) = |\dot{\rho}(\tau, F)| = \left| \frac{1}{2E} \int_{-\infty}^{\infty} \dot{S}(t; 0, 0) \dot{S}^*(t; \tau, F) dt \right| =$$

$$= \left| \frac{1}{2E} \int_{-\infty}^{\infty} \dot{S}(t) \dot{S}^*(t - \tau) \exp(-j2\pi Ft) dt \right|, \quad (2.75)$$

как функции двух переменных τ, F . Эта функция, часто называемая *функцией неопределенности* (ФН) Вудворда, играет исключительно важную роль в теории сигналов. Геомет-

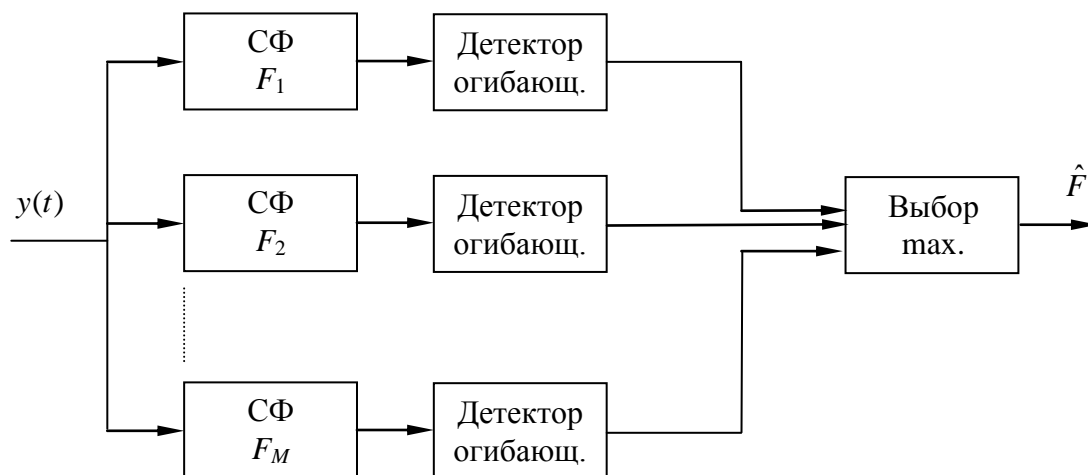


Рис. 2.23. Схема МП оценивания временной задержки и частоты на основе СФ.

рически ее можно изобразить в виде трехмерной поверхности над плоскостью τ, F , имеющей максимум, равные единице, в начале координат: $\rho_0(\tau, F) \leq \rho_0(0, 0) = 1$. Примерный вид функции неопределенности дает рис. 2.24.

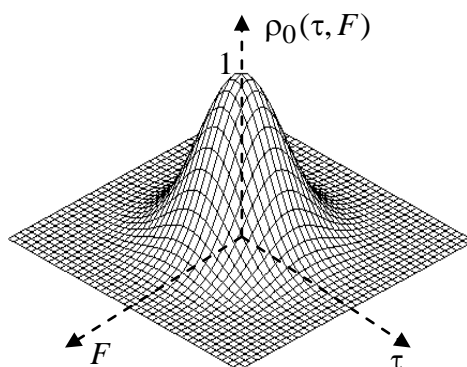


Рис. 2.24. Пример функции неопределенности.

При $F = 0$ сигнальные копии рассогласованы только во времени и, в соответствие с этим, функция неопределенности (2.75) превращается в АКФ (2.72), взятую по абсолютной величине: $\rho_0(\tau, 0) = \rho_0(\tau)$. С другой стороны, в отсутствии временного сдвига $\tau = 0$ сигнальные копии рассогласованы только по частоте, и функция неопределенности становится спектром (2.73) квадрата амплитудного закон модуляции сигнала: $\rho_0(0, F) = \rho_0(F)$. Другими словами, АКФ сигнала и спектр квадрата огибающей представляют собой сечения функции неопределенности вертикальными плоскостями $\tau = 0, F = 0$ соответственно. Возвращаясь к схеме, представленной на рис.2.23, можно заметить, что в случае отсутствия шума сигнал в i -м канале воспроизводит в реальном времени сечение функции неопределенности вертикальной плоскостью $F = F_i$.

Как следует из вышеприведенных рассуждений единственным свободным от энергетических затрат источником повышения точности частотно-временного оценивания служит остроты функции неопределенности. ФН должна достаточно быстро спадать в любом направлении плоскости τ, F при отсутствии априорных сведений о возможных значениях τ, F . Одним из практических способов, пригодных для характеристики крутизны $\rho_0(\tau, F)$, является использование ее горизонтального сечения (называемого *диаграммой неопределенности (ambiguity diagram)*) на некотором фиксированном уровне, например, равном 0.5. Рис.2.25 служит примером такого сечения. Поскольку протяженность двух основных сечений $\rho_0(\tau)$ и $\rho_0(F)$ определяется временем корреляции ($\tau_c \approx 1/W$) и частотной протяженностью огибающей ($F_e \approx 1/T$), то эти характеристики автоматически определяют размеры сечения по осям τ, F . Очевидно, что чем острее функция неопределенности, тем меньше площадь диаграммы неопределенности. Последняя пропорциональна произведению $\tau_c F_e \approx 1/WT$, что влечет за собой следующее утверждение: *только сигналы с распределенным спектром позволяют повысить точность оценивания временного запаздывания (частоты) без ухудшения точности измерения частоты (времени запаздывания)*. Действительно, для любого простого сигнала $WT \approx 1$ и, следовательно, $\tau_c F_e \approx 1$, так что достичь значительной крутизны функции неопределенности вдоль одного направления (например, τ) невозможно иначе, как одновременным растяжением ее вдоль другого

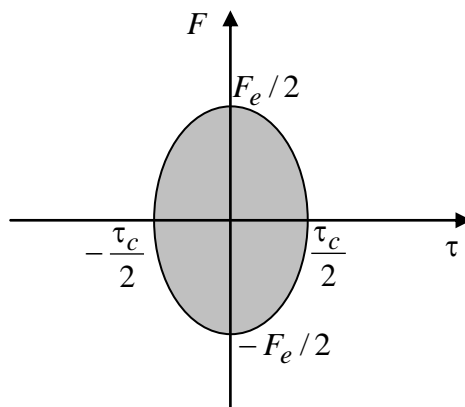


Рис. 2.25. Сечение функции неопределенности.

направления (F).

При использовании сигналов с распределенным спектром существует возможность избежать противоречия между протяженностями τ_c, F_e функции неопределенности или, что эквивалентно, между длительностью сигнала и шириной его полосы, аналогично тому, как это имело место в 2.12. Применение особым образом выбранного закона угловой

модуляции определяет полосу сигнала W (время корреляции τ_c) и посредством этого точность временного измерения, тогда как выбор соответствующей длительности сигнала может быть осуществлен независимо, что гарантирует необходимую точность частотного измерения.

Сравнение утверждения, к которому только что пришли, со сделанными в предыдущих главах, показывает, что среди всех до сих пор рассмотренных классических задач приема одновременное оценивание времени и частоты является первой, в которой обращение к философии распределенного спектра является категорически необходимым. Не существует других средств для бесконфликтного выполнения требований высокой точности оценивания, как по времени, так и по частоте иных, чем привлечение сигналов с распределенным спектром.

2.15. Разрешение сигналов.

Для многих практических приложений чрезвычайно типичной является ситуация, когда принятый сигнал фактически представляет собой суперпозицию многих копий исходного «чистого» сигнала, каждая из которых характеризуется своими значениями амплитуды, фазы, временем запаздывания и сдвигом частоты. Перекрываясь во времени, эти копии интерферируют между собой, образуя достаточно сложный результирующий сигнал, что часто затрудняет извлечение необходимой информации. Процедура, направленная на разделение интерферирующих сигнальных копий или нейтрализацию их взаимного вредного взаимодействия получила наименование *разрешения сигналов*. В зависимости оттого, что является конечной целью, могут существовать различные пути решения данной задачи в рамках классического подхода к приему сигналов. Все из них, тем не менее, снова базируются на категории расстояния между копиями сигнала или сходства.

Для лучшего восприятия идеи решения данной проблемы обратимся к наиболее показательному случаю *временного разрешения сигналов*. Начнем с полосного сигнала $s(t) = \text{Re}[\dot{S}(t)\exp(j2\pi f_0 t)]$ и предположим, что на вход приемника поступают две его копии, имеющие одинаковые амплитуды, но различное время запаздывания $\tau_1 = 0$ и $\tau_2 = \tau$ на входе приемника. Тогда колебание суперпозиции

$$s_r(t) = s(t) + s(t - \tau) = \text{Re}\{[\dot{S}(t) + \dot{S}(t - \tau)]\exp(j2\pi f_0 t)\}$$

строго определяется взаимным временным сдвигом τ , а не длительностью сигнала T и несущей частотой, как и «тонкими» деталями модуляции сигнала. Рис. 2.26 демонстрирует несколько ситуаций на примере простого импульса с плавной огибающей: диаграмма *a* отвечает «собственно чистому сигналу», тогда как три остальных соответствуют различным значениям τ , выраженным в терминах несущей частоты f_0 .

В случае, когда различие во времени τ превышает длительность сигнала T (диаграмма *b*), разрешение сигналов не вызывает проблем, поскольку копии сигнала полностью разнесены во времени и наблюдатель способен определить, что на входе присутствуют две копии. Очевидно, что в данной ситуации отсутствуют интерференционные эффекты и, если, помимо временной позиции, некоторые другие параметры сигнала представляют интерес, то информация о них может быть извлечена из каждой копии без всяких затруднений.

Другая ситуация имеет место, когда сигналы полностью перекрываются. Тогда интерференция может привести либо к усилению результирующего сигнала (если сигнальные копии близки по фазе, диаграмма *c*), либо к его ослаблению (если отличие по фазе между копиями близко к π , диаграмма *d*). Возможные трудности для наблюдателя очевидны, учитывая, что принятое колебание всегда искажено канальным шумом. Он либо с трудом сможет распознать присутствие в принятом сигнале более чем одной копии (и быть в полной уверенности, что интерференция не исказила извлеченную информацию

(c)), либо окажется в ситуации чрезмерно низкого отношения сигнал-шум (d).

Сценарии подобного рода широко распространены в системах, связанных с передачей и сбором информации. Физическим основанием рассмотренного явления служат ограниченная полоса канала и многолучевое распространение сигнала. Последнее детально будет обсуждаться в 3.5. Теперь же лишь кратко рассмотрим некоторые характерные при-

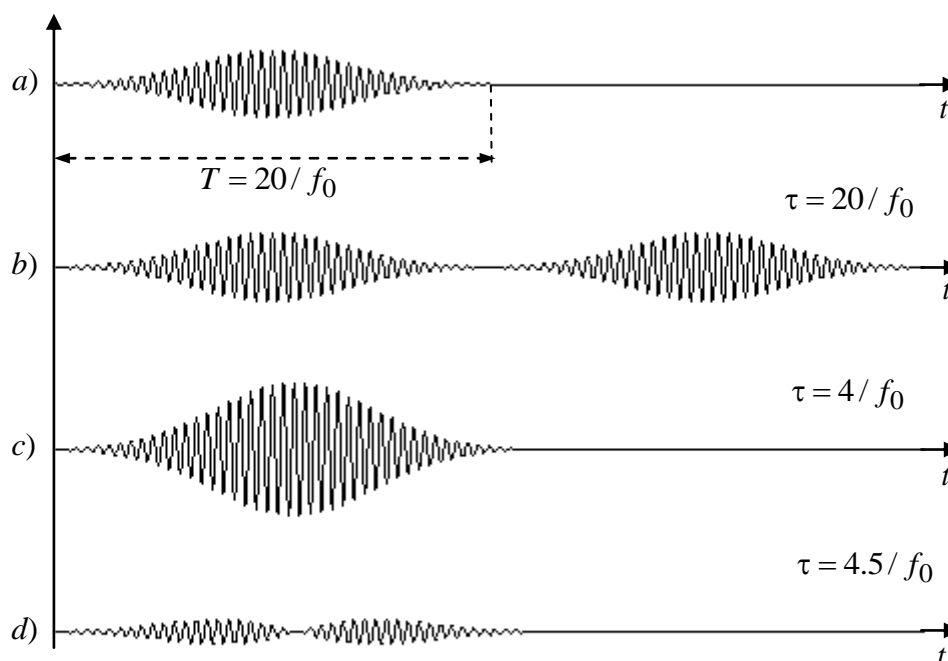


Рис. 2.26. Иллюстрация к задаче временного разрешения.

меры. В цифровой связи наиболее часто встречающимися и опасными являются следующие два вида канальных искажений: *межсимвольная интерференция* – *intersymbol interference* (МСИ) и *фединг* (*fading*). Они имеют общую природу, поскольку являются результатом линейного суммирования множества взвешенных и задержанных во времени копий сигнала. В том случае, когда диапазон задержек сравним со временем корреляции сигнала или, что эквивалентно, передаточная функция канала является полностью неравномерной в пределах спектра сигнала, принятая суперпозиция оказывается существенно искаженной относительно исходного сигнала, что типично для МСИ или *частотно-селективного искажения* (*frequency-selective fading*). Неселективный или *медленный* (*flat*) фединг имеет место тогда, когда задержки копий сигнала отличаются незначительно в сравнении со временем корреляции сигнала (передаточная функция канала близка к равномерной), однако достаточно, чтобы вызвать паразитный взаимный фазовый сдвиг, приводящий к ощутимым потерям в мощности результирующего сигнала. Очевидно, что задача разрешения, т.е. эффективного разделения сдвинутых во времени копий сигнала, непосредственно связана с задачей подавления МСИ.

Аналогичным образом, в радиолокации принимаемое колебание может быть образовано множеством сдвинутых во времени копий сигнала, образованных в результате отражения излученного импульса от множества целей. Одним из основных предназначений радиолокатора является определение числа целей, расположенных в определенном направлении пространства, и измерение расстояний до них. Сразу же становится очевидным, что сценарий, описываемый рис.2.26, b, наиболее прост с рассматриваемой точки зрения, тогда как ситуации, когда взаимное расстояние между целями достаточно мало,

приводят к перекрытию отраженных сигналов (рис.2.26, *c, d*), что и подтверждает необходимость задачи разрешения.

Множественность путей распространения сигналов типична и для навигационных задач. Многие системы наземного базирования основаны на использовании ионосферного отражения колебаний длинно- и средневолнового диапазона. Вследствие этого земная поверхность и ионосферный слой образуют волной канал, по которому распространение сигналов может осуществляться различными путями. Аналогичная проблема характерна и для систем космического базирования, поскольку помимо прямого сигнала со спутника на вход приемника могут поступать сигналы, образованные отражением от окружающих объектов (например, корабельных мачт или палубных сооружений). Все указанные ситуации хорошо укладываются в рамки задачи разрешения.

Анализируя рис. 2.26, не удивительно прийти к заключению, что достижение хорошего временного разрешения возможно на пути использования коротких сигналов. Очевидно, что укорочение сигнала может служить решением задачи, однако остаются в силе ограничения, достаточно полно обсужденные в 2.12. Достижение высокой разрешающей способности невозможно добиться простым уменьшением длительности сигнала, поскольку для сохранения необходимого отношения сигнал-шум требуется обеспечить пропорциональное увеличение пиковой мощности. Следовательно, ограничения на пиковую мощность делают указанную простейшую стратегию не столь привлекательной.

К счастью, существует альтернативный и более элегантный путь решения задачи разрешения сигналов во времени, уже рассмотренный ранее. Точно так же, как и в случае измерения временного положения сигнала, разрешающая способность определяется не длительностью самого сигнала, а протяженностью его АКФ, поскольку очищение сигнала от шума согласованным фильтром может рассматриваться как составная часть любой процедуры приема. Следовательно, требование малости времени корреляции τ_c или остроты АКФ диктуется необходимостью разрешения во времени, повторяя тем самым обоснованность выдвижения подобного требования, впервые сформулированного в задаче оценивания временного запаздывания. Выполнив это требование, можно допустить перекрытие сдвинутых во времени копий сигнала и их интерференцию, однако, если временной сдвиг между копиями τ превышает время корреляции τ_c , отклики на выходе согласованного фильтра не будут перекрываться, т.е. могут быть разрешены. В итоге снова приходим к идее, высказанной в 2.12. Необходимая энергия сигнала (отношение сигнал-шум) обеспечивается большой длительностью сигнала T , позволяя удовлетворить ограничениям на пиковую мощность, тогда как достаточно малое время корреляции $\tau_c \ll T$ достигается соответствующим выбором внутренней угловой модуляции. Тогда эффект временного сжатия в согласованном фильтре способен обеспечить разделение перекрывающихся сигнальных копий, т.е. их разрешение. Потенциально это достижимо только при использовании сигналов с распределенным спектром, для которых $WT \approx T/\tau_c \gg 1$.

Следовательно, можно дословно повторить вывод, сделанный в 2.12, в приложении к разрешению во времени. Если отсутствуют ограничения на пиковую мощность, задача разрешения сигналов *не требует обязательного обращения* к технологии распределенного спектра. Однако, распределение спектра является насущной потребностью при наложении жестких ограничений на пиковую мощность.

Рис. 2.27 демонстрирует две показательные ситуации, смоделированные в среде программирования Matlab. Колонка *a*) содержит простой одиночный импульс $s(t)$ с гладкой огибающей, его сдвинутую на τ копию $s(t - \tau)$, значительно перекрывающуюся с исходным сигналом $s(t)$, суперпозицию $s(t)$ и $s(t - \tau)$ на входе согласованного фильтра и отклик фильтра. Колонка *b*) содержит аналогичные диаграммы для импульса с распределенным спектром (с линейной частотной модуляцией) той же длительности и энергии. При одинаковой в обоих случаях временной задержки сигнальной копии второй вариант демонстрирует отличное качество разрешения – две сигнальные копии полностью разде-

ляются тогда, как первый вариант свидетельствует о невозможности разрешения сигналов.

Выше представленная концепция разрешения во времени легко может быть распространена и на другие параметры. Если перекрываются несколько сигнальных копий, отличающихся только частотным сдвигом, то возникает задача разрешения по частоте, решение которой определяется коэффициентом корреляции (2.73), аналогично случаю оценивания частоты. Очевидно, что в данном случае отсутствуют причины к применению технологии распределенного спектра (см. параграф 2.13). Если же копии сигналов отличаются как временным, так и частотным сдвигом, то процедура частотно-временного разрешения и ее качество определяются функцией неопределенности (2.75). Как и в случае частотно-временного измерения (см. 2.14), данная ситуация требует обращения к технологии распределенного спектра: *ни один из других способов* не может гарантировать получение остроконечной функции неопределенности в любом направлении плоскости τ, F .

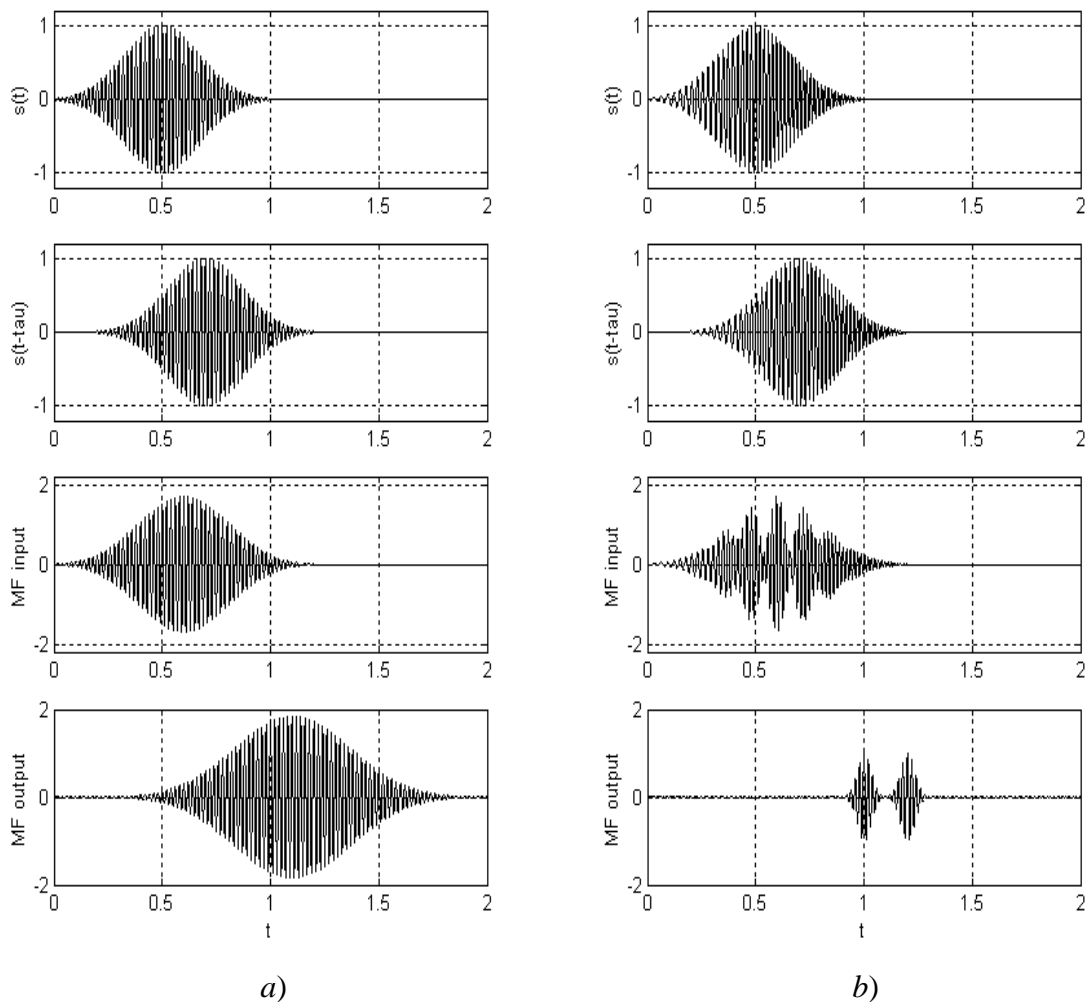


Рис 2.27. Задача разрешения во времени: простой (а) и широкополосный (б) сигналы.

И в заключение параграфа упомянем также о многочисленных ситуациях пространственного разрешения, когда сигнальные копии, поступающие на приемную антенну, приходят с различных направлений и задача наблюдателя заключается в раздельной обработке каждой копии. Именно сама антенна (или антенны, включая и передающую)

играет роль пространственного сигнала, а синтез «сигнала» состоит в наиболее эффективной комбинированной обработке принятых колебаний элементами антенны. Многие положения процедуры разрешения во времени приложимы также и к проблеме пространственного разрешения.

2.16. Заключение.

В данной главе кратко были рассмотрены основные процедуры приема сигналов: обнаружение, различение, оценивание параметров и разрешение. В соответствии с классической формулировкой приема во всех рассмотренных задачах была принята модель гауссовского аддитивного шума, и в рамках этого идеализированного подхода основной целью являлось выяснение необходимости привлечения технологии распределенного спектра. Выводы, приведенные в таблице 2.1, содержат перечень параметров сигнала, которые определяют качество выполнения каждой из специфических процедур, и указания на привлечение распределенности спектра в случае, когда желательно улучшение характеристик процедуры без увеличения энергетических затрат.

Как показывает содержимое таблицы, будет неправильно сказать, что классические задачи приема в значительной степени обращены к философии распределенного спектра. Только одновременное частотно-временное оценивание и разрешение дают явный импульс к ее применению. Это может показаться очень странным и оспаривающим утверждения о широкой популярности распределенности спектра в настоящее время. Как будет показано в следующих главах, эти основания вполне обоснованы и явно проявляют себя всякий раз, когда исследования основываются на более реалистичной модели канала, чем несколько «академичной» гауссовской, или вызывают необходимость привлечения некоторых иных, дополнительных критериев качества.

Таблица 2.1. Роль сигналов с распределенным спектром в классических задачах приема.

Задача.	Параметры сигнала, определяющие качественные характеристики.	Необходимость привлечения сигналов с распределенным спектром.
Обнаружение, амплитудные и фазовые измерения.	Отношение сигнал-шум (только энергия сигнала).	Нет необходимости.
Бинарная передача данных ($M=2$)	Отношение сигнал-шум, коэффициент корреляции.	Нет необходимости.
M -ичная передача данных, $M>2$	Отношение сигнал-шум, коэффициент корреляции всех сигналов.	Нет необходимости, но иногда привлекательно в реализационном плане.
Измерение запаздывания и разрешение во времени.	Отношение сигнал-шум, полоса сигнала.	Нет необходимости при отсутствии ограничений на мощность, в противном случае необходимо.
Измерение частотного сдвига, разрешение по частоте.	Отношение сигнал-шум, длительность сигнала.	Нет необходимости.
Частотно-временные измерения, разрешение по времени и частоте.	Отношение сигнал-шум, полоса и длительность сигнала.	Необходимо.

Задачи.

2.1. Заданы три сигнала $s_1(t)$, $s_2(t)$ и $s_3(t)$, изображенные на рис. 2.28.

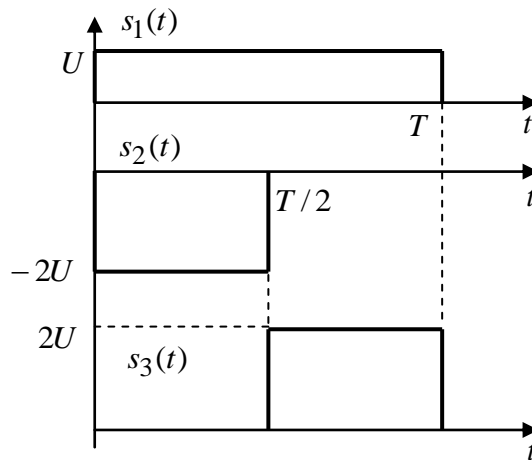


Рис. 2.28. Множество из трех сигналов.

Во сколько раз максимальное расстояние в данном множестве сигналов больше его минимального значения?

2.2. Наблюдение $y(t)$ на выходе АБГШ–канала имеет вид, представленный на рис. 2.29:

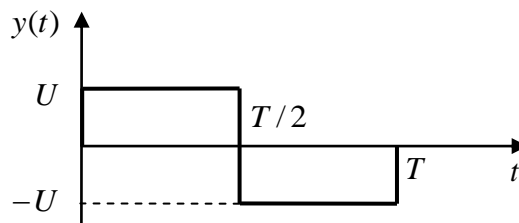


Рис. 2.29. Сигнал на выходе канала.

Сигналы на входе канала заданы в задаче 2.1. Каким будет решение оптимального приемника?

2.3. Источник генерирует данные со скоростью $R = 10$ кбит/сек. Каждый бит данных источника передается по АБГШ–каналу с помощью бинарной ФМ. Допустимой является полоса $W = 10$ МГц. Рационально ли использовать сигналы с полосой $W = 10$ МГц?

2.4. Какие сигналы с бинарной ФМ являются предпочтительными для передачи по АБГШ–каналу:

- прямоугольные импульсы с пиковой мощностью 1000 Вт и полосой 100 КГц;
- прямоугольные импульсы такой же длительности с пиковой мощностью 900 Вт и полосой 10 МГц?

2.5. Вычислить энергетические потери для пар сигналов, используемых для бинарной передачи данных по АБГШ–каналу (см. рис.2.30), по отношению к оптимальной паре.

2.6. При дифференциальной бинарной фазовой манипуляции (DBPSK) значение бита передается с чередованием или без полярности двух последовательных импульсов: импульсы одинаковой полярности соответствуют нулю, а различной – единице. Сравнить в первом приближении DBPSK и BPSK по уровню потребляемой энергии (основываясь только на минимуме расстояния) и ширине занимаемой полосы при условии равенства скоростей.

2.7. При квадратурной ФМ (ФМ-4) два бита (4 сообщения) передаются четырьмя сигналами с начальными фазами: $0, \pi, \pm \pi/2$. Является ли данный вариант оптимальным способом передачи двух бит? Если нет, указать наилучший способ и оценить соответствующий ему асимптотический выигрыш от кодирования по сравнению с ФМ-4.

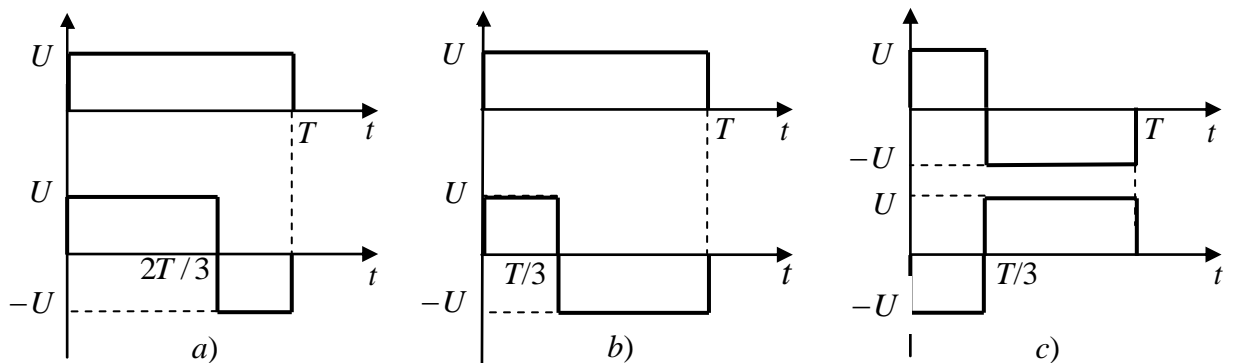


Рис. 2.30. Три сигнальные пары.

2.8. Можно ли построить 10 эквидистантных сигналов, для которых коэффициент корреляции между двумя любыми был бы равным $-1/7$? Каково максимально возможное число сигналов с указанным коэффициентом корреляции?

2.9. Вычислить и построить зависимость от M энергетических потерь (в дБ) множества M ортогональных сигналов по сравнению с множеством M оптимальных сигналов для АБГШ канала. Определить значение потерь в асимптотике при увеличении M .

2.10. Асимптотический энергетический выигрыш при ортогональном кодировании в сравнении с не кодированной передачей для случая $M = 2$ стремиться к 0,5 или -3 дБ, т.е. является отрицательным, демонстрируя потери, а не выигрыш. Как это можно объяснить практически?

2.11. Асимптотический энергетический выигрыш при ортогональном кодировании в сравнении с не кодированной передачей для случая $M = 4$ стремиться к 1 или 0 дБ, т.е. отсутствует совсем. Дать физическое объяснение результату.

2.12. Сообщения ($M = 8$) передаются с использованием ФМ-8, т.е. идентичными радиоимпульсами с 8-ью различными эквидистантными начальными фазами. Является ли этот вариант передачи 8 сообщений по гауссовскому каналу наилучшим при отсутствии ограничений на ширину полосы? Если нет, то каковы энергетические потери варианта с ФМ-8 по сравнению с оптимальным множеством из M сигналов?

2.13. Сравнить (на основании минимума расстояния) эффективность M -ичной ФМ с ортогональным кодированием по энергетическим затратам (при заданной вероятности ошибки) и занимаемой полосе.

2.14. В обратном канале сотовой системы радиосвязи стандарта IS-95 осуществляется ортогональное кодирование блоков из 6-ти бит. Скорость передачи составляет 28.8 кбит/сек. Оценить величину полосы, занимаемую кодированными сигналами (не учитывая дальнейшего расширения спектра сигнала с помощью длиной M -последовательностью).

2.15. Некоторая система цифровой связи занимает полосу $W = 1.2288$ МГц. Какое максимальное число M ортогональных сигналов может быть использовано для передачи данных со скоростью 38.4 кбит/сек?

2.16. В некоторой системе осуществляется передача данных по гауссовскому каналу со скоростью 10 кбит/сек. Проектировщик системы планирует обеспечить энергетический выигрыш в 6 дБ по сравнению с передачей без кодирования. Достижима ли эта цель на основе ортогонального кодирования, если допустимой считается полоса в 320 КГц?

2.17. Некоторой системе отведена полоса, равная 10.24 МГц, при необходимой скорости передачи в 100 кбит/сек. Какой величины может достичь в этой системе потенци-

альный асимптотический выигрыш от кодирования?

2.18. По гауссовскому каналу необходимо передавать данные со скоростью 100 кбит/сек на несущей в 2 ГГц. Реалистично ли рассчитывать на энергетический выигрыш $G = 10$ дБ при использовании ортогональных сигналов?

2.19. Построить матрицу Адамара размерности 16.

2.20. Какое из следующих преобразований сохранит/нарушит основное свойство матрицы Адамара (ортогональность строк):

- перестановка строк;
- перестановка столбцов;
- одновременное изменение знака у всех элементов;
- одновременное изменение знака у всех элементов нескольких строк;
- одновременное изменение знака у всех элементов нескольких столбцов;
- изменение знака только у верхнего крайнего слева элемента.

2.21. Матрица Адамара \mathbf{H}_M размера $M = 2^m$ построена из исходной матрицы $\mathbf{H}_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ с помощью правила Сильвестра. Из матрицы \mathbf{H}_M удален первый столбец, а строки вновь полученной матрицы используются при формировании сигналов для M -ичной передачи. Какого вида ансамбль сигналов получается в результате такой операции? Какую полосу они занимают в сравнении с ортогональным ансамблем?

2.22. Сигнал с ЛЧМ (линейной частотной модуляцией) задается соотношением

$$s(t) = \begin{cases} A \cos \left(2\pi f_0 t + \frac{\pi W_d (t - \tau)^2}{T} + \varphi \right), & |t - \tau| \leq T/2, \\ 0, & |t - \tau| > T/2 \end{cases}$$

где A – амплитуда, f_0 – несущая частота, W_d – девиация, T – длительность, τ – время запаздывания и φ – начальная фаза сигнала. Разделить эти шесть параметров на энергетические и нет. (Для любого полосного сигнала $f_0 T \gg 1, W \ll f_0$).

2.23. Необходимо измерить неэнергетический параметр сигнала λ . Зависимость коэффициента корреляции $\rho(\lambda)$ сигнальных копий, отличающихся по величине λ , для трех различных случаев представлена на рис.2.31. В каком случае точность измерения λ будет наивысшей и почему?

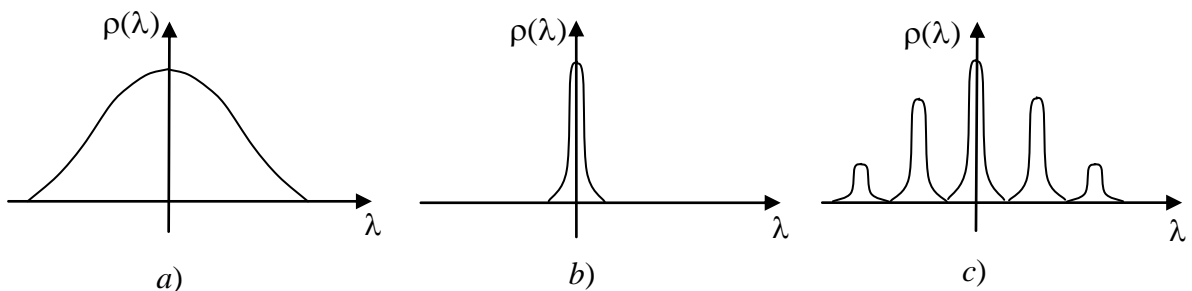


Рис. 2.31. Зависимость коэффициента корреляции от измеряемого параметра.

2.24. Одновременно должны быть измерены два неэнергетических скалярных параметра λ_1, λ_2 . Коэффициент корреляции двух копий сигнала с различными парами значений λ_1, λ_2 определяется $\rho(\lambda_1, \lambda_2)$ и геометрически представляет собой поверхность в трехмерном пространстве. Сечение этой поверхности на некотором уровне (например, 0.5) горизонтальной плоскостью представлено на рис. 2.32 для трех типичных случаев. Какой

из них обеспечивает наибольшую точность одновременного оценивания λ_1, λ_2 при условии отсутствия априорных сведений об их значениях и почему?

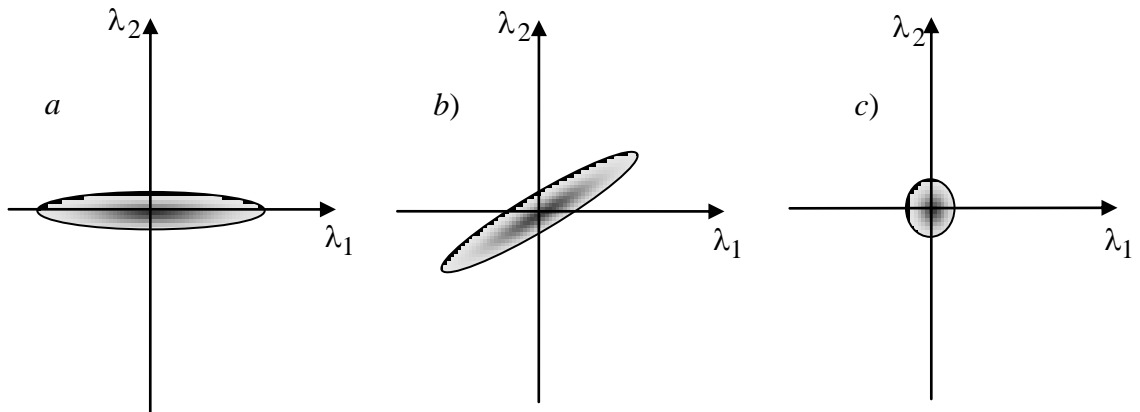


Рис.2.32. Горизонтальное сечение поверхности $\rho(\lambda_1, \lambda_2)$.

2.25. Следует измерить амплитуду A сигнала $s(t; A) = As(t)$. Образцы сигнала $s(t)$ представлены на рис. 2.33. В каком из случаев точность измерения A будет наилучшей и почему?

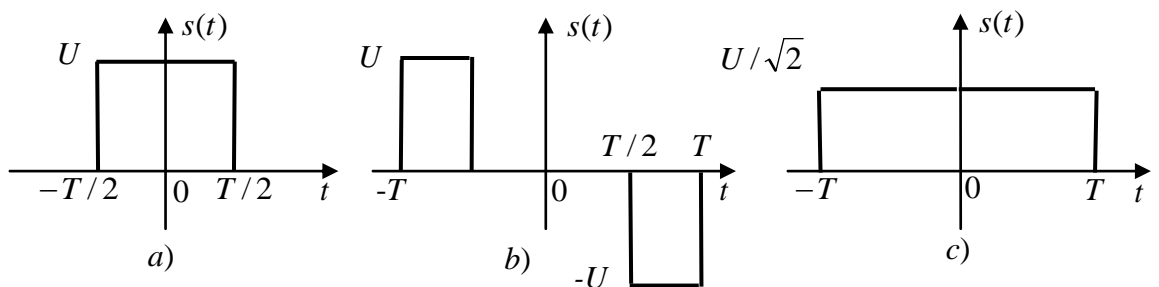


Рис.2.33. Три варианта образцов сигнала.

2.26. Измеряется амплитуда A сигнала $s(t; A) = As(t)$. Точность полученной оценки оказалась неудовлетворительной. Во сколько раз следует увеличить длительность исходного сигнала $s(t)$, чтобы вдвое уменьшить среднеквадратическое отклонение оценки A при сохранении неизменными всех остальных параметров $s(t)$?

2.27. Измеряется амплитуда A сигнала $s(t; A) = As(t)$. Амплитуда исходного сигнала $s(t)$ увеличивается вдвое, тогда как его длительность в два раза уменьшается. Что произойдет с величиной среднеквадратического отклонения оценки A ?

2.28. Необходимо измерить начальную фазу полосного сигнала. Что произойдет со среднеквадратическим отклонением оценки фазы, если

- удвоить несущую частоту сигнала;
- удвоить длительность сигнала;
- уменьшить в два раза амплитуду;
- удвоить амплитуду, а длительность уменьшить в четыре раза?

2.29. Необходимо измерить начальную фазу полосного сигнала. Три варианта огибающей сигнала представлены на рисунке к задаче 2.25. В каком случае точность измерения фазы будет наибольшей и почему?

2.30. Измеряется начальная фаза φ ЛЧМ сигнала, определенного в задаче 2.22. Изменение каких параметров A, f_0, W, τ, T , и в каком направлении повлияет на точность

оценки фазы? Что произойдет со среднеквадратическим отклонением оценки фазы, если все значения A , f_0 , W увеличить в $\sqrt{2}$ раз, а T , τ уменьшить в два раза?

2.31. Изобразить автокорреляционные функции сигналов, представленных на рис. 2.34.

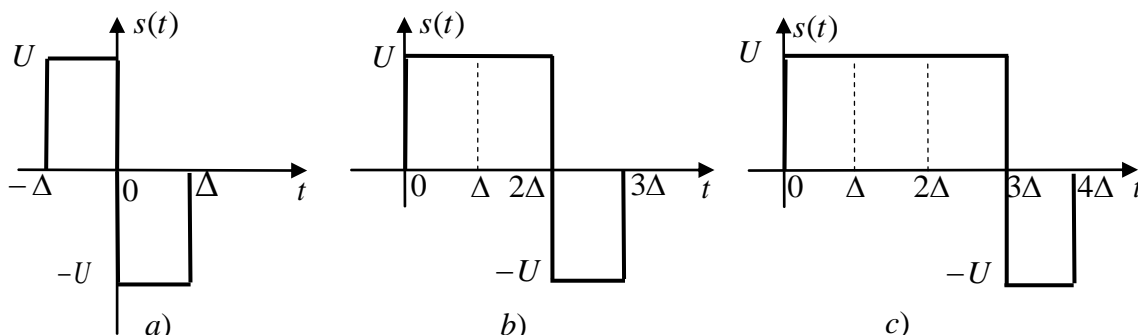


Рис. 2.34. Три варианта формы сигнала.

2.32. Полосный бинарный ФМ сигнал состоит из трех последовательных прямоугольных импульсов длительностью Δ каждый. Начальные фазы первых двух равны нулю, а третьего $-\pi$. Изобразить автокорреляционную функцию сигнала.

2.33. Задан фильтр, согласованный с прямоугольным видеоимпульсом длительности Δ . Какого вида устройства следует добавить к нему, чтобы получить согласованный фильтр с сигналами на рис. 2.34? Изобразить отклик фильтра, согласованного с сигналом (с), при подаче на вход именно этого сигнала.

2.34. Задан фильтр, согласованный с прямоугольным радиоимпульсом длительности Δ . Какие дополнительные к нему устройства образуют согласованный с сигналом из задачи 2.32 фильтр? Изобразить отклик согласованного фильтра на сигнал из задачи 2.32.

2.35. Заданы автокорреляционные функции трех альтернативных сигналов, изображенные на рис. 2.35. Какой из сигналов является наилучшим для измерения временной задержки и почему?

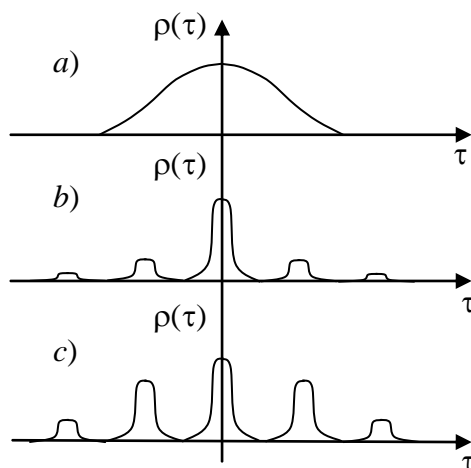


Рис. 2.35. Примеры автокорреляционных функций.

2.36. Параметры простого импульсного сигнала обеспечивают стандартное отклонение измерения временной задержки, равное 0.5 мксек и отношение сигнал-шум (SNR) на выходе согласованного фильтра $q = \sqrt{2E/N_0} = 10$. Оценить грубо длительность сигнала.

2.37. В некоторой РЛС используется простой импульсный сигнал. Проектировщик системы планирует уменьшить пиковую мощность в сто раз без ухудшения SNR и в то же время уменьшить в десять раз стандартное отклонение измеряемого времени запаздывания. Каким должно быть частотно-временное произведение сигнала в усовершенствованной системе?

2.38. В некоторой РЛС измеряется расстояние посредством излучения импульса с ЛЧМ (см. задачу 2.22) с частотно-временным произведением $WT = 10^3$. В результате поломки модулятора РЛС стала излучать немодулированные импульсы той же пиковой мощности и длительности. Что произойдет со стандартным отклонением измеряемой дальности?

2.39. В некоторой системе необходимо уменьшить в десять раз стандартное отклонение частотного измерения. Мощность сигнала может быть увеличена только в 25/16 раза (1.94 дБ). Насколько должна измениться длительность сигнала?

2.40. В некоторой системе длительность сигнала увеличена в четыре раза без изменения мощности сигнала. Что произойдет со стандартным отклонением оценки частоты?

2.41. Горизонтальные сечения функции неопределенности четырех сигналов представлены на рис. 2.36, на котором все размеры, отмеченные символами (τ_{c1} и др.), одинаковы. Какой из сигналов является наилучшим для:

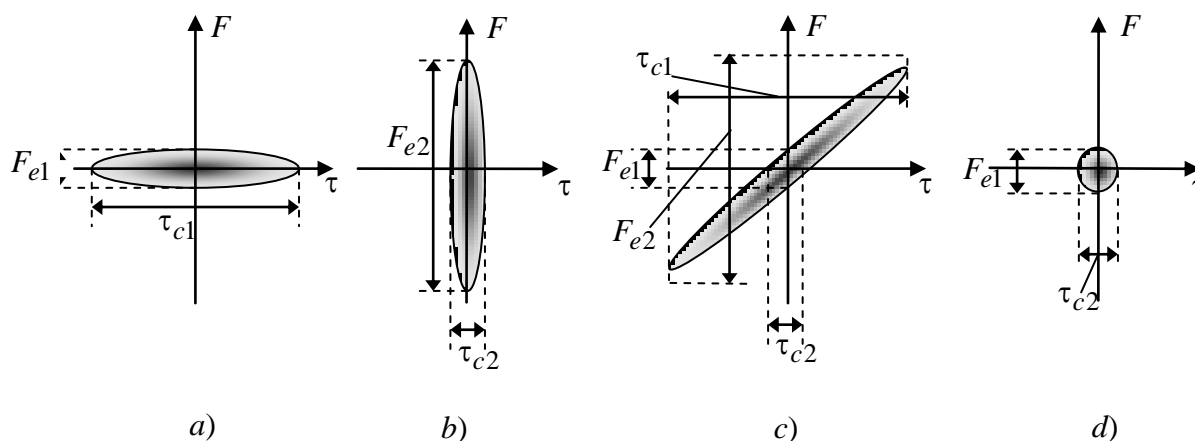


Рис. 2.36. Горизонтальные сечения функции неопределенности.

- a) измерения только времени запаздывания;
 - b) измерения только частотного сдвига;
 - c) одновременного измерения временного запаздывания и частотного сдвига
- и почему?

2.42. В ходе модернизации некоторой системы, работавшей первоначально с простым сигналом, излучаемая мощность была уменьшена на 6 дБ. Одновременно была увеличена в четыре раза длительность сигнала, а простой сигнал заменен широкополосным с частотно-временным произведением $WT = 100$. Что произошло со стандартным отклонением оценок временного запаздывания и частоты в сравнении с первоначальным вариантом?

2.43. Какой из сигналов в задаче 2.41 является наилучшим для разрешения:

- a) по времени;
 - b) по частоте;
 - c) по времени и частоте одновременно
- и почему?

2.44. Горизонтальные сечения функции неопределенности четырех сигналов представлены на рис. 2.37, на котором все размеры, отмеченные одинаковыми символами, равны. Какой из сигналов является наилучшим для разрешения:

- a) по времени;
 - b) по частоте;
 - c) по времени и частоте одновременно
- и почему?

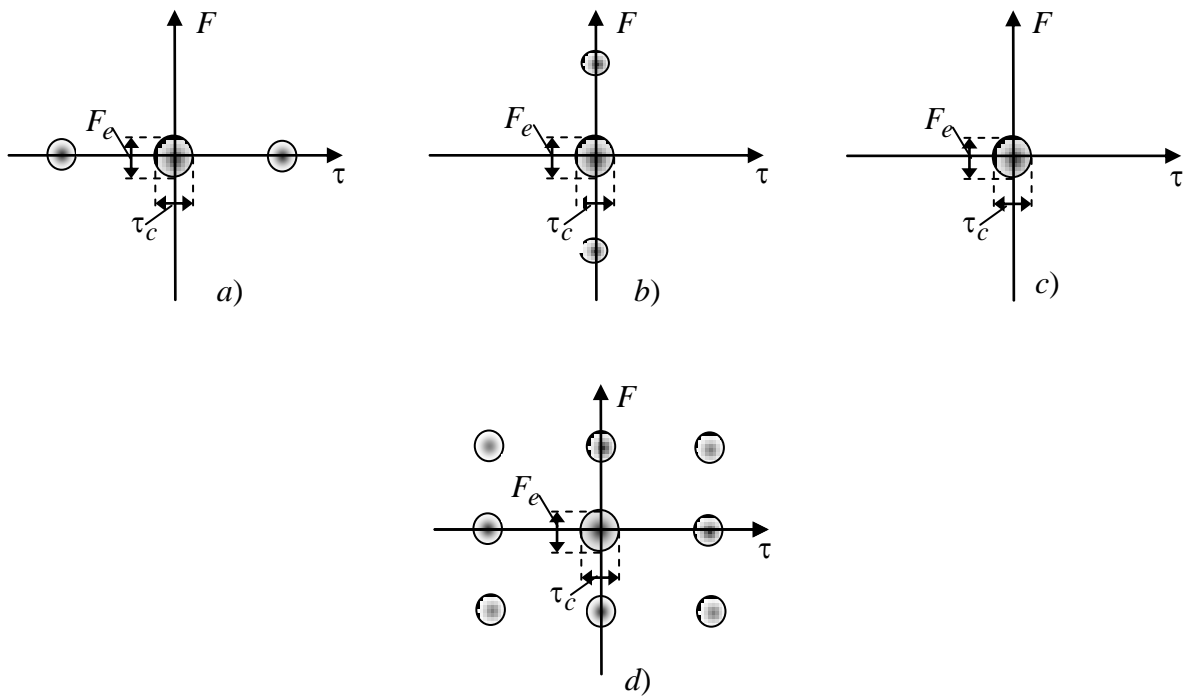


Рис. 2.37. Горизонтальные сечения функции неопределенности.

3. Достоинства технологии распределенного спектра.

3.1. Иммунитет к воздействию помех.

Окружающая обстановка, в которой конкретная система передает и извлекает информацию, не является дружественной для выполнения указанных целей. На приемной стороне полезному сигналу могут сопутствовать наряду с тепловым шумом и искажения различной природы. В частности, источниками искажений могут служить другие соседствующие системы, если диапазон их рабочих частот достаточно близок к рассматриваемой в качестве «полезной». Следуя повсеместно принятой терминологии, будем называть этот вид искажений помехами.

В повседневной практике встречается большое разнообразие типов помех в связи с чем, как правило, принимаются специальные меры, противодействующие их разрушительному действию. В данной главе будет показано, что расширение спектра является мощным инструментом нейтрализации помех. Исчерпывающее исследование поведения систем, подверженных комбинированному воздействию помех и теплового шума, требовало бы вычисления интегральных характеристик, таких как вероятность ошибки или точность измерения. Конкретные результаты этой достаточно кропотливой работы можно найти в книгах, например [5, 6] и, в основном, многочисленных специализированных статьях (см., например, библиографию в [3]), однако наша задача много скромнее и состоит в демонстрации общей идеи борьбы с помехами путем использования сигналов с пространственным спектром. По этой причине ограничимся только простейшим оцениванием отношения мощностей сигнала к общему мешающему воздействию. В противном случае, будем полагать, что действует простейший, но и наиболее характерный вид помехи, описываемый гауссовским случайным процессом, спектр которого перекрывается со спектром сигнала. Иногда такой упрощенный подход является адекватным, как это имеет место при передаче данных с помощью BPSK или ASK, когда вероятность ошибки зависит только от выше упомянутого отношения, что предполагает возможность аппроксимации мешающего воздействия гауссовским законом. Другие ситуации (как, например, M-ичная передача данных или измерение параметров) не настолько очевидны, и характеристики качества определяются не только величиной отношения мощностей. Однако данное отношение остается индикативным, позволяющим судить о потенциальных достоинствах распределенности спектра. Рассмотрим две основных модели помех, начиная со случая узкополосной помехи.

3.1.1. Узкополосная помеха.

Данный тип помех наиболее характерен для ситуаций, когда некоторая соседствующая система или системы не имеют враждебных намерений по отношению к рассматриваемой, и создают помехи только в результате своего нормального функционирования. На рис. 3.1 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. Узкополосная помеха, кроме того, может быть определена как частично-полосная, гармоническая и т.д. [3, 5, 6], однако в нашей ситуации конкретное значение ее полосы не играет роли.

Первоначально предположим, что рассматриваемая система не предпринимает никаких специальных мер для противодействия помехи за исключением только соответствующего выбора сигнала. Подобного рода сценарий означает, что при проектировании

системы предусмотрена возможность присутствия помехи, и она принята в расчет на этапе выбора сигнала, однако система не является адаптивной, и не принимаются специальные меры путем выбора закона модуляции или алгоритмов обработки, приспособленных к

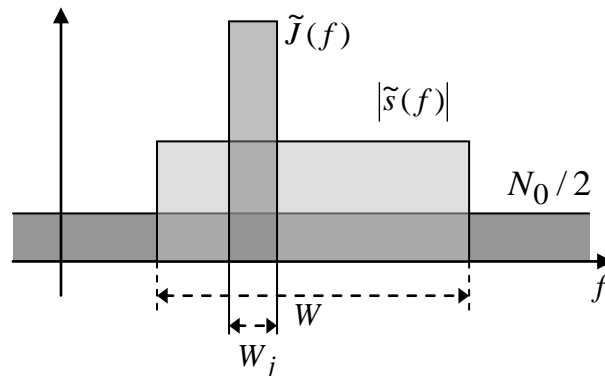


Рис. 3.1. Спектр сигнала, помехи и фонового АБГШ.

существующей помеховой обстановке. Другими словами, всегда используется только фильтр, согласованный с АБГШ, не взирая на наличие или отсутствие помехи на входе приемного устройства.

При определении отношения q_I^2 мощностей сигнала и мешающего воздействия (SIR) на выходе согласованного фильтра отметим тот факт, что при прямоугольном спектре сигнала ($|\tilde{s}(f)|$ равен константе \tilde{s} в пределах полосы сигнала W и нулю вне ее) амплитудная передаточная функция фильтра равномерна в полосе сигнала W и равна нулю вне ее. Без потери общности положим, что ее ненулевое значение равняется единице. Следовательно помеха, трактуемая как случайный процесс, проходит на выход фильтра без изменения своей мощности J , тогда как отфильтрованное значение мощности шума составит N_0W . С другой стороны, фильтр согласован с сигналом и, значит, когерентно суммирует все гармонические составляющие сигнала, обеспечивая максимальное значение амплитуды на выходе $A_{out} = \int_{-\infty}^{\infty} |\tilde{s}(f)| df = 2W\tilde{s}$, 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

$$q_I^2 = \frac{A_{out}^2}{J + N_0W} = \frac{4W^2\tilde{s}^2}{J + N_0W} = \frac{2E}{N_0 + J/W}. \quad (3.1)$$

Из последнего равенства можно увидеть, что, несмотря на конкретное значение полосы помехи W_j , отношение SIR на выходе согласованного фильтра ведет себя так, как если бы мощность помехи была равномерно распределена в полосе W сигнала (не помехи), 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. 3.2 иллюстрирует вид спектра на выходе режекторного фильтра. Наблюдается полное отсутствие помех, однако, вы-

резаются 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)

$$q_J^2 = \frac{2E(1-W_j/W)}{N_0} = q^2(1-W_j/W), \quad (3.2)$$

где $q^2 = 2E/N_0$ – «истинное» отношение мощностей сигнала и шума SNR на выходе согласованного фильтра в отсутствии помехи.¹

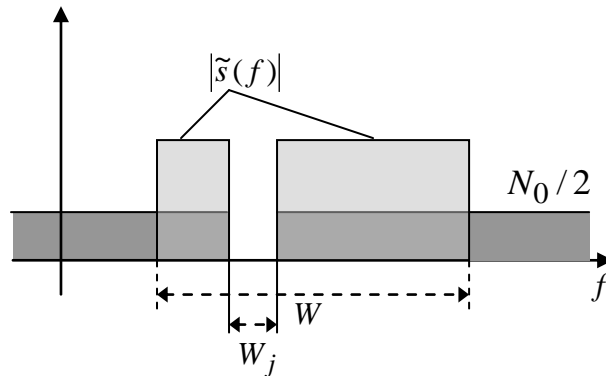


Рис. 3.2. Модель спектра на выходе режекторного фильтра.

Из анализа соотношений (3.1)–(3.2), 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 . Однако, если пиковая мощность сигнала P ограничена и не может быть увеличена, то расширение полосы сигнала не реализуется тривиальным укорочением длительности сигнала, поскольку в этом случае произойдет уменьшение энергии сигнала, а также отношений SNR и SIR. 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*.

3.1.2. Заградительная помеха.

Во многих военных сценариях и разведывательных мероприятиях часто встречается ситуация, когда помеха создается преднамеренно как средство радиоэлектронного противодействия. В подобных ситуациях постановщик помехи может предполагать, что по-

¹ Еще раз подчеркнем, что отношение SNR не является универсальной характеристикой качества. Она отвечает этой роли при рассмотрении передачи данных с помощью BPSK или ASK, но, в общем случае, вырезание полосы затрагивает не только энергию сигнала, но и корреляционные свойства тоже. Так, например, ортогональность сигналов может быть нарушена при частичном вырезании полосы. Для исследования подобного рода эффектов необходим более детальный анализ.

давленная 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. 3.3). 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 (3.1). В этом случае, однако, враждебные намерения

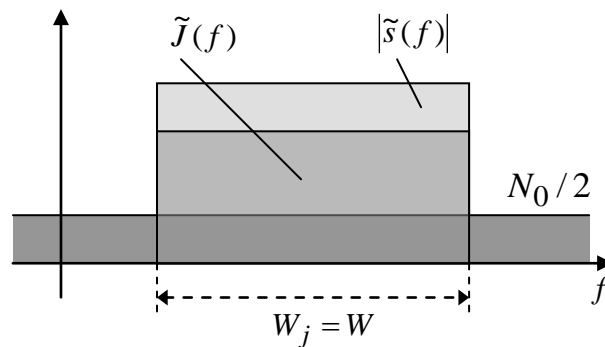


Рис 3.3. Спектр сигнала, заградительной помехи и АБГШ.

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

$$q_J^2 \approx \frac{2EW}{J} = \frac{2P(WT)}{J}. \quad (3.3)$$

Из (3.3) следует, что при наложении ограничений, как на пиковую мощность полезного сигнала, так и на мощностной ресурс постановщика помех, *единственным способом повышения иммунитета системы к заградительной помехе* является привлечение сигналов с большим значением частотно-временного произведения WT , т.е. широкополосных сигналов.

Формула (3.3) 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 (processing gain)*.

Для визуального подтверждения вышеупомянутого заключения приведем два рисунка, полученных моделированием в среде Matlab. Рис. 3.4 иллюстрирует согласованную фильтрацию простого сигнала с прямоугольной огибающей (колонка *a*) и широкополосного сигнала (ЛЧМ, $WT \approx 50$) (колонка *b*) одинаковой длительности и энергии. Одна и та же непрерывная помеха добавляется к обоим сигналам (вторая строка). В то время как простой сигнал полностью замаскирован помехой и явным образом не виден на выходе согласованного фильтра, широкополосный сигнал, отчетливо наблюдается на выходе после временного сжатия в согласованном фильтре.

На рис. 3.5, где колонки соответствуют прежним сигналам, верхняя диаграмма от-

вечает их спектрам. Диаграммы во второй строке демонстрируют спектр мощности двух случайных реализаций различных заградительных помех с одинаковой средней мощно-

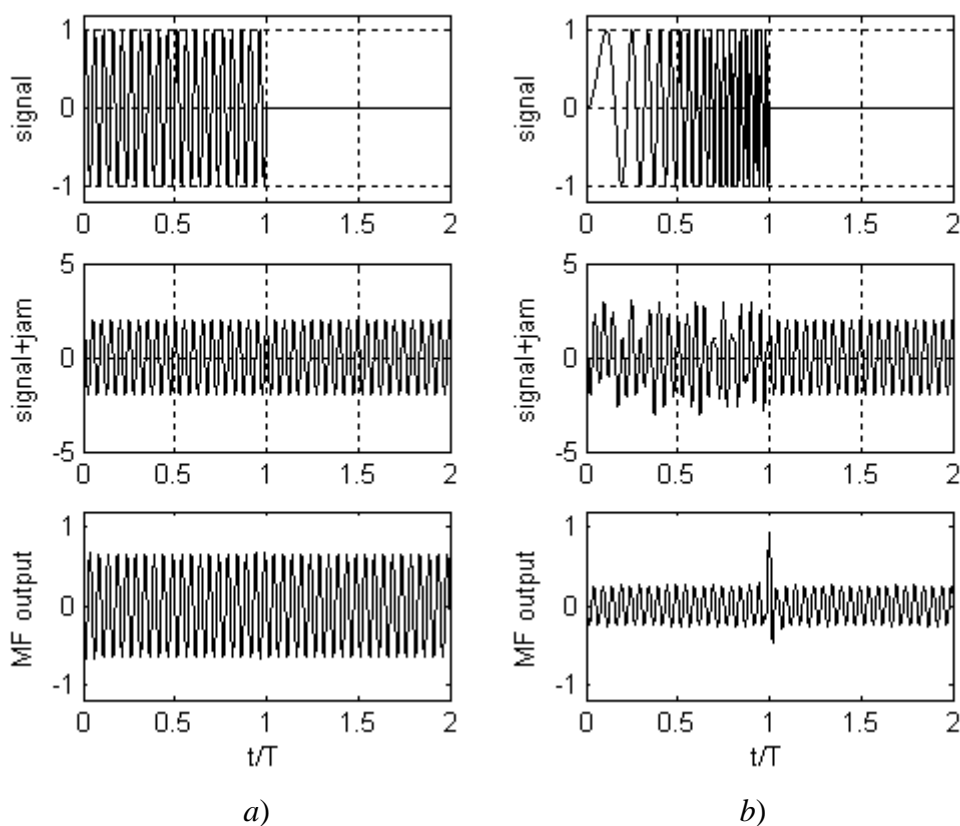


Рис. 3.4. Примеры согласованной фильтрации сигнала и помехи.

стью в полосе сигнала. Вследствие этого средний уровень спектра в колонке *b)* примерно в 50 раз ниже уровня в колонке *a)*. Третий ряд демонстрирует пример наблюдаемого колебания, причем интенсивность помехи примерно одинакова для обоих сигналов, которые достаточно хорошо ею скрыты. Что касается нижних диаграмм, то они снова недвусмысленно подтверждают превосходство широкополосного сигнала в устойчивости к заградительной помехе.

Завершая параграф, еще раз 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. Заинтересованных в данной проблеме можно отослать к многочисленным публикациям, в которых они убедятся, что какими бы сложными не были исследуемые системы и стратегии, основная тенденция остается прежней: распределенный спектр повышает потенциал иммунитета к воздействию помех.

3.2. Низкая вероятность обнаружения.

Как уже отмечалось ранее, перспективность распределенного спектра впервые была признана разработчиками военных и разведывательных систем и, в соответствие с результатами предыдущего параграфа, одной из причин этого являлась высокая стойкость широкополосных сигналов к воздействию помех. Роль другого стимула к подобному ис-

пользованию будет рассмотрена в данном параграфе.

При радиоэлектронном противодействии эффективная помеха может быть организована только после обнаружения присутствия противостоящей системы в эфире 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 (intended)*) старается действовать по возможности скрытно и избежать обнаружения своего сигнала, тогда как вторая (*перехват-*

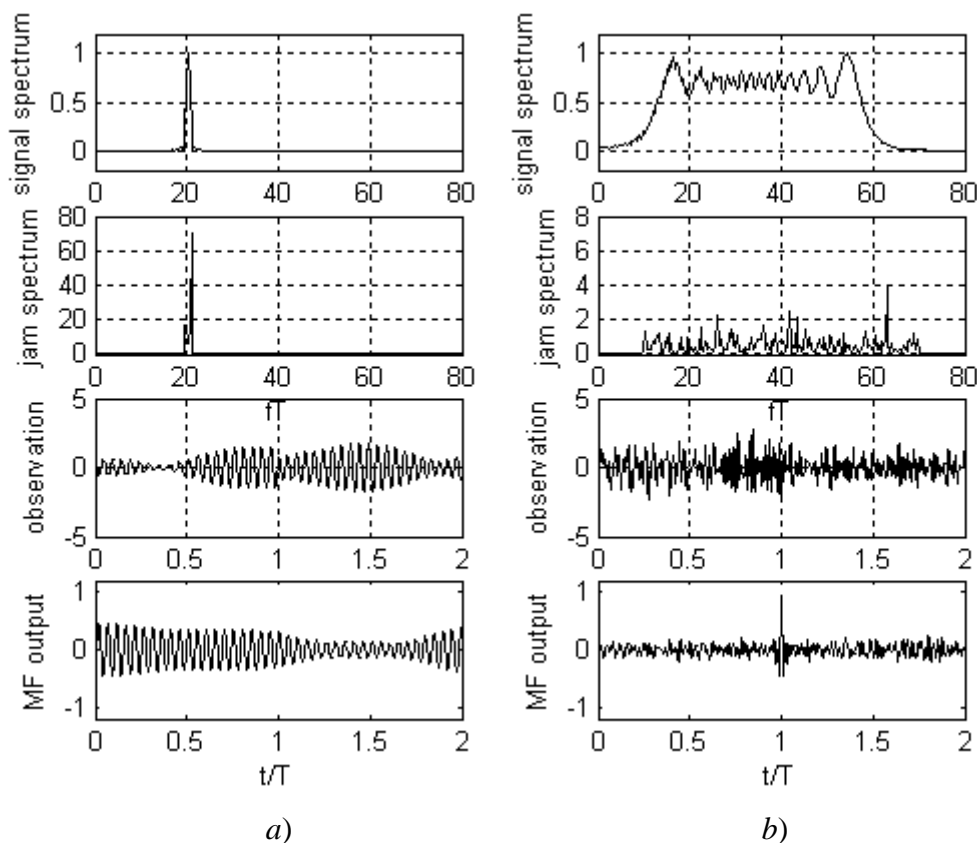


Рис. 3.5. Очистка сигнала от заградительной помехи.

чик или *eavesdropper*) 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.

Существует множество стратегий и методов, которые гипотетически могут находиться в распоряжении системы–перехватчика. Они могут быть достаточно сложными и трудно поддающимися анализу (см. [6, 9] и приведенную в них библиографию). 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. На рис. 3.6 изображена структурная схема подобного энергетического детектора. Полосовой фильтр с полосой W_i , пропускающей весь спектр сигнала или только его часть, фильтрует наблюдение с целью устранения внеполосного шума. Затем квадратичный амплитудный детектор осуществляет оценку мгновенной мощности, которая в дальнейшем интегриру-

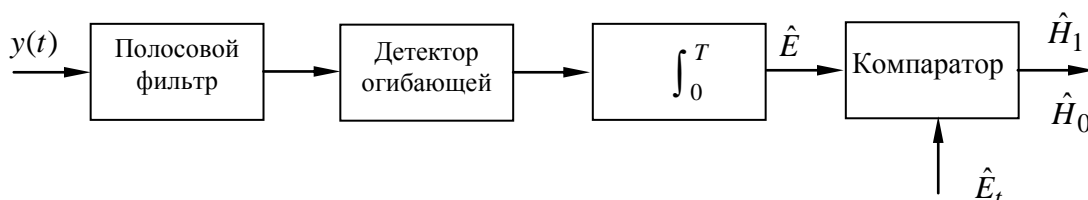


Рис. 3.6. Энергетический детектор.

ется для выработки оценки энергии \hat{E} в пределах интервала наблюдения T_{ob} . Полученная оценка энергии сравнивается затем с порогом E_t , а выполнение неравенства $\hat{E} \geq E_t$ влечет принятие решения о наличии в наблюдении сигнала наряду с естественным шумом, тогда как противоположное свидетельствует об отсутствии сигнала. 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. Это позволяет идеализировать ситуацию с априорными сведениями перехватчика и предполагать, что он осведомлен о возможном месте появления энергии на плоскости время–частота. Поскольку наблюдение вне длительности сигнала не содержит информации о его присутствии, можно положить $T_{ob} = T$, как это было сделано в схеме на рис.3.6.

На рис. 3.7 представлена прямоугольная аппроксимация спектра сигнала вместе с равномерной спектральной плотностью мощности естественного АБГШ (a) и амплитудно-частотная передаточная функция полосового фильтра радиометра (b). 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 $\sigma_n^2 = N_0W_i$, а при его наличии – $\sigma_n^2 + \sigma_s^2 = (N_s + N_0)W_i = (P/W + N_0)W_i$.

Найдем среднее и дисперсию на выходе детектора огибающей. Прежде всего, выходное напряжение u_d детектора равняется мгновенной входной мощности. Следовательно, математическое ожидание \bar{u}_d напряжения u_d в отсутствие сигнала есть просто средняя мощность отфильтрованного шума $\bar{u}_{d0} = \sigma_n^2$, тогда как при наличии сигнала его значение возрастает до $\bar{u}_{d1} = \sigma_n^2 + \sigma_s^2$. Очевидно, что именно приращение \bar{u}_d , обусловленное сигналом и равное

$$\overline{\Delta u_d} = \bar{u}_{d1} - \bar{u}_{d0} = \sigma_s^2 = N_s W_i = P W_i / W, \quad (3.4)$$

позволяет перехватчику надеяться зарегистрировать факт появления сигнала. С другой стороны, этот полезный (с точки зрения перехватчика) эффект маскируется случайными флюктуациями u_d , величина которых оценивается дисперсией $D\{u_d\}$. Последняя характеристика может быть найдена на основании того факта, что мгновенная мощность полосного процесса совпадает с мгновенным значением его огибающей Y , возведенной в квадрат и умноженной на 0.5, так что $u_d = Y^2 / 2$, откуда $D\{u_d\} = D\{Y^2\} / 4$. В свою очередь, дисперсия любой случайной величины может быть вычислена как среднее значение ее квадрата минус квадрат ее среднего значения [1,13,14]:

$$D\{Y^2\} = \overline{Y^4} - \left(\overline{Y^2}\right)^2. \quad (3.5)$$

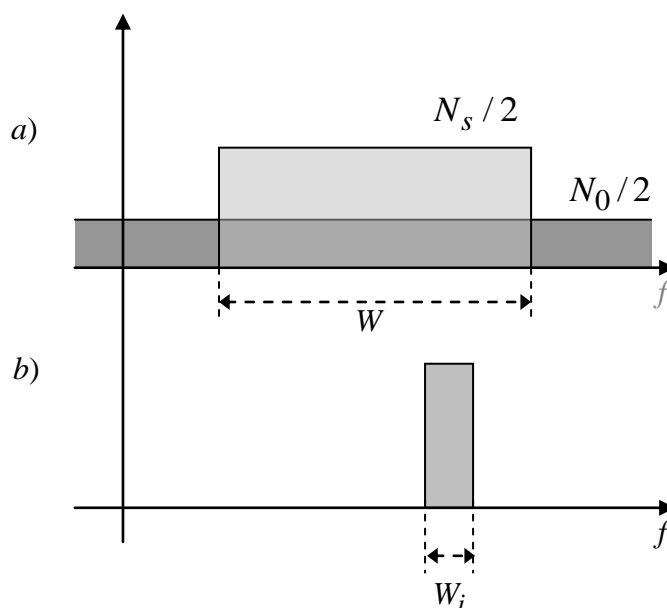


Рис 3.7. Спектр наблюдения на входе радиометра и ширина полосы полосового фильтра.

Теперь, используя тот факт, что огибающая Y гауссовского случайного полосного процесса с дисперсией σ^2 имеет релеевский закон распределения [1]

$$W(Y) = \begin{cases} \frac{Y}{\sigma^2} \exp\left(-\frac{Y^2}{2\sigma^2}\right), & Y \geq 0, \\ 0, & Y < 0 \end{cases}$$

а ее четные моменты находятся посредством интегрирования [13]

$$\overline{Y^{2n}} = \int_0^{\infty} Y^{2n} W(Y) dY = (2\sigma^2)^n \int_0^{\infty} \left(\frac{Y^2}{2\sigma^2}\right)^n \exp\left(-\frac{Y^2}{2\sigma^2}\right) d\left(\frac{Y^2}{2\sigma^2}\right) = n! (2\sigma^2)^n.$$

Подстановка полученного результата в (3.5) дает $D\{Y^2\} = 8\sigma^4 - 4\sigma^4 = 4\sigma^4$. В случае отсутствия сигнала на входе приемник перехватчика в данном выражении следует положить $\sigma^2 = \sigma_n^2$. Строго говоря, при появлении сигнала отфильтрованное наблюдение может отличаться от гауссовского процесса, делая применимость только что полученного результата сомнительной. Данное замечание, однако, не имеет особой важности в рассматриваемом

мом вопросе, поскольку скрытная (intended) система спроектирована наилучшим образом для сокрытия своего сигнала под тепловым шумом и имеются веские основания полагать, что сигнал оказывает пренебрежимо малый эффект на дисперсию мгновенной мощности и, таким образом, на дисперсию отклика детектора. Поэтому, независимо от присутствия сигнала, значение дисперсии u_d можно считать неизменной, т.е.

$$D\{u_d\} = D\{Y^2\}/4 = \sigma_n^4 = (N_0W_i)^2. \quad (3.6)$$

С целью получения оценки среднего значения \bar{u}_d и регистрации его увеличения, обусловленного присутствием сигнала, интегратор в схеме на рис. 3.6 осуществляет усреднение во времени величины отклика детектора на интервале наблюдения T . Для того, чтобы сделать постоянную составляющую на выходе детектора заметной на фоне случайных флуктуаций необходимо их сгладить посредством интегрирования. Данная процедура возможна только тогда, когда флуктуации отклика детектора около \bar{u}_d достаточно быстры и изменяют свою полярность много раз за период T , что и позволяет обеспечить их компенсацию проведением усредняющего действия. Другими словами, число статистически независимых отсчетов n_s отклика детектора на интервале T должно быть достаточно большим. Протяженность во времени (время корреляции τ_c) автокорреляционной функции случайного процесса служит достоверной начальной аппроксимацией минимального интервала времени между отсчетами, начиная с которого отсчеты могут считаться независимыми. Поскольку отфильтрованное наблюдение занимает полосу W_i , время корреляции может быть оценено как $\tau_c \approx 1/W_i$, что обеспечивает $n_s \approx W_iT$ независимых отсчетов.

Хотя фактически процедура интегрирования может быть выполнена как непрерывная операция, ее результат оказывается достаточно близким с получаемым путем суммирования n_s независимых отсчетов [6, 9], что реализуется значительно проще особенно средствами цифровой техники. Для более точного выполнения анализа необходимо найти плотность распределения выходного значения интегратора \hat{E} для обеих гипотез (отсутствия и наличия сигнала) и затем, интегрируя по решающим областям, получить две вероятности: *ложной тревоги* и *правильного обнаружения*. Эта плотность вероятности подчиняется закону распределения хи-квадрат, который достаточно сложен и не поддается прозрачному физическому толкованию. Однако снова можно использовать тот факт, что сигнал является слабым и для его надежного обнаружения требуется интегрирование большого числа отсчетов n_s . Тогда центральная предельная теорема позволяет полагать, что плотность распределения величины \hat{E} на выходе интегратора является гауссовской и, следовательно, нахождение среднего и дисперсии \hat{E} достаточно для выполнения вычислений ранее упомянутых вероятностей.

Когда на входе присутствует только АБГШ, среднее и дисперсия \hat{E} определяются, как $\bar{\hat{E}} = n_s \bar{u}_{d0}$, $D\{\hat{E}\} = n_s D\{u_d\} = n_s \sigma_n^4$, где второй результат следует из (3.6) и статистической независимости интегрируемых отсчетов. Аналогично, когда наблюдаются АБГШ и сигнал, $\bar{\hat{E}} = n_s \bar{u}_{d1}$, однако дисперсия остается неизменной, поскольку сигнал слаб. Таким образом, плотности распределения на выходе интегратора, отвечающие гипотезам H_0 (сигнал отсутствует) и H_1 (сигнал присутствует) представимы в виде

$$W(\hat{E}|H_i) = \frac{1}{\sqrt{2\pi n_s \sigma_n^2}} \exp\left[-\frac{(\hat{E} - n_s \bar{u}_{di})^2}{2n_s \sigma_n^4}\right], i = 0, 1.$$

Когда \hat{E} превосходит порог E_t , ложная тревога происходит тогда, когда сигнал в действительности отсутствует, а правильное обнаружение, если сигнал достиг приемника.

Следовательно, вероятности P_f и P_d указанных событий определяются соответственно как

$$p_f = P(\hat{E} \geq E_t | H_0) = \int_{E_t}^{\infty} W(\hat{E} | H_0) d\hat{E} = Q\left(\frac{E_t - n_s \bar{u}_{d0}}{\sqrt{n_s} \sigma_n^2}\right)$$

и

$$p_d = P(\hat{E} \geq E_t | H_1) = \int_{E_t}^{\infty} W(\hat{E} | H_1) d\hat{E} = Q\left(\frac{E_t - n_s \bar{u}_{d1}}{\sqrt{n_s} \sigma_n^2}\right).$$

Перепишав второе соотношение в виде

$$p_d = Q\left(\frac{E_t - n_s \bar{u}_{d0}}{\sqrt{n_s} \sigma_n^2} - \frac{n_s (\bar{u}_{d1} - \bar{u}_{d0})}{\sqrt{n_s} \sigma_n^2}\right), \quad (3.7)$$

можно увидеть, что если допустимый уровень вероятности ложной тревоги задан заранее, то первая дробь в скобках (3.7) фиксирована, и вероятность обнаружения полностью определяется отношением

$$q_i = \frac{n_s (\bar{u}_{d1} - \bar{u}_{d0})}{\sqrt{n_s} \sigma_n^2} = \frac{n_s \bar{\Delta u}_d}{\sqrt{n_s D\{u_d\}}}. \quad (3.8)$$

Фактическое содержание последнего соотношения очевидно: это отношение сигнал-шум по напряжению на выходе интегратора, показывающее пропорцию между полезным (увеличением математического ожидания, обусловленное появлением сигнала) и мешающим (стандартное отклонение случайных флюктуаций) компонентами \hat{E} .

Применяя (3.4) и (3.6) совместно с соотношением $n_s = W_i T$ в (3.8), получаем

$$q_i = \sqrt{n_s} \frac{P}{WN_0} = \sqrt{W_i T} \frac{P}{WN_0}.$$

Это соотношение позволяет 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

$$q_i = \frac{P\sqrt{T}}{\sqrt{WN_0}} = \frac{q^2}{2\sqrt{WT}}, \quad (3.9)$$

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* . Обращение к рис. 3.7 позволяет дать физическое обоснование данного заключения. Расширение спектра сигнала с постоянной энергией и длительностью уменьшает уровень его спектральной плотности мощности, скрывая ее под спектром естественного теплового шума.

Пример 3.2.1. Рассмотрим систему, которая спорадически и достаточно редко передает одно из 64 сообщений с использованием ортогональных сигналов. Для обеспечения вероятности ошибочного приема не хуже, чем 10^{-3} , система работает при отношении сигнал-шум (SNR) порядка 7 дБ на бит или 15 дБ на 6-ти битовое сообщение (см. рис. 2.9). Таким образом, $q^2 = 15$ дБ, а преобразовав отношение сигнал-шум (по напряжению) пере-

хватчика (3.9) в дБ, имеем

$$(q_i)_{dB} = 2(q^2)_{dB} - 20\lg 2 - 10\lg WT.$$

При использовании системой широкополосных сигналов, для которых выигрыш от обработки составляет величину $WT = 1000$, для перехватчика рабочее отношение сигнал-шум стремится к $(q_i)_{dB} = -6$ дБ или $q_i = 1/2$, что является совершенно недостаточным для надежного обнаружения скрытной системы за один сеанс ее связи. Если, например, для перехватчика допустима вероятность ложной тревоги $p_f = 10^{-3}$, то, согласно (3.7), вероятность правильного обнаружения составляет $p_d \leq 5 \cdot 10^{-3}$, т.е. является чрезвычайно малой и не представляет серьезной угрозы скрытной системе. \square

Завершая параграф, отметим, что рассмотренные преимущества распределенного спектра в настоящее время широко используются не только военными потребителями и спецслужбами. То обстоятельство, что широкополосный сигнал практически незаметен для оборудования систем радиоконтроля, серьезным образом влияет на лицензионную политику. В частности, число коммерческих систем, активно работающих в эфире, не прибегая к получению лицензии, становится все шире, и в некоторых регионах выделяются специальные участки частотного диапазона для подобного безлицензионного использования.

3.3. 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. Одним из основных требований, стоящих перед подобными системами, является стремление избежать несанкционированного доступа к обслуживанию, адресованного только разрешенным пользователям, или искажения передаваемой информации. Глобальная спутниковая навигационная система GPS Navstar служит наглядным примером подобного рода. В ней используется два дальномерных канала (см. 11.1): открытый (обозначаемый как C/A) и специальный (P). Сигнал, применяемый во втором канале, позволяет с высокой точностью определять свое местоположение, и правительство США, руководящее системой, не считает рациональным разрешать безусловный доступ к этому каналу. Для его защиты от несанкционированного использования предприняты специальные меры, касающиеся закона модуляции сигнала.

В дисциплинах, имеющих дело с безопасностью информации, степень защиты данных определяется числом равновероятных конкурирующих *ключей*, с помощью которых противостоящий криптоаналитик (eavesdropper) старается взломать шифротекст, т.е. зашифрованные данные. В применении к структуре сигнала каждый из таких ключей есть ничто иное, как закон модуляции, который, как правило, повторяется с периодом T . Предположим, что сигнал построен на основе M -ичного алфавита, т.е. возможны M реализаций индивидуального сигнального элемента (чипа). Если полоса, отводимая системе, равна W , то общее сигнальное пространство имеет размерность, определяемое как WT (игнорируя возможность его удвоения при переходе к полосным сигналам, см. 2.3-2.5), т.е. закон модуляции может быть сконструирован посредством WT чипов. Очевидно, что величина M^{WT} определяет общее число различных законов модуляции, т.е. число конкурирующих ключей, и, значит, системный конструктор, отвечающий за секретность модуляционного формата разрабатываемой системы, должен использовать сигналы с достаточно большим выигрышем от обработки.

Пример 3.3.1. Сигнал P-канала (P-код) в системе GPS Navstar является бинарным ($M = 2$) с полосой $W \approx 10$ МГц. Его структура имеет регулярный характер и повторяется с периодом $T = 7$ суткам. Будучи скрытым под тепловым шумом, этот сигнал не может быть восстановлен путем посимвольного приема и только знание его тонкой структуры позволяет эффективно очистить его от АБГШ. Для раскрытия структуры данного сигнала необходимо проанализировать 2^{WT} альтернатив, а поскольку $WT = 7 \times 86400 \times 10^7 > 10^{12}$, то число ключей больше, чем 2 в степени 10 с двенадцатью нулями, что невозможно и вообразить. По этой причине P-код считается не раскрываемым, и за 20 лет истории GPS не появилось ни одной статьи с сообщениями об успешных попытках его взлома. \square

Заклучим параграф еще одним заявлением о достоинствах широкополосности:

данная технология в значительной степени способствует криптографической защите структуры сигнала.

3.4. 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 [15]. Мотивацией подобного поведения являются не только этические принципы, но и 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 (см. 3.2), 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 [16].

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 (см. 3.1) снова применимы. Следовательно, *широкополосная философия хорошо ложится в русло обеспечения ЭМС.*

3.5. Эффекты распространения в беспроводных системах.

Для исследования следующего достоинства распределенного спектра необходимо приобрести дополнительные сведения об эффектах распространения волновых колебаний в беспроводных каналах, и данный параграф будет посвящен знакомству с этой областью. Прежде всего, ключевым параметром, определяющим характеристику любой задачи приема, является интенсивность сигнала, или отношение сигнал-шум (SNR). Конечно, энергия сигнала или его мощность во всех предшествующих формулах, устанавливающих вероятность ошибки, дисперсию оценки и т.п., характеризовали уровень сигнала на входе приемника. Следовательно, необходимо иметь возможность предсказывать интенсивность сигнала в некоторой точке пространства, удаленной от передающей антенны, учитывая эффекты, сопровождающие распространение электромагнитной волны.

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. Вследствие их воздействия принимаемый сигнал искажается не только аддитивным шумом (АБГШ), но и *мультипликативной* помехой, название которой вытекает из того факта, что она изменяет интенсивность сигнала, т.е. перемножается с амплитудой сигнала.

3.5.1. Распространение в свободном пространстве.

Начнем рассмотрение с идеализированной модели распространения в свободном пространстве (см. рис. 3.8), где отсутствуют препятствия между передающей и приемной антеннами, и излученная волна распространяется по единственно возможному пути, называемому *линией прямой видимости (line-of sight (LOS))*.

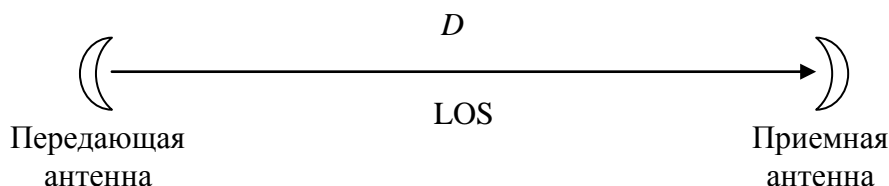


Рис. 3.8. Модель распространения в свободном пространстве.

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 $P_r = P_t A_r / 4\pi D^2$. Если же передающая антенна является направленной, то в сторону приемной антенны излучается мощность, которая в G_t раз превосходит всенаправленную, а величина G_t называется *коэффициентом усиления мощности* передающей антенны. В этом случае принятая мощность также увеличивается в G_t раз. Для представления принятой мощности в симметричной форме воспользуемся зависимостью между 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 [4]

$$P_r = P_t G_t G_r \left(\frac{\lambda_w}{4\pi D} \right)^2, \quad (3.10)$$

показывающей, что ослабление мощности сигнала по линии прямой видимости в свободном пространстве обратно пропорционально квадрату расстояния.

Модель свободного пространства может быть непосредственно применена для каналов связи, среда распространения в которых представляет собой открытое пространство, например, между космическими носителями или летательными аппаратами, наземным центром контроля и космическим носителем и т.п. Среда распространения наземных систем намного менее благоприятна и основное влияние на интенсивность сигнала оказывают два следующих компонента: *затенение (shadowing)* и *фединг (замирение) вследствие многолучевого распространения (multipath fading)*.

3.5.2. Затенение.

Затенение обусловлено деталями ландшафта, препятствующими распространению по прямой линии: возвышенностями, лесами, чащами, зданиями и т.п. Вследствие их влияния интенсивность сигнала падает с расстоянием значительно быстрее, чем это предсказывает соотношение (3.10). Очевидно, что нерегулярный характер земной поверхности делает невозможной или бесполезной попытку создать некоторую универсальную теоретическую модель затенения. С целью накопления необходимых сведений об общем характере зависимости между принимаемой мощностью и длиной пути распространения было выполнено множество натуральных испытаний, на основании которых был предложен целый ряд эмпирических моделей [17-19]. Специалистами в области мобильной связи одной из наиболее популярных признана модель Окумура–Хаты. Согласно последней зависимость средней принятой мощности \bar{P}_r подчиняется соотношению $\bar{P}_r = kP_t / D^e$, в котором конкретное значение показателя e зависит от типа подстилающей поверхности, изменяясь от 3 (для сельской местности) до 5 (для плотной городской застройки), а коэффициент k – определяется частотным диапазоном и высотой антенн [2, 6, 15, 19]. Принимаемая мощность, определяемая данной моделью, дает лишь очень грубую отправную точку, получаемую путем усреднения по различным положениям приемника, находящегося на одинаковом удалении D от передатчика. Флюктуации P_r по дуге радиуса D с центром в месте расположения передатчика значительны и хорошо описываются логнормальным законом распределения, означающем, что распределение принятой мощности, выраженной в децибелах $x = 10 \lg P_r$, является гауссовским (нормальным):

$$W(x) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp \left[-\frac{(x - \bar{x})^2}{2\sigma_x^2} \right].$$

Общепринятым, согласно библиографическим источникам, считается, что величина стандартного отклонения σ_x величины $10 \lg P_r$ в последнем соотношении лежит в диапазоне 6–12 дБ.

Уменьшение мощности, обусловленное затенением, носит статический характер и даже тогда, когда приемник находится в движении, изменение мощности во времени, как правило, происходит сравнительно медленно вследствие значительного размера компонентов ландшафта (десятки и сотни метров). По этой причине затенение часто рассматривается как *крупномасштабный фединг (замирение) (large scale or long term fading)*.

3.5.3. Замирения вследствие многолучевого распространения.

Обратимся теперь ко второму фактору, влияющему на интенсивность принимаемого сигнала: *многолучевое распространение (multipath propagation)*. Фактически излучен-

ный сигнал может достичь приемной антенны разнообразными путями. Путь прямого распространения может оказаться либо одним из них, либо отсутствовать при крайне сложном распространении, тогда как другие пути образуются в результате отражения излученной волны различными объектами. В качестве подобных отражателей могут выступать здания, вышки, летательные аппараты, подстилающая поверхность и многое другое (см. рис.3.9).

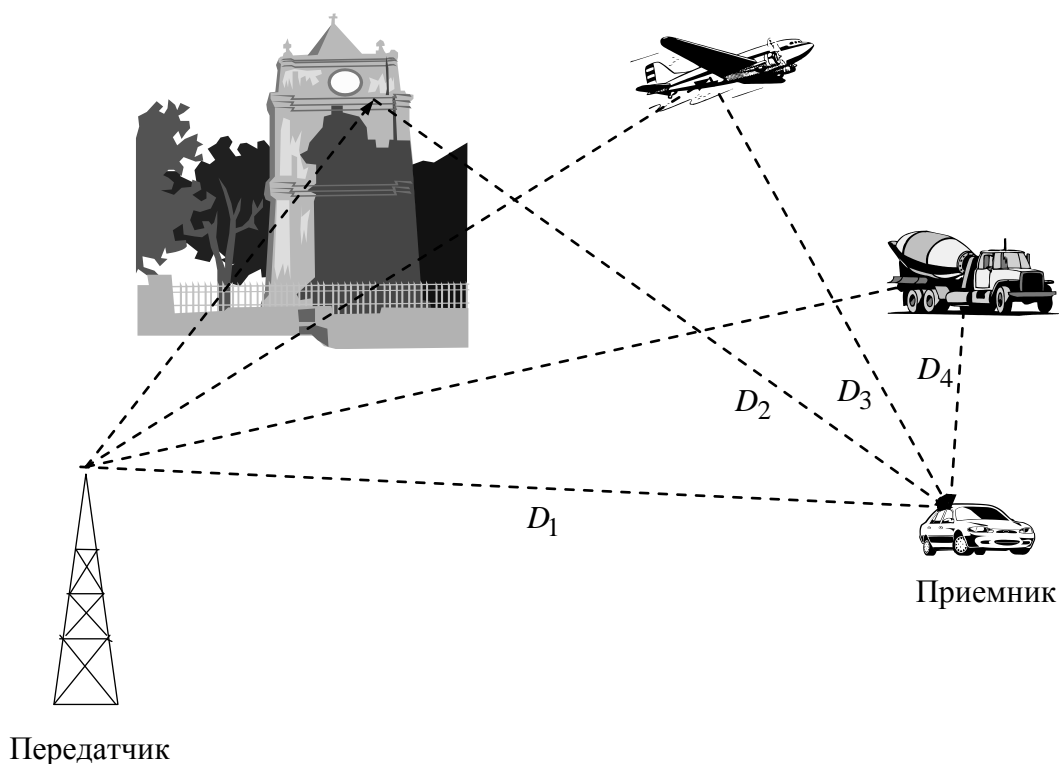


Рис.3.9. Иллюстрацию многолучевого распространения.

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

$$\dot{S}_r(t) = \sum_i A_i \dot{S}(t - \tau_i) \exp(j\varphi_i), \quad (3.11)$$

где, без потери общности, вещественная амплитуда исходного сигнала принята равной единице. В том случае, когда *задержка распространения (delay spread)* τ_{ds} , т.е. максимальное значение взаимной задержки между сигналами различных путей, не превосходит длительности сигнала, все многолучевые сигналы перекрываются и интерферируют между собой. Для лучшего понимания этого явления первоначально рассмотрим простейший сценарий распространения, имеющий место в мобильной связи, телевизионном вещании и др.

Пример 3.5.1. Рис. 3.10 демонстрирует ситуацию, когда излученный сигнал достигает приемника по путям, обусловленным двумя отражателями (зданиями, машинами и т.п.), путь распространения по прямой (LOS) полностью блокирован препятствием (зданием), в результате чего в (3.11) присутствуют только две составляющие. Отражатели ориентированы в пространстве таким образом, что они образуют вторичные волны, направленные друг на друга. Приемник, расположенный на линии, соединяющей отражатели, будет наблюдать суперпозицию двух интерферирующих колебаний, разность фаз φ кото-

рых определяется отношением разности путей распространения $\delta = D_1' + D_1'' - D_2' - D_2''$ к длине волны λ_w : $\varphi = 2\pi\delta/\lambda_w$. Считая амплитуды отраженных сигналов в точке приема, равными A_1, A_2 , результирующая амплитуда A_r может быть найдена с помощью теоремы

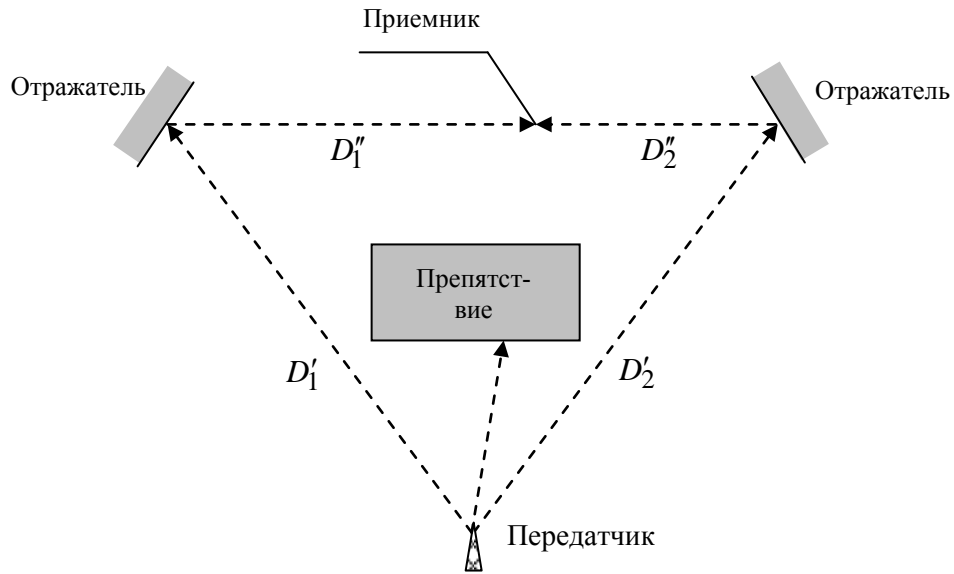


Рис. 3.10. Вариант с двумя путями распространения.

косинусов как $A_r = \sqrt{A_1^2 + A_2^2 + 2A_1A_2 \cos \varphi}$ (см. фазовую диаграмму на рис. 3.11, *a*). Периодичность A_r , как функции φ , свидетельствует о ее периодичности в зависимости от разности распространения δ . При движении приемника вдоль линии, соединяющей отражатели, его смещение на величину $\lambda_w/2$ в любом направлении изменяет δ на одну длину волны λ_w , так что φ изменяется на 2π , и значения A_r в точках, разнесенных на $\lambda_w/2$, одинаковы. Другими словами, интерференция двух наложившихся колебаний образует стационарную волну с периодом $\lambda_w/2$. Двигаясь вдоль тестовой линии, на входе прием-

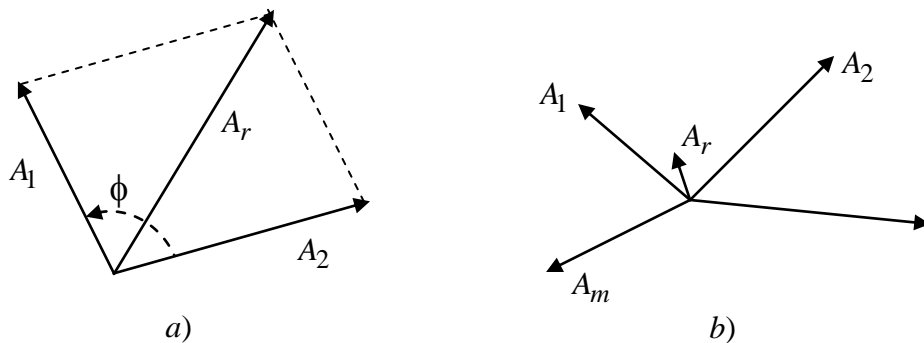


Рис. 3.11. Интерпретация эффектов многолучевости фазовой диаграммой.

ника будут чередоваться максимумы $A_1 + A_2$ и минимумы $|A_1 - A_2|$ амплитуды наблюдения каждые $\lambda_w/2$ метров. Если амплитуды отраженных сигналов близки по величине (достаточно вероятный случай), то результирующая мощность $P_r = A_r^2/2 = (A_1^2 + A_2^2)/2 + A_1A_2 \cos \varphi$ спадает почти (или в точности) до нуля, когда прием-

ник проходит узлы колебания. Именно этот феномен получил название многолучевого фединга (замирания, обусловленного многолучевым распространением). Поскольку пространственное расстояние между соседними пиками P_r сравнимо с длиной волны, то для систем, работающих в метровом и дециметровом диапазоне, временные циклы изменения P_r на входе мобильного приемника будут достаточно короткими (типичны значения секунд). □

График изменения P_r в зависимости от времени, представленный на рис. 3.12, отвечает значениям параметров, типичным для мобильной связи: $\lambda_w \approx 0.3$ м и скорость движения приемника $V_r = 60$ км/час. Как видно из диаграммы, даже при столь незначительной скорости передвижения изменение принимаемой мощности происходит достаточно быстро. Это служит объяснением тому, что многолучевому федингу присваивают другое название: *мелкомасштабный фединг* (*short-term fading* или *small-scale fading*).

Очевидно, что только что рассмотренный пример искусственно упрощен, чтобы рассмотреть указанный эффект наиболее явным образом. В действительности, число од-

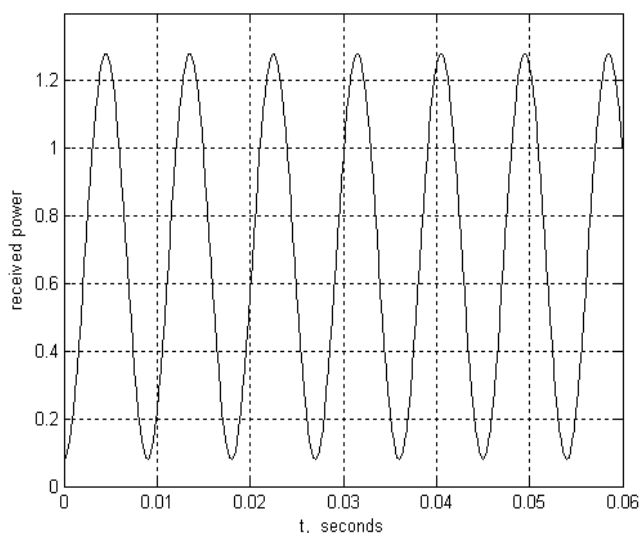


Рис. 3.12. Временной профиль принятой мощности при 2-х лучевом фединге.

новременно принимаемых многолучевых сигналов L может быть достаточно большим, в результате чего интерферирующий образец становится более сложным. Фазовая диаграмма на рис. 3.11, *b*, иллюстрирует именно этот случай. Хаотический характер распределения отражателей или рассеивателей в окружении приемника делает интерферирующий образец непредсказуемым, а более подходящим становится его статистическое описание.

Пример 3.5.2. На рис.3.13 представлен временной профиль принимаемой мощности, полученный в результате моделирования в среде Matlab обстановки распространения с пятью отражателями, расположенными равновероятно в пределах квадрата со стороной, равной начальному расстоянию между передатчиком и приемником в $D = 30$ км. Принимаемая мощность нормирована к средней. Длина волны и скорость объекта составляют 0.3 м и 60 км/час соответственно. Нерегулярный характер изменения мощности достаточно очевиден, также как наличие глубоких провалов в интенсивности принимаемого сигнала. □

Согласно центральной предельной теореме суперпозиция независимых и примерно равных по вкладу случайных слагаемых стремится к гауссовскому распределению по мере роста их числа. Следовательно, множество распространяющихся по разным путям сигнала-

лов, подчиняющихся приведенным условиям, 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. Однако огибающая данного процесса подчиняется релеевскому распределению (см. параграф 3.2) и, таким образом, приходим к модели канала с релеевскими замираниями (*Rayleigh fading channel*). В результате принятая амплитуда A_r является не детерминированной, а случайной, удовлетворяющей релеевскому закону плотности распределения

$$W(A_r) = \begin{cases} 2A_r \exp(-A_r^2), & A_r \geq 0, \\ 0, & A_r < 0. \end{cases} \quad (3.12)$$

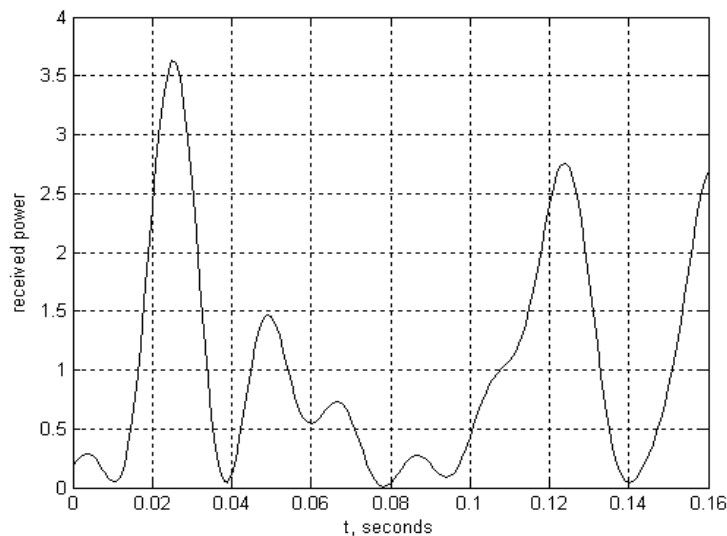


Рис. 3.13. Профиль принятой мощности при 5-ти лучевом фединге.

Поскольку в произведении $A_r \dot{S}(t)$ «истинная», действительно измеряемая амплитуда поделена между двумя сомножителями, что может быть сделано произвольным образом, в соотношении (3.12) предполагается соответствующая нормировка, устанавливающая средний квадрат A_r равным единице: $\overline{A_r^2} = 1$. График зависимости плотности распределения вероятности (3.12) представлен на рис. 3.14.

Релеевские замирания характерны для многих систем, включая, наряду со связными, также локационные, навигационные и др. Присущие им глубокие провалы в интенсивности сигнала, как правило, не нейтрализуются спорадическим увеличением A_r в случае прихода многолучевых сигналов с близкими фазами. В результате эффект релеевского замирания носит достаточно вредный характер.

3.5.4. Анализ характеристик.

Рассмотрим в качестве примера передачу бинарных данных по релеевскому каналу с медленными и амплитудными (*flat*) замираниями. Первое определение означает, что помеховый образец остается стабильным в течение многих символов и текущая опорная фаза может быть удалена из принятого сигнала путем усреднения за соответствующий интервал времени. Другими словами, случайность сигнала не исключает бинарную ФМ из допустимых вариантов. 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*. Для понимания уместности использования слова “flat” обратимся к (3.11) и отметим, что адекватной моделью многолучевого канала может служить линия задержки с отводами, отвечающими значениям τ_i , и весовыми коэффициентами $A_i \exp(j\varphi_i)$. Передаточная

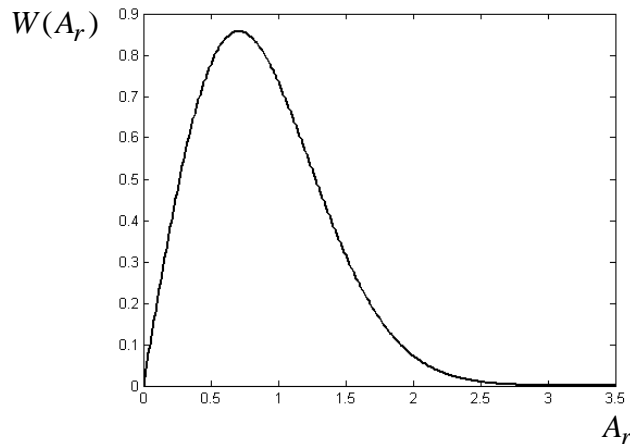


Рис. 3.14. Релеевская плотность распределения.

функция такой системы сильно зависит от задержек отводов и при $\tau_{ds} \ll T_b$ достаточно равномерна (flat) в полосе сигнала, так что все частотные компоненты сигнала искажаются одинаково, а форма сигнала остается неизменной. Единственным видом искажений, которым в этом случае подвержен сигнал вследствие многолучевого распространения, являются релеевские флюктуации амплитуды, описываемые (3.12).

Рис. 3.15 служит иллюстрацией введенным определениям. Диаграммы, построенные в результате моделирования в среде Matlab, демонстрируют медленные амплитудные замирания (a) в противопоставлении быстрым амплитудным замираниям (b) для случая колоколообразных символьных импульсов. Второй тип замираний характеризуется быстрым изменением во времени интерференционного образца, так что искажения последовательных символов становятся практически независимыми.

Пусть энергия принятого сигнала, отвечающая случаю $A_r = 1$, равняется E . Тогда энергия сигнала с другим значением амплитуды составит $E(A_r) = A_r^2 E$, а средняя энергия снова будет E вследствие ранее принятой нормировки: $\overline{E(A_r)} = \overline{A_r^2} E = E$. В случае, когда амплитуда принятого сигнала полагается фиксированной и равной A_r , для вычисления условной вероятности ошибки $P_e(A_r)$ может быть использовано соотношение (2.19):

$$P_e(A_r) = Q\left(\sqrt{\frac{2E(A_r)}{N_0}}\right) = Q\left(A_r \sqrt{\frac{2E}{N_0}}\right) = Q(A_r q_b),$$

где отношение сигнал-шум $q_b = \sqrt{2E/N_0}$ отвечает сигналу с энергией $E(A) = E$, т.е. с амплитудой $A_r = 1$. Фактическая амплитуда A_r является случайной и изменяется от одного сеанса приема к другому в соответствии с релеевской плотностью распределения (3.12). Естественно тогда характеризовать качество передачи данных величиной $P_e(A_r)$, усредненной по всем A_r . Сохраняя теперь термин «вероятности ошибки» и обозначение P_e за этим математическим ожиданием, имеем

$$\begin{aligned}
P_e = \overline{P_e(A_r)} &= \int_0^{\infty} P_e(A_r) W(A_r) dA_r = 2 \int_0^{\infty} Q(A_r q_b) A_r \exp(-A_r^2) dA_r = \\
&= \frac{2}{\sqrt{2\pi}} \int_0^{\infty} \int_0^{\infty} A_r \exp\left(-\frac{x^2}{2}\right) \exp(-A_r^2) dx dA_r,
\end{aligned}$$

где использовано определение дополнительной функции ошибки $Q(\cdot)$. Поменяв порядок интегрирования, получаем

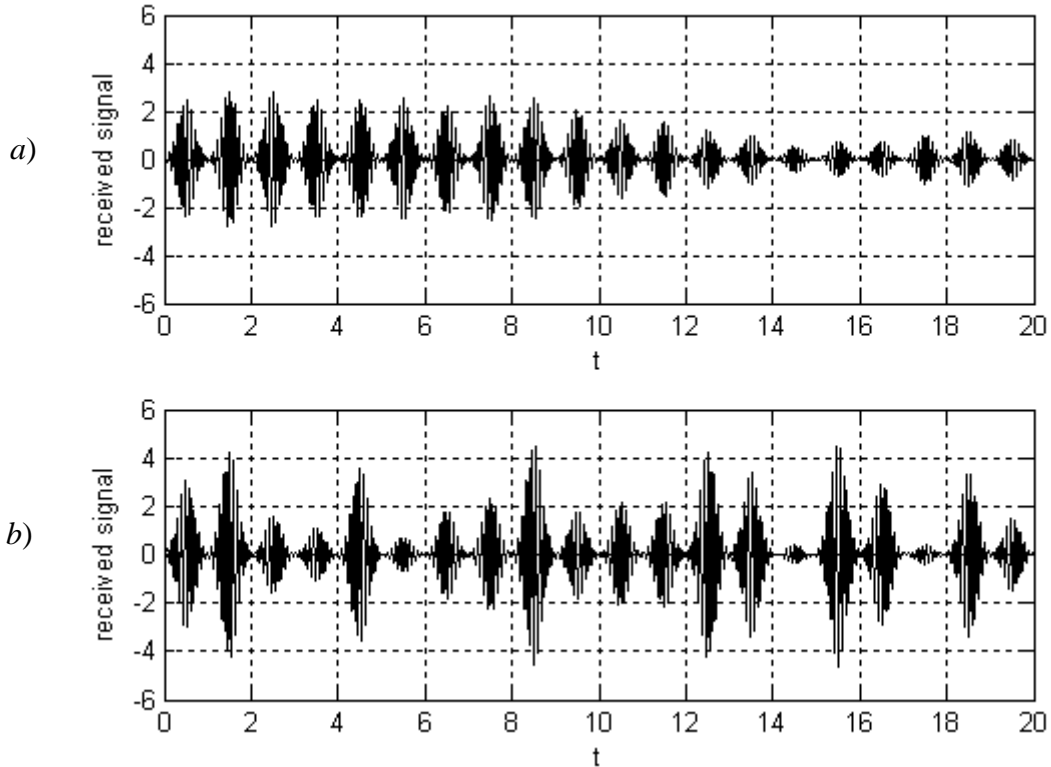


Рис. 3.15. Медленные (а) и быстрые (b) замирания.

$$\begin{aligned}
P_e &= \frac{1}{\sqrt{2\pi}} \int_0^{\infty} \left[\int_0^{x/q_b} 2A_r \exp(-A_r^2) dA_r \right] \exp\left(-\frac{x^2}{2}\right) dx = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} \left[1 - \exp\left(-\frac{x^2}{q_b^2}\right) \right] \exp\left(-\frac{x^2}{2}\right) dx = \\
&= \frac{1}{\sqrt{2\pi}} \int_0^{\infty} \exp\left(-\frac{x^2}{2}\right) dx - \frac{1}{\sqrt{2\pi}} \int_0^{\infty} \exp\left[-\frac{(q_b^2 + 2)x^2}{2q_b^2}\right] dx,
\end{aligned}$$

где первый член равен $Q(0) = 1/2$, а второй приводится к аналогичной форме после умножения его на $(\sqrt{q_b^2 + 2})/q_b$. Окончательно имеем

$$P_e = \frac{1}{2} \left(1 - \frac{q_b}{\sqrt{q_b^2 + 2}} \right). \quad (3.13)$$

Для количественной оценки степени вредного влияния замираний обратимся к рис. 3.16, на котором представлены вероятности ошибки передачи бинарных ФМ сигналов по гауссовскому и релеевскому каналам. Как следует из графиков, вероятность ошибки

$P_e = 10^{-3}$ гарантируется для АБГШ канала при отношении сигнал-шум на бит примерно равном 10 дБ, тогда как для релейского канала подобная достоверность передачи требует энергетических затрат не менее 27 дБ, т.е. в 50 раз выше. Подобные внушительные энергетические потери, обусловленные замираниями, возрастают еще больше при повышении

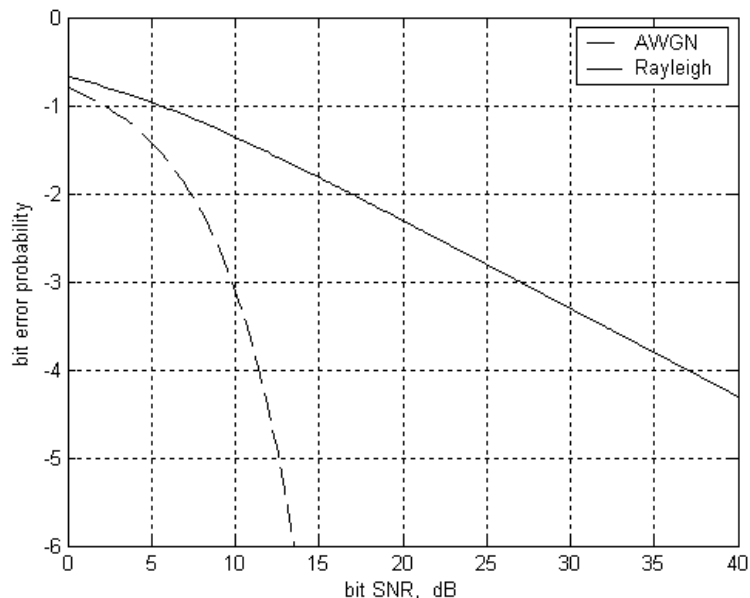


Рис. 3.16. Вероятность ошибки на бит ($\lg P_e$) для АБГШ и релейского каналов

достоверности передачи и становятся близкими к 25 дБ (300 раз) при $P_e = 10^{-4}$.

Физическое толкование весьма вредного эффекта замираний достаточно очевидно. Резкие спорадические провалы интенсивности сигнала, вызванные помехами многолучевости, присущи релейскому каналу. Интегрирование плотности вероятности (3.12) показывает, например, что вероятность уменьшения A_r до уровня ниже 0.4 (снижение отношения сигнал-шум на 8 дБ) составляет 0.15. Тогда из графика на рис. 3.16 (сплошная линия) следует, что в подобных условиях вероятность ошибки не может быть ниже 0.1 при исходном отношении сигнал-шум 10 дБ. Поскольку доля таких сеансов приема равна 0.15, то ее вклад в полную (среднюю) вероятность ошибки будет не меньше, чем $0.1 \times 0.15 = 0.015$, т.е. в 15 раз больше, чем значение, соответствующее исходному отношению сигнал-шум. Данный эффект никаким образом не может быть компенсирован за счет возможных благоприятных сеансов с большим отношением сигнал-шум, поскольку их вклад в полную вероятность ошибки никогда не будет отрицательным.

В случае, когда замирания оказываются *частотно-селективными*, многолучевое распространение потенциально может иметь даже более драматические последствия. Указанный термин в противопоставлении к определению “flat” (ровный) описывает ситуацию, при которой передаточная функция канала не будет равномерной в полосе сигнала. Подобная ситуация имеет место тогда, когда задержка распространения охватывает несколько излученных бит, так что на выходе канала предшествующий бит перекрывается с текущим. Для противодействия данной межсимвольной интерференции применяются специальные фильтры (эквалайзеры), которые спрямляют неравномерную передаточную функцию канала. С другой стороны, при соответствующем использовании частотная селективность может служить дополнительным источником борьбы с замираниями, позволяя организовать многолучевое разнесение, рассматриваемое в параграфе 3.7.

3.6. Разнесение.

Основная идея борьбы с деструктивными эффектами многолучевости заключается в использовании *разнесения* (*diversity*), которое означает организацию нескольких независимых каналов передачи или *ветвей* (*branches*). Вследствие этого, несмотря на то, что каждая из ветвей остается подверженной релеевским замираниям, вероятность того, что интерференционные (помеховые) образцы во всех из них будут одновременно неудачными, определяется правилом умножения вероятностей и, таким образом, не может оказаться значительной. Обратимся к цифрам примера, приведенного в конце последнего параграфа, и предположим, что каким-либо способом организованы две независимые идентичные ветви. Тогда вероятность одновременного снижения мощности сигнала в обоих из них до указанного уровня составит $0.15^2 = 2.25 \cdot 10^{-3}$, т.е. станет заметно меньше по сравнению с вероятностью неблагоприятных условий в индивидуальной ветви. С увеличением числа ветвей это выигрыш от разнесения становится все более и более существенным. Параллельная работа ветвей может рассматриваться так, как будто они обеспечивают друг другу ослабление ухудшений, вызванных замираниями.

Другими словами, известно, что ухудшение характеристик, связанное с многолучевыми замираниями, полностью обусловлено глубокими падениями в отношении сигнал-шум, которые время от времени происходят. Следовательно, 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*.

3.6.1. Методы комбинирования.

На приемной стороне могут быть применены различные варианты комбинирования результатов обработки сигналов, пришедших по различным ветвям. 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 $i = 1, 2, \dots, n_d$. Каким тогда будет наилучший алгоритм линейной обработки, обеспечивающий максимально возможное результирующее отношение сигнал-шум? Любое линейное объединение откликов ветвей $u_i, i = 1, 2, \dots, n_d$ представляется в виде взвешенной суммы $\sum_{i=1}^{n_d} w_i^* u_i$, где w_i^* – весовые коэффициенты, в общем случае комплексные. Тогда результирующее отношение сигнал-шум по мощности q_r^2 представляет собой отношение величины детерминированной компоненты этой суммы к дисперсии ее шумовой компоненты. Последнее есть просто взвешенная с коэффициентами $|w_i|^2$ сумма дисперсий шума ветвей, поскольку сами они являются независимыми. Таким образом, имеем

$$q_r^2 = \frac{\left| \sum_{i=1}^{n_d} w_i^* A_i \exp(j\varphi_i) \right|^2}{\sum_{i=1}^{n_d} |w_i|^2 \sigma_i^2}. \quad (3.14)$$

Сумма в числителе соотношения (3.14) может рассматриваться как скалярное произведение двух n_d -мерных векторов, компонентами которых являются $w_i \sigma_i$ и $(A_i / \sigma_i) \exp(j\varphi_i)$ соответственно. Не существует модуля любого скалярного произведе-

ния большего произведения длин векторов, что известно как неравенство Шварца [1, 2]. Следовательно

$$q_r^2 \leq \frac{\left(\sum_{i=1}^{n_d} |w_i|^2 \sigma_i^2 \right) \left(\sum_{i=1}^{n_d} \frac{A_i^2}{\sigma_i^2} \right)}{\sum_{i=1}^{n_d} |w_i|^2 \sigma_i^2} = \sum_{i=1}^{n_d} \frac{A_i^2}{\sigma_i^2} = \sum_{i=1}^{n_d} q_i^2, \quad (3.15)$$

где $q_i = A_i / \sigma_i$ – отношение сигнал-шум по напряжению в i -й ветви разнесения. При оптимальном выборе весов

$$w_i^* = \frac{A_i}{\sigma_i} \exp(-j\varphi_i),$$

неравенство (3.15) обращается в равенство, т.е. достигается максимально возможное результирующее отношение сигнал-шум. Выбор таких весов, как это легко заметить, 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. В литературе подобная техника комбинирования известна как *техника с максимальным отношением (maximal-ratio technique)* [5,18].

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.

На практике находят применение и другие методы комбинирования, поскольку обработка с максимальным отношением требует дополнительного оборудования (может оказаться необходимым введение некоторого специального пилотного сигнала для измерения отношения сигнал-шум и фазы в ветви разнесения и т.п.). Альтернативными методами являются метод комбинирования с равными весами и метод выбора ветви с максимальным отношением сигнал-шум. Первый достигает эффективности способа с максимальным отношением только при одинаковом отношении сигнал-шум во всех ветвях разнесения. Выигрыш второго 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.

Рассмотрим теперь традиционные способы образования независимых ветвей разнесения.

3.6.2. Организация ветвей разнесения.

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

- *пространственное разнесение (space diversity)*;
- *разнесение по частоте (frequency diversity)*;
- *разнесение во времени (time diversity)*;
- *поляризационное разнесение (polarization diversity)*;

– *многолучевое разнесение (multipath diversity)*.

Пространственное разнесение предполагает создание нескольких независимых путей распространения за счет использования многих антенн, что объясняет другое распространённое название этого метода: *антенное разнесение (antenna diversity)*. 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 на входе приемника. Ко-

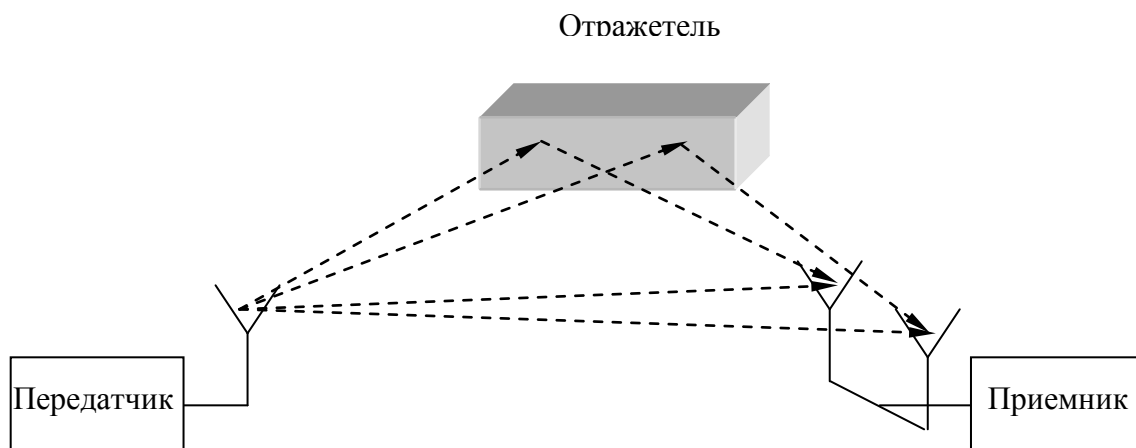


Рис. 3.17. Организация разнесения на приеме.

гда несколько антенн применяются на приемной стороне (см. рис. 3.17) (*разнесение на приеме – receive diversity*), 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, описанной выше.

Разнесение с помощью передающих антенн (*transmit diversity*) не реализуется столь прямолинейно. Во-первых, как следует из рис. 3.18, ограниченный полный энергетический ресурс передатчика необходимо распределить между несколькими передающими ан-

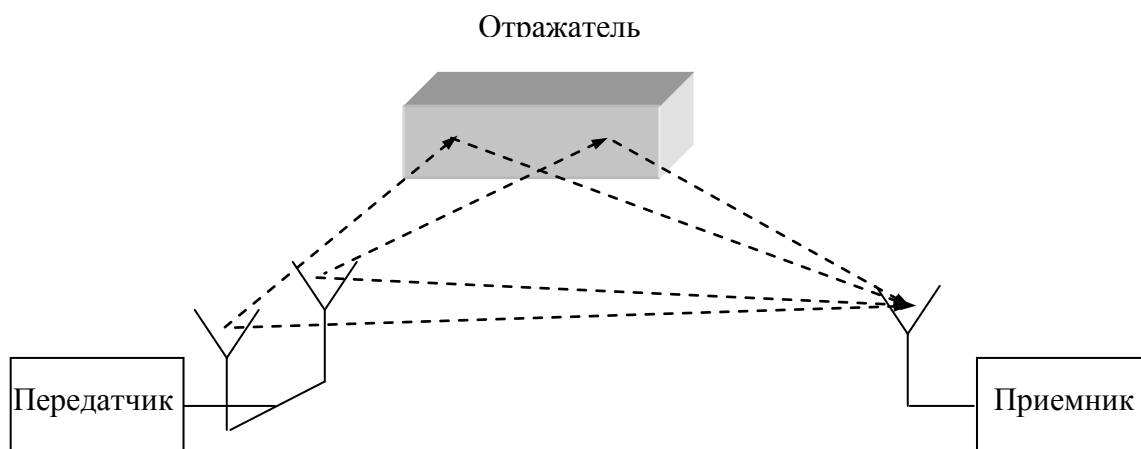


Рис. 3.18. Организация разнесения на передаче.

теннами. 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* – *space-time coding* (см. параграф 10.3).

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* (*coherence bandwidth*). 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. Как уже подчеркивалось в предыдущем параграфе, 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. Можно сказать, что разнесение по частоте использует частотную селективность замираний на благое дело. Рис 3.19 предоставляет элементарное пояснение данной идее. Колебания с двумя длинами волн λ_{w1} и λ_{w2} , распространяющиеся двумя одинаковыми маршрутами, характеризуются идентичным геометрическим приращением пути распространения δ . Однако разность фаз между сигналами двух путей для каждой длины волны будет строго индивидуальна и равна соответственно $2\pi\delta/\lambda_{w1}$ и $2\pi\delta/\lambda_{w2}$. Когда одна из этих разностей фаз приведет к ослаблению результирующего сигнала, другая может повести себя не столь деструктивно. При множестве параллельных путей распространения настоящее статистическое обоснование набирает силу и разность частот, превосходящая канальную полосу когерентности, обеспечивает независимость ветвей разнесения в этой схеме. Соответствующий выбор частот в рассматриваемой схеме разнесения гарантирует разделение ветвей в приемнике с помощью полосной фильтрации.

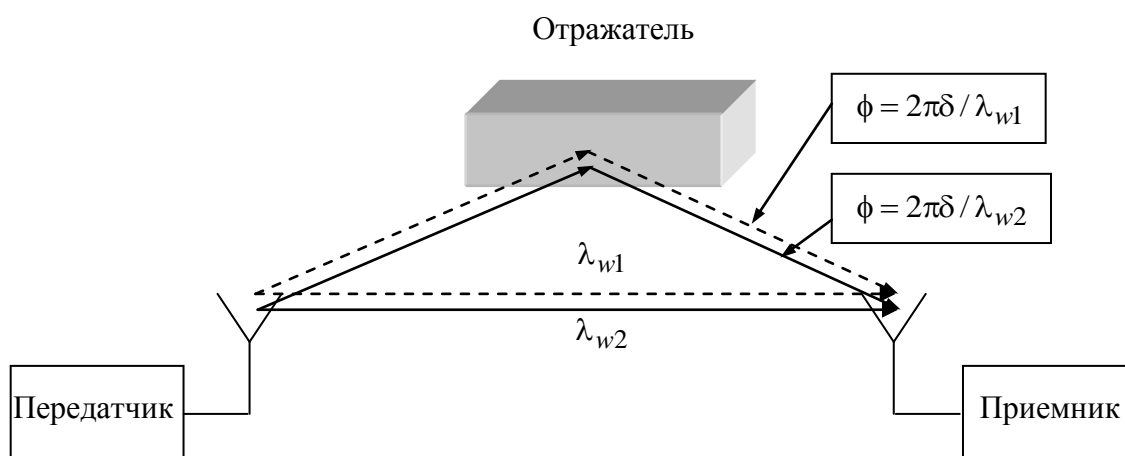


Рис. 3.19. Разнесение по частоте.

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 (*coherence time*), интервал времени, в течение которого мощность принимаемого сигнала остается приблизительно стабильной. Снова обратим внимание на дуальность частоты и времени: диапазон коррелированности в частотной области (полоса когерентности – *coherence bandwidth*) обратно пропорционален распределению во времени (распределению задержки – *delay spread*), тогда как во временной области он (время когерентности – *coherence time*) обратно пропорционален 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. С незначительными изменениями этот принцип используется в телекоммуникационных системах в форме *перемежения* (*interleaving*).

Поляризационное разнесение, основанное на различии профилей многолучевости колебаний с разной поляризацией, не нашло до сих пор широкого применения. Что же касается последнего пункта приведенного перечня, то в рассматриваемом контексте он играет особую роль и будет обсужден отдельно в следующем параграфе.

3.7. Многолучевое разнесение и Rake–приемники.

Обычная схема разнесения по частоте, рассмотренная выше, использует параллельную передачу одного и того же сигнала на нескольких несущих, расстояние между которыми превосходит полосу когерентности канала. Как правило, подобный метод используется тогда, когда спектр сигнала уже полосы когерентности канала, что предполагает только медленные замирания. Другой версией разнесения по частоте является многолучевое разнесение, использующее сигналы со спектром, значительно превышающим полосу когерентности. Вследствие этого замирания становятся частотно-селективными, допускающими в принципе разрешение по времени многолучевых сигналов. Таким образом, 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 вида (3.11) прошел фильтр, согласованный с сигналом $s(t)$. Тогда, учитывая линейность фильтрации и связь между АКФ сигнала и откликом согласованного фильтра (см. 2.11), выражение для комплексной огибающей $\dot{S}_{rf}(t)$ сигнала на выходе фильтра имеет вид

$$\dot{S}_{rf}(t) = \sum_i A_i \dot{R}(t - \tau_i - T) \exp(j\varphi_i), \quad (3.16)$$

где $\dot{R}(\tau)$ – АКФ комплексной огибающей $\dot{S}(t)$ исходного сигнала, определенного соотношением (2.66), а T – как обычно, 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 сиг-

нал должен иметь узкую АКФ. Как показало обсуждение в параграфе 2.15, 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 могут быть найдены только среди сигналов с распределенным спектром и, значит, в перечень достоинств распределенного спектра может быть добавлено еще одно: *возможность построения схемы многолучевого разнесения*.

Метод многолучевого разнесения уникален в том смысле, что он радикально изменяет отношение к эффектам многолучевости, которые на первый взгляд представляются

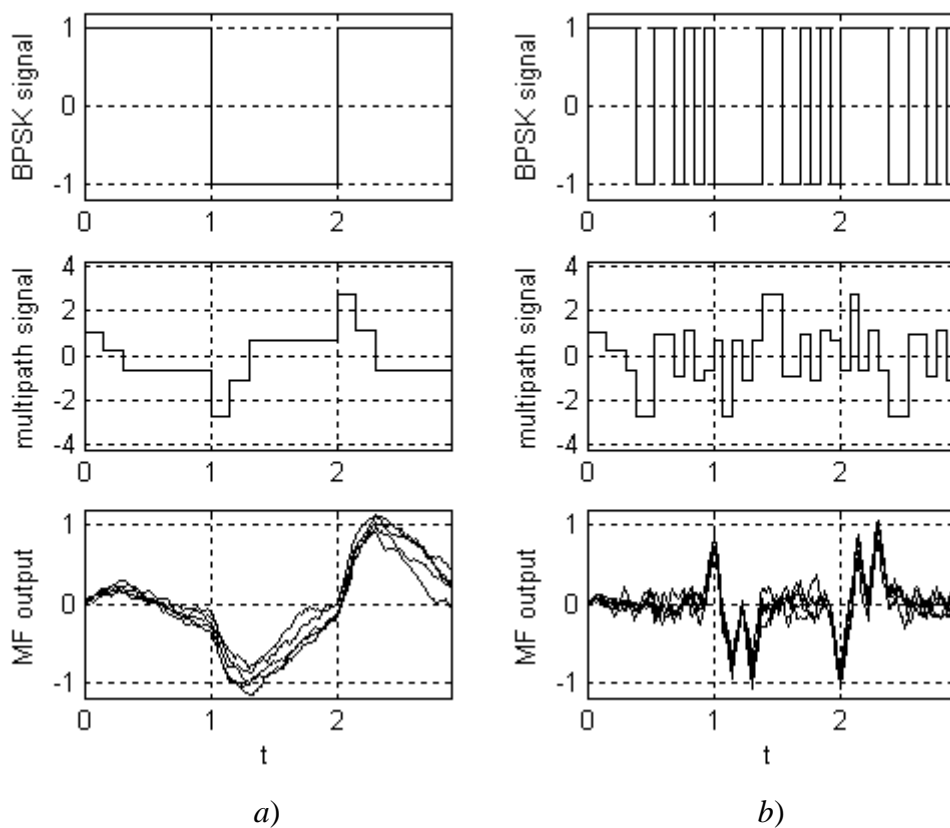


Рис. 3.20. Эффекты многолучевости при бинарной ФМ передаче:
a) простые сигналы, *b)* сигналы с распределенным спектром.

безоговорочно вредными. Как показывает ранее приведенное обсуждение, отражение колебаний содержит также и полезный компонент. 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.

Пример 3.7.1. Обратимся теперь к иллюстрации рассмотренной схемы разнесения в

приложении к цифровой связи, представленной на рис. 3.20. В среде Matlab осуществлено моделирование передачи бинарной ФМ с длительностью бита T по каналу, в котором три пути характеризуются взаимными задержками $\tau_2 - \tau_1 = \tau_3 - \tau_2 \approx 0.15T$. Значения амплитуд многолучевых сигналов составляют $A_1 = 1, A_2 = 0.8, A_3 = 0.9$, а фазы второго и третьего сигналов противоположны фазе первого. Для большей прозрачности на рис. 3.20 показаны только низкочастотные эквиваленты (комплексные огибающие) всех сигналов. Левая колонка (а) отвечает передаче одного бита простым прямоугольным импульсом, причем нулевое значение бита передается импульсом положительной полярности (верхняя диаграмма). Второй график 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 битах вряд ли возможен.

Как пример противоположного плана колонка (b) отвечает методу передачи, при котором значение каждого бита определяет полярность широкополосного сигнала, в частности, кода Баркера длины $N = 13$ (см. параграф 6.4). 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

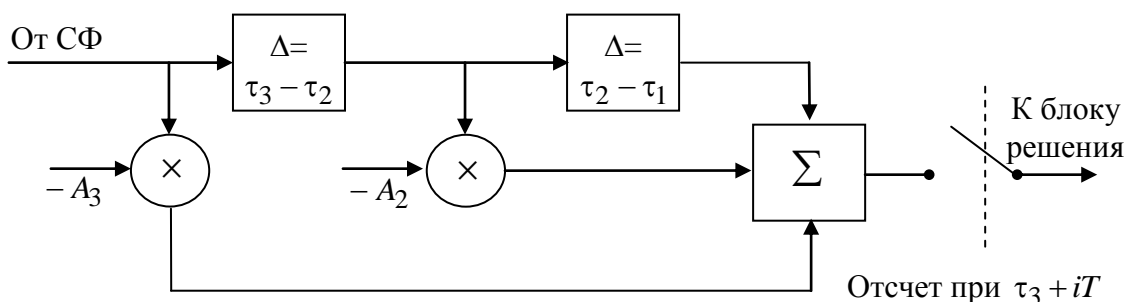


Рис. 3.21. Устройство комбинирования с максимальным отношением.

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. 3.21, 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 пер-

вому пути.

Выигрыш в отношении сигнал-шум наглядно демонстрирует рис. 3.22, на котором верхняя диаграмма повторяет сигнал на выходе согласованного фильтра рис. 3.20, *b*, 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

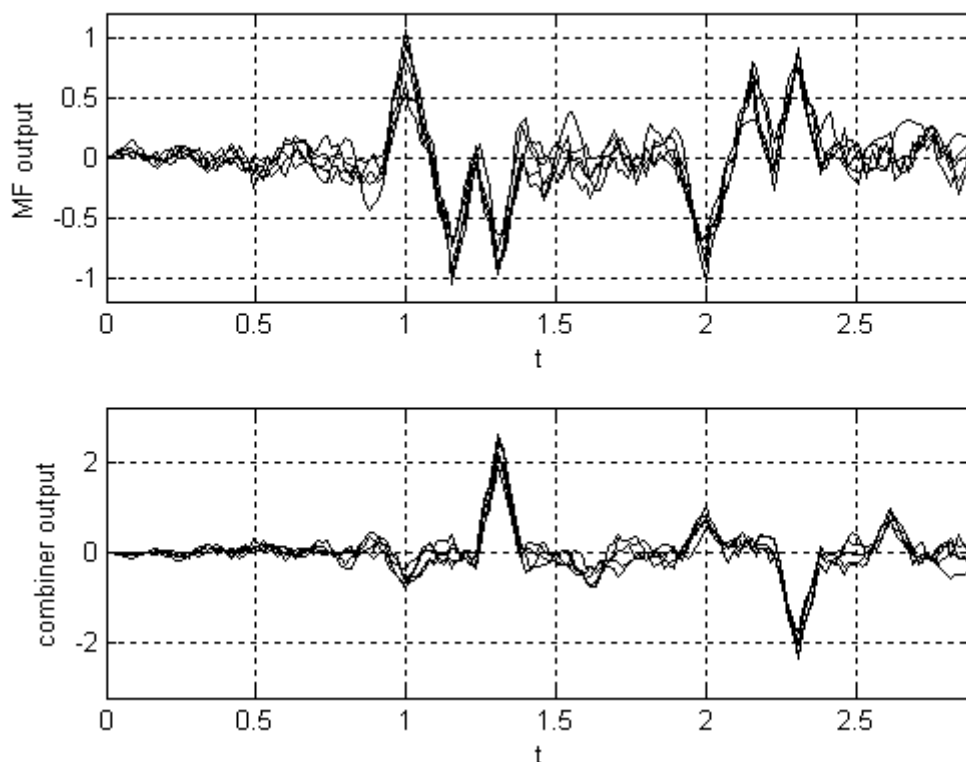


Рис. 3.22. Сигналы на выходе согласованного фильтра и сумматора.

решения о значении переданных бит данных.

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 (см. рис. 3.20 и 3.22) напоминают в некоторой степени зубцы популярного садового инструмента.

Известны многочисленные версии аппаратной реализации RAKE алгоритма. Одна из них предполагает использование n_d параллельных корреляторов вместо согласованного фильтра, где n_d определяется числом зубцов, т.е. числом используемых ветвей разнесения. 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. Преимуществом подобной структуры относительно схемы с согласованным фильтром следует считать то, что для сложного закона модуляции широкополосного сигнала построение коррелятора часто является более осуществимым, чем согласованного фильтра, поскольку, в отличие от последнего, первый

вычисляет только единственное значение корреляции (см. параграф 2.11). При использовании комбинирования с максимальным отношением RAKE приемник на основе корреляторов представим схемой, изображенной на рис. 3.23 (версия для видеосигналов). Для подчеркивания практической роли RAKE разнесения достаточно упомянуть стандарты сотовой связи 2-го (IS-95) и 3-го (UMTS, cdma2000) поколений, использующих этот принцип как составную часть своей идеологии.

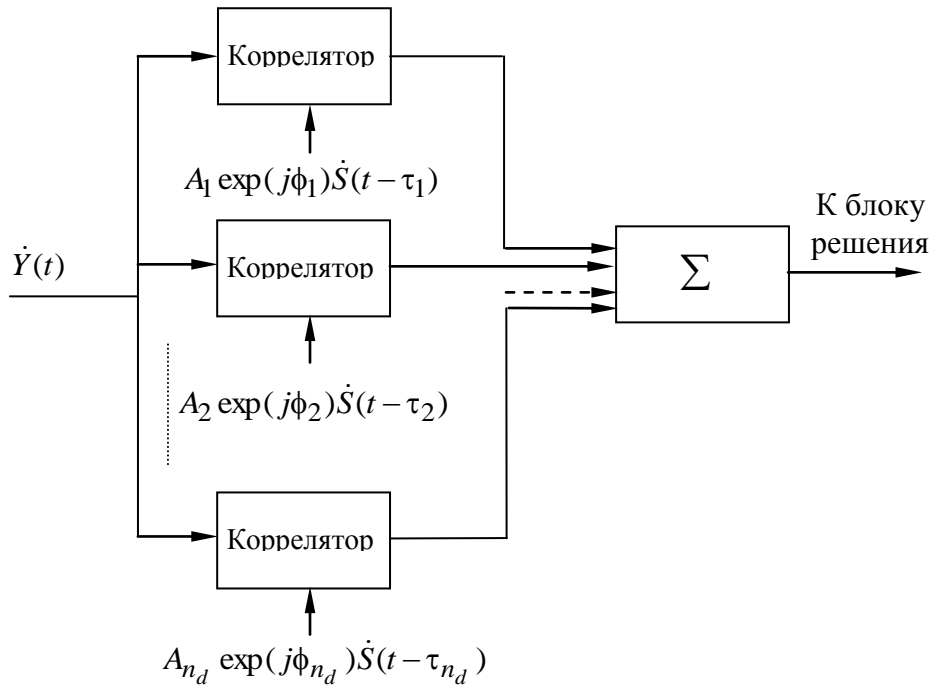


Рис. 3.23. Корреляционный RAKE приемник.

Задачи.

Общие замечания:

- если форма спектра сигнала и/или помехи не определена, то считать ее прямоугольной;
- термин «согласованный фильтр» используется для обозначения фильтра, согласованного с сигналом на фоне АБГШ;
- в случае рассмотрения заградительной помехи считать уровень АБГШ незначительным.

3.1. От системы требуется обеспечить максимально возможное отношение мощности полезного сигнала к общей мощности АБГШ и помехи. На входе присутствует узкополосная помеха. Какая из двух стратегий является лучшей: игнорирование помехи или использование режекторного фильтра, если:

- а). мощность помехи равна мощности АБГШ в пределах полосы сигнала, а ширина полосы помехи составляет половину от ширины полосы сигнала;
- б). мощность помехи на 6 дБ меньше мощности АБГШ в пределах полосы сигнала, а ширина полосы помехи составляет четвертую часть от ширины полосы сигнала;
- в). мощность помехи на 3 дБ больше мощности АБГШ в пределах полосы сигнала, а ширина полосы помехи составляет четвертую часть от ширины полосы сигнала.

3.2. Спектры сигнала и узкополосной помехи представлены на рис. 3.24. Какая центральная частота помехи является наиболее опасной в случае, когда приемник игнорирует помеху и когда использует режекторный фильтр? Почему?

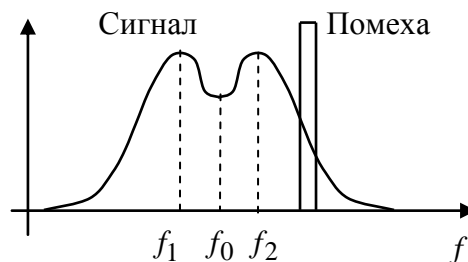


Рис. 3.24. Спектр сигнала и помехи.

3.3. В некоторой системе отношение сигнал–мешающее воздействие по мощности (SIR) на выходе согласованного фильтра проигрывает в 101 раз отношению сигнал–шум по мощности (SNR), тогда как система может сохранить свою работоспособность только при SIR по мощности в два раза меньшем SNR по мощности. Что следует изменить в сигнале при сохранении постоянного значения SNR по мощности, если:

- а). разрешено использовать только простые сигналы?
- б). пиковая мощность сигнала не может возрасти (каким будет выигрыш в обработке сигнала в этом случае)?

3.4. В некоторой системе режекторный фильтр нейтрализует узкополосную помеху. В результате этой операции отношение сигнал–шум (SNR) на выходе согласованного фильтра уменьшается на 3 дБ.

- а). Что следует сделать с длительностью и амплитудой простого сигнала, если допустимым считается уменьшение SNR по мощности только на два процента (в отсутствие режекторного фильтра величина SNR фиксирована)?

б). Возможно ли уменьшить ухудшение ниже 2 процентов без увеличения мощности сигнала? Если да, то каким должен быть выигрыш от обработки сигнала?

3.5. Система может работать с отношением сигнал-шум (SNR) не менее 10 дБ. В результате воздействия заградительной помехи SNR уменьшилось до -3 дБ. Какие параметры сигнала могут быть изменены для нейтрализации действия помехи, если:

- а). Разрешено использовать только простые сигналы с фиксированной энергией?
- б). Могут быть использованы только простые сигналы с одинаковой пиковой мощностью?
- с). Фиксированы только величина пиковой мощности и энергии сигнала?
- д). Фиксирована только пиковая мощность сигнала, а его полоса может быть увеличена только в 10 раз?

Найти выигрыш от обработки сигнала для случаев (с) и (д).

3.6. В рамках борьбы с передатчиком заградительной помехи система увеличила длительность сигнала в 4 раза с одновременным уменьшением мощности сигнала в два раза и расширением полосы в 50 раз. Постановщик помехи способен увеличить свою мощность не более чем на 13 дБ. Какая из сторон окажется победителем в этом противостоянии?

3.7. Сигнал занимает две разнесенные подполосы одинаковой ширины W . Полная энергия сигнала распределена между частотными участками в пропорции 9:16. На приемной стороне два согласованных фильтра осуществляют обработку обеих подполос, а их выходы комбинируются оптимальным образом для максимизации результирующего отношения сигнал-шум. На систему воздействует передатчик заградительной помехи. Какое из распределений его мощности между подполосами сигнала будет наиболее вредным?

3.8. Отношение сигнал-шум на выходе согласованного фильтра скрытного приемника равно 14 дБ. Выигрыш от обработки сигнала составляет $WT = 400$. Найти:

- а). отношение спектральных мощностей сигнала и АБГШ;
- б). отношение сигнал-шум на выходе интегратора радиометра перехватчика.

3.9. В системе передачи данных с бинарной ФМ требуется обеспечить вероятность ошибки на бит $P_e = 1.5 \cdot 10^{-3}$. Разработчик системы рассчитывает, что отношение сигнал-шум для радиометра перехватчика не превысит -10 дБ на передаваемый бит. Какое значение выигрыша от обработки на один бит следует признать удовлетворительным?

3.10. Для радиометра перехватчика отношение сигнал-шум (SNR) составляет -10 дБ на длительность сигнала, тогда как для скрытного приемника SNR должно быть 12 дБ. Длительность сигнала равна $T = 100$ мкс. Каким должна быть его минимальная ширина полосы?

3.11. Сравниваются две системы передачи данных с бинарной ФМ, реализующие технологию распределенного спектра. Первая из них использует бинарную модуляцию для расширения полосы, обеспечивая выигрыш от обработки, равный 100, на один бит данных. Вторая применяет троичную модуляцию, приводящую к выигрышу, равному 50, на бит данных. Какая из систем характеризуется лучшей устойчивостью к раскрытию закона модуляции при условии равенства скоростей передачи данных?

3.12. Сравниваются две широкополосные системы передачи данных с бинарной ФМ, обладающих одинаковой скоростью. В первой из них отношение сигнал-шум (SNR) на бит данных на входе скрытного приемника составляет 12 дБ, тогда как для радиометра

перехватчика SNR равно -12 дБ на бит. Для второй системы аналогичные параметры принимают значения 16 и -4 дБ соответственно. Какая из систем обладает большим иммунитетом к раскрытию закона модуляции?

3.13. Разработчик системы учитывает требования ЭМС при разработке новой системы. Максимальное отношение сигнал-шум, обеспечиваемое новой системой во всей зоне покрытия старыми, составляет 20 дБ. Любая прежняя система работает с удовлетворительным качеством, если дополнительная спектральная плотность мощности не превышает -10 дБ по сравнению со спектром АБГШ. Каким должен быть минимальный выигрыш от обработки у новой системы?

3.14. Имеются две широкополосные системы, занимающие одну и ту же полосу и работающие внутри одной и той же географической области. Максимальное отношение сигнал-шум (SNR) в области пересечения их зон покрытия составляет 20 и 17 дБ соответственно. Для совместимости их работы избыточная спектральная плотность мощности, обусловленная излучением другой системы, должна быть на -7 дБ ниже уровня АБГШ. Определить минимальный выигрыш от обработки для каждой из систем.

3.15. Для нормального функционирования системы, использующей колебания с длиной волны $\lambda_w = 30$ см, необходимое отношение сигнал-шум составляет $q^2 = 14$ дБ. Длительность сигнала $T = 100$ мкс, а шумовая температура приемника $\theta_n = 1000$ °К. Передающая антенна имеет коэффициент усиления, равный 5 дБ, тогда как приемная – всенаправленная. Найти необходимую излучаемую мощность для модели свободного пространства распространения, если зона охвата системой должна иметь радиус не менее 30 км. Насколько увеличиться эта мощность в условиях, типичных для систем мобильной связи, с экспонентой ослабления от расстояния, равной 3.84 , в отсутствие других дополнений к модели свободного пространства?

3.16. Система сохраняет работоспособность, если принятое отношение сигнал-шум по напряжению уменьшается не более чем в 4 раза ниже среднего предсказанного уровня. Определить вероятность выхода системы из строя вследствие крупномасштабного замирания со стандартным отклонением по мощности в 9 дБ.

3.17. Имеются два пути распространения: по прямой линии (LOS) и через отражатель, расположенный в 3 -х км от LOS и равноудаленный от передатчика и приемника. Определить период в метрах и временной интервал между последовательными провалами в мощности в установившемся колебании на входе приемника, расположенном на расстоянии 12 км от передатчика и движущегося с постоянной скоростью 60 км/час, при длине волны 30 см.

3.18. Возможно ли применение бинарной ФМ в канале с быстрыми многолучевыми релейскими замираниями? Какой способ бинарной передачи приемлем в этом случае?

3.19. Передача бинарных данных осуществляется по каналу с логнормальным крупномасштабным и релейским мелкомасштабным замиранием. Вследствие крупномасштабного замирания мощность сигнала в децибелах флюктуирует около значения в 27 дБ со стандартным отклонением в 12 дБ. Достижима ли в этом канале вероятность ошибки на бит, равная 10^{-3} , без помехоустойчивого кодирования?

3.20. Имеются две релейских ветви разнесения с идентичной средней энергией сигнала. Используя в качестве критерия среднее отношение сигнал-шум по мощности на

выходе устройства комбинирования, сравнить по энергетическому выигрышу две схемы комбинирования: с максимальным отношением и выбора ветви с максимальным SNR.

3.21. Сигнал системы занимает полосу в 60 КГц. Распределение задержки в канале составляет 20 мкс. Общая полоса не превышает 300 КГц. Сколько частотных ветвей разнесения может быть организовано при этих данных?

3.22. Имеются четыре пути распространения с длинами в 5 км, 5.4 км, 5.55 км и 6 км. В системе данные передаются со скоростью $R = 20$ кбит/с. Приблизительно оценить полосу сигнала и выигрыш от обработки, необходимые для организации 4-х зубчатого RAKE приемника.

3.23. Минимальная разность в длине путей распространения в канале составляет 300 м. Задержка распространения в канале лежит в пределах 10 мкс. В системе осуществляется передача данных с использованием QPSK со скоростью 20 кбит/с. Какова необходимая полоса и выигрыш от обработки сигнала для организации RAKE приемника с максимально возможным числом зубцов?

3.24. С выхода релеевского канала RAKE приемник разделяет результирующий сигнал на n_d сигналов равной энергии без замираний для применения комбинирования с максимальным отношением. Насколько среднее отношение сигнал-шум по мощности на выходе устройства комбинирования отличается от варианта, когда RAKE алгоритм не применяется? Как объяснить выигрыш в энергии от использования RAKE метода в свете данного ответа?

4. Многопользовательская среда. Множественный доступ с кодовым разделением.

4.1. Многоабонентские системы и проблема множественного доступа.

Большое число современных беспроводных систем относятся к многопользовательским. В *многопользовательской* системе множество линий связи размещаются в пределах общего частотно-временного ресурса, так что любому индивидуальному абоненту разрешалось передавать или принимать свою собственную информацию параллельно с другими пользователями и независимо от них. Показательным примером многопользовательской системы, в которой единственный передатчик осуществляет передачу данных множеству абонентов, служит канал «вниз» спутниковой системы или наземной сотовой системы. Каждый пользовательский приемник в такой системе должен обеспечить выделение информации, адресованной непосредственно данному абоненту, из наблюдаемого сигнала, который содержит данные, направленные множеству пользователей. Другим подобным примером может служить канал «вверх» спутниковой или наземной сотовой системы, в котором осуществляется параллельная работа множества передатчиков, а единственный приемник должен обнаружить данные конкретного абонента в наблюдаемом результирующем сигнале.

При проектировании любой многопользовательской системы основной проблемой является выбор способа *множественного доступа*, т.е. возможности одновременного использования многими абонентами канала связи с минимальным взаимным влиянием. Для математического описания подобной задачи предположим, что данные k -го пользователя образуют последовательность $\mathbf{b}_k = (b_{k,0}, b_{k,1}, \dots)$, где $b_{k,i}$ отвечает i -му символу потока данных k -го пользователя. Данная последовательность тем или иным способом модулирует специфический сигнал k -го пользователя $s_k(t)$, образуя модулированный сигнал $s_k(t; \mathbf{b}_k)$. При распространении по каналу каждый такой сигнал может приобрести амплитуду A_k и временное запаздывание τ_k и после суммирования с сигналами других пользователей образует общий или *групповой* сигнал, достигающий приемника, вида

$$s(t; \mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_K) = \sum_{k=1}^K A_k s_k(t - \tau_k; \mathbf{b}_k),$$

где K – число активных, т.е. реально передающих данные пользователей, а аргументы после точки с запятой в групповом сигнале подчеркивают его зависимость от данных всех активных пользователей. Очевидно, что групповой сигнал сопровождается канальным шумом $n(t)$, так что результирующее наблюдение представимо как

$$y(t) = s(t; \mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_K) + n(t) = \sum_{k=1}^K A_k s_k(t - \tau_k; \mathbf{b}_k) + n(t). \quad (4.1)$$

Приемник должен выделить данные пользователя из наблюдения $y(t)$. Согласно общим положениям, представленным в разделе 2.1, ключевая роль в решении о принимаемых данных $\mathbf{b}_k, k = 1, 2, \dots, K$ в случае гауссовского канала принадлежит евклидову расстоянию (его квадрату) между наблюдением $y(t)$ и различными копиями группового сигнала $s(t; \mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_K)$, соответствующими всем возможным комбинациям данных K пользователей

$$d^2(\mathbf{s}, \mathbf{y}) = \int_0^T [y(t) - s(t; \mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_K)]^2 dt. \quad (4.2)$$

Подстановка (4.1) в (4.2) и раскрытие скобок приводит к соотношению

$$d^2(\mathbf{s}, \mathbf{y}) = \|\mathbf{y}\|^2 - 2 \sum_{k=1}^K A_k z_k(\mathbf{b}_k) + \sum_{k=1}^K \sum_{l=1}^K A_k A_l \int_0^T s_k(t - \tau_k; \mathbf{b}_k) s_l(t - \tau_l; \mathbf{b}_l) dt, \quad (4.3)$$

где $z_k(\mathbf{b}_k)$ – корреляция (скалярное произведение) наблюдения $y(t)$ и задержанного на τ_k k -го пользовательского сигнала, промодулированного последовательностью данных \mathbf{b}_k

$$z_k(\mathbf{b}_k) = \int_0^T y(t) s_k(t - \tau_k; \mathbf{b}_k) dt. \quad (4.4)$$

Как правило, принятию решения о последовательностях данных предшествует операция оценивания интенсивности сигналов и их задержек, так что параметры $A_k, \tau_k, k = 1, 2, \dots, K$ в соотношениях (4.3)-(4.4) можно полагать известными с достаточной точностью. Тогда оптимальная (максимально правдоподобная или по минимуму расстояния) стратегия восстановления пользовательских данных состоит в подстановке всех возможных реализаций последовательностей $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_K$ в (4.3) и выборе тех из них, которые совместно минимизируют квадрат расстояния (4.3)¹. Данное решающее правило, называемое *многопользовательским алгоритмом*, может оказаться совершенно не практичным в ситуации, когда число пользователей K измеряется десятками или более. В качестве примера рассмотрим простейший вариант синхронной системы с нулевыми взаимными задержками $\tau_k = 0, k = 1, 2, \dots, K$ и передачей бинарных данных. При интервале наблюдения, отвечающем длительности бита, восстановление собственных бит $K = 40$ пользователей потребовало бы проверки $2^{40} > 10^{12}$ битовых образцов всех пользователей, что представляется абсолютно нереальным с реализационной точки зрения. К задаче многопользовательского обнаружения вновь обратимся в разделах 7 и 10.

Так называемый *обычный* или *однопользовательский* приемник осуществляет альтернативное решающее правило, заключающееся в отдельной оценке каждой из последовательностей данных \mathbf{b}_k на основании максимизации корреляции (4.4). Очевидно, что данная стратегия совпадает с оптимальной (многопользовательской), если и только если третье слагаемое в (4.3) полностью не зависит от последовательностей данных $\mathbf{b}_k, k = 1, 2, \dots, K$. Выполнение последнего условия возможно при использовании схемы модуляции, обладающей следующими свойствами:

- а) энергия пользовательских сигналов не зависит от передаваемых данных (PSK, FSK);
- б) все сигналы пользователей ортогональны вне зависимости от передаваемых данных.

Оба эти требования формализуются соотношением

$$\int_0^T s_k(t - \tau_k; \mathbf{b}_k) s_l(t - \tau_l; \mathbf{b}_l) dt = E \delta_{kl}. \quad (4.5)$$

Называя этот метод множественного доступа ортогональным и возвращаясь к материалу, изложенному в разделах 2.3-2.4, вспомним, что максимальное число ортогональных сигналов ограничено общей размерностью сигнального пространства и при общей полосе W_t и временном ресурсе T_t может составлять не более $2W_t T_t$ ортогональных полосных сигналов. Для определения максимально достижимого числа пользователей в ор-

¹ Это правило остается оптимальным, даже если приемник предназначен для восстановления только определенной (k -й) информационной последовательности данных, что, например, имеет место в канале «вниз» мобильной сотовой системы. После оценивания данных всех пользователей приемник просто отбрасывает как ненужные данные всех пользователей кроме k -го.

тогональной схеме множественного доступа ограничимся M -ичной цифровой передачей данных с фиксированной скоростью R бит/сек. Учитывая, что все пользовательские сигналы должны быть ортогональны на интервале времени, равном длительности M -ичного символа, приходим к соотношению $T_t = (\log_2 M)/R$. Следовательно, максимальная размерность сигнального пространства составляет $2W_t T_t = (2W_t \log_2 M)/R$. Если $M = 2$ (вариант BPSK), каждый пользователь занимает только одномерное подпространство сигнального пространства, поскольку для передачи одного бита необходимы только два противоположных импульса (т.е. два коллинеарных вектора) (см. рис. 2.5, а). В этом случае максимальное число пользователей совпадает с общей размерностью сигнального пространства. При $M > 2$ каждому пользователю требуется двумерное подпространство (т.е. плоскость, см. рис. 2.6, с), и, согласно (4.5), все подобные плоскости должны быть ортогональны, так что максимальное число пользователей становится в два раза меньше полной размерности сигнального пространства. Сочетание двух этих результатов дает верхнюю границу максимального числа пользователей в ортогональной схеме множественного доступа

$$K = \begin{cases} \frac{2W_t}{R}, & M = 2, \\ \frac{W_t \log_2 M}{R}, & M > 2 \end{cases} \quad (4.6)$$

В следующих трех параграфах будут кратко обсуждаться традиционные пути реализации ортогонального множественного доступа.

4.2. Множественный доступ с частотным разделением.

Одним из наиболее простых вариантов выполнения требования (4.5) служит использование сигналов пользователей, спектры которых не перекрываются. Данная идея полностью родственна проблеме ортогонального кодирования частотным сдвигом, рассмотренной в 2.7.2. Этот метод множественного доступа, называемый *множественным*

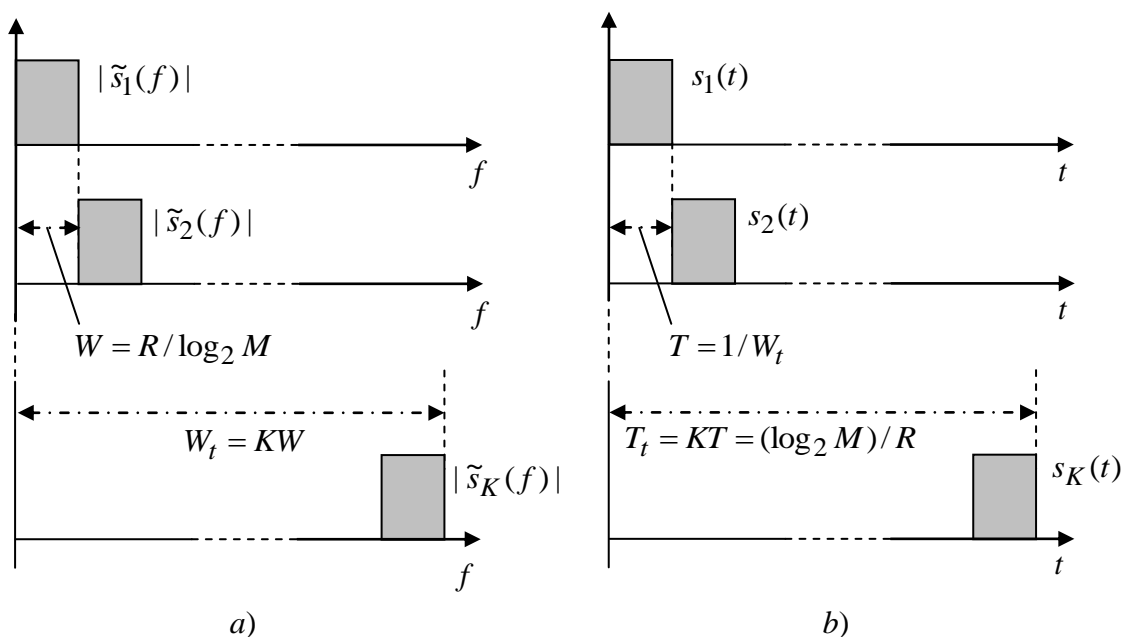


Рис. 4.1 Множественный доступ с частотным (а) и временным (б) разделением

доступом с частотным разделением (МДЧР) (*frequency division multiple access (FDMA)*),

поясняется рис. 4.1, *a*. Если для передачи данных со скоростью R используется M -ичная ФМ, тогда длительность символа данных будет $T_t = (\log_2 M)/R$, так что каждому пользовательскому сигналу должна отводиться полоса, не меньшая чем $W = 1/T_t = R/\log_2 M$. Тогда в полной выделенной полосе W_t могут располагаться не более чем $W_t/W = (W_t \log_2 M)/R$ не перекрывающихся спектров. Если $M > 2$, то данное соотношение в точности определяет максимальное число пользователей, что иллюстрирует рис 4.1, *a*. Если же $M = 2$ и обеспечивается фазовая когерентность, каждая из упомянутых полос может быть использована двумя потребителями, несущие частоты которых отличаются только квадратурным фазовым сдвигом. В результате потенциальное число пользователей в схеме FDMA устанавливается границей (4.6). На практике неидеальность устройств фильтрации, дрейф опорного генератора и доплеровский сдвиг частот могут привести к частичному перекрытию соседних спектров, т.е. возникновению взаимной помехи между различными пользовательскими сигналами. Необходимость нейтрализации этих негативных эффектов и сохранения разделения сигналов абонентов вынуждает введение защитных интервалов между соседними спектрами, что снижает достижимое число пользователей в сравнении с граничным (4.6).

Метод FDMA является старейшим и классическим способом множественного доступа, используемым как в аналоговых, так и цифровых беспроводных системах (радио- и телевидение, мобильная связь и др.). Неперекрывание спектров сигналов гарантирует их ортогональность и, следовательно, возможность разделения сигналов пользователей вне зависимости не только от данных, но и временного запаздывания, благодаря чему отсутствует необходимость в синхронизации сигналов. Последнее часто рассматривается как серьезное достоинство FDMA (более детально этот факт будет рассмотрен в 4.5).

4.3. Множественный доступ с временным разделением.

Другой популярной схемой ортогонального множественного доступа является *множественный доступ с временным разделением* (МДВР) (*time division multiple access* (TDMA)), в котором пользовательские сигналы не перекрываются во временной области (рис. 4.1, *b*). Данная идея снова заимствована у способа ортогонального кодирования временным сдвигом (см. 2.7.1). В случае M -ичной ФМ данное условие означает, что полный выделенный временной ресурс $T_t = (\log_2 M)/R$ (в системах с TDMA его часто называют *кадром* (*frame*)) делится на не перекрывающиеся слоты длительностью T . Если $M > 2$ (данный вариант представлен на рис. 4.1, *b*), то каждый слот может быть использован только одним пользователем, и длительность передаваемого им символа данных не может быть меньше величины, обратной полосе $1/W_t$. Поэтому общее число пользователей ограничено сверху величиной $T_t/T = (W_t \log_2 M)/R$. Если же $M = 2$, то при фазовой когерентности два пользователя могут использовать один и тот же слот, применяя сдвинутые в квадратуре несущие. В итоге вновь приходим к границе (4.6) для максимального значения K , что демонстрирует теоретическую эквивалентность FDMA и TDMA относительно потенциального числа абонентов.

Метод TDMA находит применение в различных системах, например, в системах мобильной связи второго поколения (GSM, IS-136 и др.). Не смотря на кажущуюся привлекательной простоту, следует упомянуть и о недостатках этого метода. Во-первых, любой пользовательский сигнал занимает только K -ю (или возможно $K/2$ -ю) часть кадра, что влечет за собой увеличение пиковой мощности в K (или $K/2$) раз по сравнению со случаем непрерывной передачи, что вызвано необходимостью сохранения неизменной энергии сигнала, т.е. отношения сигнал-шум. Реализационные проблемы, обусловленные этим требованием, уже не раз упоминались ранее. Во-вторых, необходима жесткая синхронизация между сигналами пользователей на входе приемника, поскольку в противном

случае они будут перекрываться, создавая взаимные помехи. В то же время в системах с подвижными пользователями, таких как канал «вверх» в мобильном телефоне, длины путей распространения между передатчиками абонентов и приемником базовой станции непрерывно изменяются, причем в достаточно большом диапазоне. Очевидно, что в подобной ситуации синхронизация пользовательских сигналов на входе приемника, хотя в принципе и возможна, однако может вызвать значительные технологические трудности. Обычный путь преодоления указанных препятствий состоит во введении защитных временных интервалов между сигналами соседних пользователей, предохраняя тем самым от возможного их перекрытия во всем диапазоне возможных задержек. Как правило, величина защитных интервалов оказывается значительной, что может в сильной степени сократить число пользователей в сравнении с границей (4.6). Острота проблемы несколько снижается, в случае, когда индивидуальный пользовательский слот содержит не единственный символ данных (т.е. бит), а пакет из n_b символов. Тогда применение защитных интервалов вызывается необходимостью разделения пакетов различных пользователей, что приводит к использованию в n_b раз меньшего защитного интервала заголовка. С другой стороны, паузы между последовательными пакетами одного и того же пользователя также увеличиваются в n_b раз. Во многих системах (подобных системам мобильного телефона, где необходимо поддерживать непрерывность обмена речевой информацией) длинные паузы не приемлемы, что обуславливает введение жесткого предела на длину пакетов.

В свете перечисленных причин «чистый» TDMA метод не находит широкого применения в современной практике. Например, в системах мобильной радиосвязи второго поколения используется комбинирование TDMA и FDMA.

4.4. Синхронный вариант множественного доступа с кодовым разделением.

Как метод FDMA, так и TDMA предполагают распределение общего доступного частотно-временного ресурса между различными пользователями, так что каждый из них

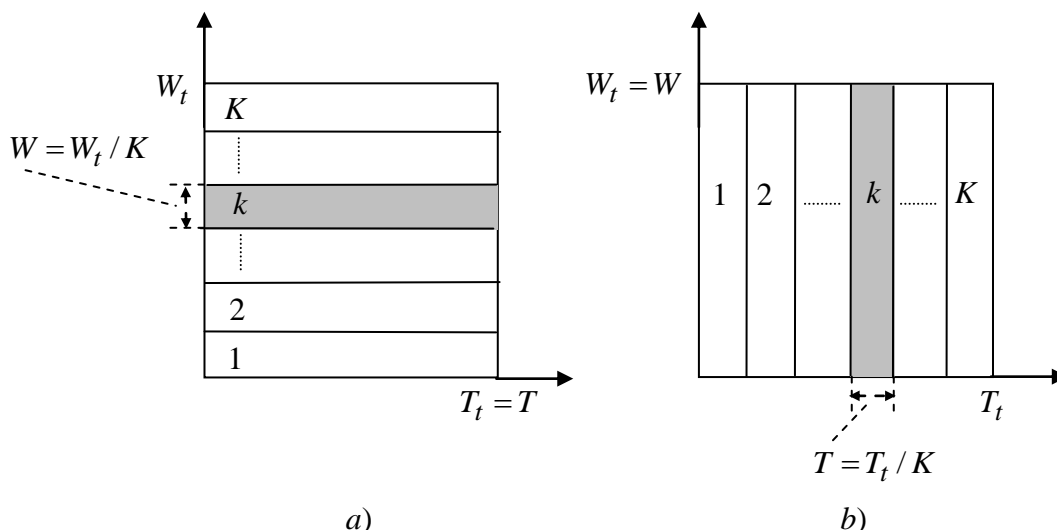


Рис.4.2 Распределение ресурса при ортогональном временном (а) и частотном (б) копировании

использует только свою собственную, специфическую для пользователя часть, и отсутствуют абоненты, совместно использующие некоторые части ресурса. При FDMA подобная фрагментация осуществляется в частотной области (рис. 4.2, а). Так, в распоряжении k –

го пользователя находится весь временной ресурс ($T = T_t$) и только часть частотного ресурса. Когда достижение максимального числа абонентов является приоритетной задачей, то $W = 1/T \approx W_t / K$. Деление на части временной области при TDMA (рис. 4.2, б) предоставляет возможность одиночному пользователю занимать весь доступный частотный ресурс ($W = W_t$), но только часть полного временного кадра ($T = 1/W \approx T_t / K$). Если необходимо максимизировать число пользователей, то в обоих этих ортогональных схемах множественного доступа фрагментация ресурса заставляет использовать каждого абонента простой сигнал, поскольку для последнего частотно-временное произведение $WT = 1$.

С другой стороны, при необходимости обслуживания большого числа пользователей K общий частотно-временной ресурс также должен быть значительным ($W_t T_t \gg 1$, см. (4.6)), и если каждый пользовательский сигнал занимает как всю доступную полосу ($W = W_t$), так и временной интервал ($T = T_t$), то приходим к необходимости применения ортогональной схемы множественного доступа, в которой все пользовательские сигналы широкополосны. В этом случае многопользовательская система будет обладать всеми достоинствами технологии распределенного спектра, рассмотренными в предыдущем разделе.

Предположим, что передача сигналов может быть организована таким образом, что на входе приемного устройства будут отсутствовать взаимные временные задержки между сигналами пользователей. Тогда, не нарушая общности, все абсолютные задержки сигналов могут быть положены равными нулю: $\tau_k = 0, k = 1, 2, \dots, K$. Возьмем произвольное семейство $W_t T_t$ ортогональных широкополосных сигналов (см. параграф 2.7.3), например, функции Уолша, и используем каждый из них как пользовательский сигнал для M -ичной ФМ передачи данных. Индивидуальный широкополосный сигнал, назначенный k -му пользователю, называется k -й *сигнатурой*. Каждая сигнатура занимает всю полосу W_t и весь временной кадр T_t (см. рис. 4.3), передавая $\log_2 M$ бит данных за интервал T_t . Если $M > 2$, то данный метод множественного доступа может обслуживать до $K = WT = W_t T_t = (W_t \log_2 M) / R$ абонентов, тогда как применение бинарной ФМ допускает удвоение числа K , позволяя двум различным пользователям использовать сдвинутые по фазе квадратурные копии одной и той же сигнатуры. Очевидно, что снова максимальное число пользователей в точности определяется соотношением (4.6), как это уже было при FDMA и TDMA.

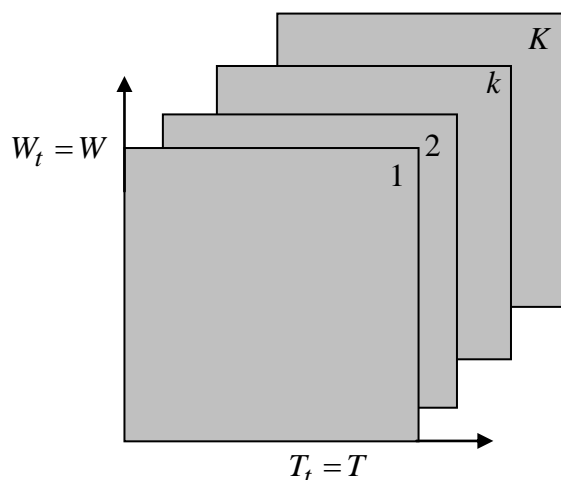


Рис.4.3. Утилизация ресурса в широкополосном методе множественного доступа.

В рассмотренном методе множественного доступа соответствующее кодирование

сигнатур обеспечивает ортогональность сигналов пользователей вместо фрагментации временной или частотной области. Именно поэтому данному методу присвоили название *множественного доступа с кодовым разделением* (МДКР) (*code division multiple access* (CDMA)). Достоинства CDMA в сравнении с классическими методами FDMA и TDMA (помехоустойчивость, низкая вероятность обнаружения, возможность привлечения RAKE алгоритма др.) автоматически следуют из широкополосной природы CDMA сигнатур. В то же время, синхронизация сигнатур является необходимым условием для обеспечения их ортогональности и разделения абонентов на приемной стороне. С целью отличия данного варианта CDMA от рассматриваемого позднее в его отношении используется термин синхронного CDMA (S-CDMA). Синхронный метод достаточно легко реализуется в системах с единственным передатчиком (подобно базовой станции в сотовой сети), одновременно излучающем потоки индивидуальных данных, каждый из которых адресован определенному пользователю (например, мобильной станции). Последнее объясняет использование S-CDMA как основы физического слоя линии «вниз» в сотовых сетях с CDMA 2-го (IS-95) и 3-го (UMTS, cdma2000) поколений. Параллельно основная идея S-CDMA используется в каналах «вниз» и «вверх» стандартах 3-го поколения для организации т. н. многокодовой передачи (см. 11.3).

4.5. Асинхронный метод с CDMA.

Во многих приложениях типичной является ситуация, когда задержка τ_k может изменяться в широком диапазоне, делая процедуру синхронизации сигнатур на входе приемника проблематичной или вообще невозможной. Показательным примером этому может служить канал «вверх» системы мобильной сотовой связи, в которой потребители перемещаются внутри соты, вследствие чего происходит постоянное изменение расстояния между ними и базовой станцией, а значит, и время поступления пользовательских сигналов на приемник базовой станции. В принципе, каждый потребитель, обладая сведениями о своем мгновенном местоположении относительно базовой станции, а значит, и задержки на распространение τ_k , имеет возможность передавать свой сигнал с коррекцией на τ_k . В результате все задержки на распространение могут быть скомпенсированы, и все пользовательские сигналы на входе приемника базовой станции будут синхронизированы. Однако подобный алгоритм работы приведет к чрезмерному усложнению оборудования, которое во многих случаях окажется экономически не обоснованным.

Проанализируем последствия асинхронного характера принимаемых сигналов пользователей. Во-первых, возможно ли сохранить ортогональность сигналов при взаимных временных сдвигах, изменяющихся в широком диапазоне? Рассмотрим два сигнала $u(t)$ и $v(t)$, и вычислим их взаимную корреляционную функцию (ВКФ) $R_{uv}(\tau)$, т.е. скалярное произведение $u(t)$ со сдвинутой во времени копией $v(t)$ как функцию аргумента τ :

$$R_{uv}(\tau) = \int_{-\infty}^{\infty} u(t)v(t-\tau)dt .$$

Применение теоремы Парсеваля дает

$$R_{uv}(\tau) = \int_{-\infty}^{\infty} \tilde{u}(f)\tilde{v}^*(f)\exp(-j2\pi f\tau)df .$$

Если целью является достижение ортогональности сигналов вне зависимости от взаимной задержки τ , то при всех значениях τ должно выполняться равенство $R_{uv}(\tau) = 0$, что, вследствие линейности преобразования Фурье, возможно, если и только если $\tilde{u}(f)\tilde{v}(f) = 0$ во всей частотной области. Последнее утверждение свидетельствует о том, что два сигнала ортогональны при любом произвольном временном сдвиге, если и

только если их спектры не перекрываются. Однако схема множественного доступа с не перекрывающимися спектрами есть FDMA! Следовательно, асинхронный вариант ортогонального множественного доступа реализуем только в виде FDMA, что достаточно часто провозглашается как одно из основных достоинств FDMA.

Однако какова плата за попытку реализации CDMA при отсутствии синхронизации сигнатур на входе приемника? Поскольку сигнатуры различных абонентов в рамках CDMA обладают перекрывающимися спектрами, то они не могут оставаться ортогональными в широком диапазоне взаимных задержек, и тогда равенство (4.5) не является справедливым при произвольных значениях τ_k, τ_l и последовательностей данных $\mathbf{b}_k, \mathbf{b}_l$. Следствием этого является возникновение межпользовательского мешающего воздействия, проявлением которого служит ненулевой отклик приемника, настроенного на k -го пользователя, от сигналов других абонентов.

Рассмотрим обычный приемник k -го пользователя. Без потери общности можно считать, что $\tau_k = 0$, переписав (4.4) как

$$z_k(\mathbf{b}_k) = \int_0^T y(t) s_k(t; \mathbf{b}_k) dt. \quad (4.7)$$

Согласно правилу одиночного наблюдателя оценка $\hat{\mathbf{b}}_k$ данных \mathbf{b}_k должна максимизировать решающую статистику $z_k(\mathbf{b}_k)$ как функцию \mathbf{b}_k . Подставив (4.1) в (4.7), выразим $z_k(\mathbf{b}_k)$ в виде

$$z_k(\mathbf{b}_k) = A_k \int_0^T s_k(t; \mathbf{b}'_k) s_k(t; \mathbf{b}_k) dt + \sum_{\substack{l=1 \\ l \neq k}}^K A_l \int_0^T s_l(t - \tau_l; \mathbf{b}'_l) s_k(t; \mathbf{b}_k) dt + \int_0^T n(t) s_k(t; \mathbf{b}_k) dt, \quad (4.8)$$

где осуществлена замена \mathbf{b}_k в (4.1) на \mathbf{b}'_k с тем, чтобы отличить подлинно передаваемые данные \mathbf{b}'_k от предполагаемых в процессе принятия решения относительно \mathbf{b}_k .

Первый и последний члены в (4.8) отвечают вкладу собственного, т.е. k -го сигнала пользователя и теплового аддитивного шума в отклик приемника k -го пользователя. В случае отсутствия сторонних абонентов ($K = 1$) второе слагаемое равнялось бы нулю, и не было бы никакого отличия от задачи, рассмотренной в главе 2. При $K > 1$ и произвольных запаздываниях сигнатур указанное слагаемое отлично от нуля, определяя вклад других пользовательских сигналов в выходной эффект k -го приемника, т.е. взаимную помеху или *помеху множественного доступа (multiple-access interference (MAI))*.

Простейшим путем оценки влияния MAI является представление всех чужих сигналов в виде случайных шумоподобных процессов, как это уже использовалось в главе 3. В любой реальной асинхронной CDMA системе должны быть предприняты меры по выравниванию уровней всех пользовательских сигналов на входе приемника с тем, чтобы избежать проблемы *близости–дальности (near-far problem)*. Последняя подразумевает, что MAI, образованная чужими пользователями, которые располагаются гораздо ближе к приемнику, чем собственный (k -й) абонент, может значительно превзойти по мощности полезный сигнал вследствие жесткой зависимости принимаемой мощности от расстояния (см. параграф 3.5.2). Поэтому можно полагать, что в результате эффективной *регуляции мощностью (power control)* все $A_k, k = 1, 2, \dots, K$, одинаковы или, другими словами, мощности всех сигналов идентичны и равны P . Тогда шумоподобный сигнал l -го стороннего пользователя, энергия которого полагается равномерно распределенной в полосе W , создаст избыточную спектральную плотность мощности шума $N_l = P/W$, которая добавится к спектру теплового шума. Поскольку наличествуют $K - 1$ независимых сторонних пользователей, то полная спектральная плотность MAI составит $N_I = (K - 1)N_l = (K - 1)P/W$.

Теперь можно записать отношение сигнал-помеха по мощности (SIR) q_I^2 , учитывающее

как MAI, так и тепловой шум

$$q_I^2 = \frac{2E}{N_0 + N_I} = \frac{2E}{N_0 + (K-1)\frac{P}{W}}. \quad (4.9)$$

Как правило, число пользователей K и/или выигрыш от обработки каждой сигнатуры является достаточно значительным для того, чтобы привлечь механизм центральной предельной теоремы и трактовать второе слагаемое в (4.8) как случайную гауссовскую величину. Такой подход оправдывает широко применяемую гауссовскую аппроксимацию помехи множественного доступа и означает, что все результаты, полученные в главе 2 для классических задач приема (вероятность ошибки, точность оценивания и др.), применимы для аналогичных задач со многими пользователями после замены отношения сигнал–шум (SNR) q^2 на отношение сигнал-помеха (SIR) q_I^2 . Например, если передача данных осуществляется с помощью BPSK, то для любого абонента вероятность ошибки на бит вычисляется с помощью соотношения (2.19), в котором q_I^2 заменяет $q^2 = 2E/N_0$.

Соотношение (4.9) дает возможность оценить максимальное число пользователей, которое может обслужить асинхронная CDMA система при заданном частотно-временном ресурсе WT . Не трудно заметить, что при многопользовательской среде отсутствие теплового шума не приведет к безошибочным решениям на приемной стороне, поскольку наличие MAI сохранит SIR конечным и равным т.н. *пороговому* (*floor*) отношению сигнал-помеха

$$q_{If}^2 = \frac{2E}{(K-1)\frac{P}{W}} = \frac{2PT}{(K-1)\frac{P}{W}} = \frac{2WT}{K-1}. \quad (4.10)$$

Последний результат показывает, что пороговое отношение SIR и, следовательно, пороговая достоверность приема полностью определяется частотно-временным произведением, т.е. выигрышем от широкополосной обработки WT , и числом пользователей. Учитывая справедливость неравенства $q_I^2 \leq q_{If}^2$, максимально возможное число пользователей может быть оценено сверху соотношением

$$K \leq \frac{2WT}{q_I^2} + 1, \quad (4.11)$$

где q_I^2 отвечает необходимому отношению SIR, гарантирующему требуемую достоверность приема в анализируемой системе. Для конкретности рассмотрим систему передачи данных с использованием BPSK или QPSK, в которой требуется обеспечить вероятность ошибки на бит не хуже, чем $P_e = 10^{-2}$. Из соотношения (2.19) или рис. 3.16 (пунктирная линия) можно увидеть, что в отсутствии замираний отношение SIR, равное 7 дБ ($q_I^2 = 5$), является достаточным для удовлетворения этого требования. Данная величина позволяет оценить потенциальное число пользователей как

$$K \leq \frac{2WT}{5} + 1. \quad (4.12)$$

В то же время, система с FDMA способна обслужить WT пользователей¹ в рамках такого же частотно-временного ресурса ($W_t = W, T_t = T$), что примерно в 2,5 раза больше, чем правая часть (4.12). Данный факт навеивает более чем скромные мысли о перспективах асинхронного варианта CDMA в сравнении с FDMA. Однако в следующем параграфе бу-

¹ Здесь не учитывается потенциальная возможность удвоения числа пользователей за счет использования той же самой несущей частоты с квадратурным сдвигом по фазе, поскольку этот вариант не осуществим вследствие отсутствия синхронизации различных абонентов.

дет продемонстрировано, что в системах, допускающих повторное использование частотного ресурса в пространственно удаленных зонах (например, сотовых системах), метод асинхронного CDMA значительно превосходит FDMA в значении максимального числа обслуживаемых потребителей.

4.6. Асинхронный вариант CDMA в сотовых сетях.

4.6.1. Проблема повторного использования ресурса и сотовые системы.

При создании новой коммерческой многопользовательской беспроводной системы ее проектировщик естественно заинтересован, чтобы она обслуживала по возможности максимальное число абонентов, находясь в то же самое время в рамках жестких фундаментальных ограничений. Первым из них является ограниченная мощность, лимитирующая размеры зоны, покрываемой одиночным передатчиком. Кривизна земной поверхности и быстрое ослабление интенсивности сигнала с расстоянием, что характерно для СВЧ диапазона, используемого системами подобными мобильному телефону (см. раздел 3.3), выводят из практического использования зоны покрытия, радиус которых превышает десятки километров. Еще одни жесткие рамки накладывает частотно-временной ресурс, т.е. полоса, занимаемая спектром, и требуемая скорость передачи данных. Так, например, полоса физического канала системы cdmaOne (IS-95) составляет $W_c = 1.25$ МГц. При скорости передачи кодированных речевых данных $R = 19.2$ бит/сек и BPSK модуляции, используемой в канале «вниз», потенциальное число активных пользователей, согласно (4.6), составляет $K = 130$. Ясно, что это число чрезвычайно мало как для покрытия густонаселенной городской зоны, так и становится еще более очевидным, если наряду с соединением телефона требуется обслуживание высокоскоростной (например, мультимедийной) передачи данных.

Эффективный путь преодоления указанных препятствий предлагается топологией сотовых сетей, включающей использование множества базовых станций, каждая из которых обслуживает свою индивидуальную зону (соту) и которые совместно покрывают всю необходимую область. Передатчик базовой станции (БС) с относительно низкой мощностью излучения сигналов пользователям или мобильным станциям (МС), которые расположены внутри обслуживаемой соты, и приемники МС образуют канал «вниз». Канал «вверх» включает передатчики МС и приемник БС. Все БС работают строго координированно, и вся сеть имеет соединение с сетями фиксированной телефонной связи и передачи данных. В том случае, когда МС, передвигаясь по зоне покрытия системы, выходит за пределы текущей соты, обслуживание данной МС автоматически переходит БС соседней соты. Процедура указанного перехода получила название *эстафетной передачи (handover)*. В рамках сотовой философии ослабление колебаний проявляет свои положительные черты, позволяя повторное употребление одних и тех же физических подканалов (например, частотных подполос при FDMA или временных слотов при TDMA) различными передатчиками при условии их взаимного разнесения на расстояние, обеспечивающее снижение интенсивности сигналов одних из них до пренебрежимо малого уровня во всей зоне покрытия других. Вследствие упомянутых доводов одним только увеличением числа ячеек можно гибко решать задачу увеличения числа пользователей и расширения зоны охвата. В малонаселенных областях макросоты (измеряемые от единиц до десятков километров) могут удовлетворить требованиям покрытия, тогда как в густонаселенных зонах может возникнуть потребность в использовании микро– (сотни метров) и даже пикосот (десятки метров). Повсеместно принято аппроксимировать одиночную ячейку сотовой сети правильным шестиугольником, так что модель сети напоминает медовые соты (см. рис. 4.4).

Оценим эффективность использования частотно–временного ресурса в сотовой системе, использующей классические схемы множественного доступа FDMA и TDMA. Для избежания лишних повторений и учитывая эквивалентность FDMA и TDMA по числу пользователей (см. 4.3) будем использовать терминологию, относящуюся только к FDMA. Очевидно, что радиус соты не может быть больше радиуса полного затухания колебания, но последний, как уже указывалось, должен быть, по крайней мере, в два раза меньше, чем расстояние между центрами сот, использующими сигналы пользователей с идентич-

ными частотами. Если первое условие не выполняется, то МС, находящаяся у границы соты, будет принимать слишком слабый сигнал БС, что не позволит установить с ней контакт. Нарушение второго условия повлечет возникновение межсотовых помех, поскольку

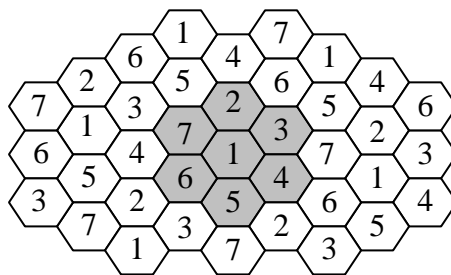


Рис. 4.4. Конфигурация сотовой сети.

опять МС, передвигающаяся вдоль границы ячейки, может принимать не только сигнал, соответствующий собственной БС, но и сигнал чужой БС, связанной с другой МС (обслуживаемой этой посторонней БС) на такой же частоте. Другими словами, множество частот всех сот вокруг любой конкретной ячейки должно отличаться от множества, используемого центральной сотой. Таким образом, возникает конфигурация, называемая *кластером*, в пределах которой отсутствует повторное использование множества частот. Регулярная структура сот, в которой распределение частот между ячейками удовлетворяет вышеприведенному условию, может существовать только для некоторых определенных размеров кластера. Наиболее типичным является 7–сотовый кластер, выделенный оттенком на рис. 4.4. Следовательно, только одна седьмая часть общего числа физических каналов (частот), допускаемых полным частотно–временным ресурсом $W_f T_t$ системы, может быть использована одиночной ячейкой. Последнее утверждение дает следующую оценку максимального числа пользователей, приходящихся на одну соту, в системе с FDMA или TDMA

$$K_c = \frac{W_f T_t}{7}, \quad (4.13)$$

где предполагается асинхронный характер работы, типичный для канала «вверх». В свете последнего результата пессимистическое заключение о перспективах асинхронного CDMA, сделанное в предыдущем параграфе, нуждается в серьезной ревизии.

4.6.2. Число пользователей на соту в асинхронном варианте CDMA.

Напомним, что асинхронный вариант CDMA системы основан на принципе распределенного спектра, и каждая сигнатура занимает весь доступный частотно–временной ресурс. Рассмотрим канал «вверх» CDMA сотовой системы, в которой все ячейки используют одну и ту же частотную полосу без распределения спектрального ресурса между ними. Другими словами, сигнатуры всех сот, включая и соседние, занимают одну и ту же спектральную полосу, а кластер состоит только из единственной ячейки. Ясно, что приемник БС некоторой конкретной соты будет принимать помехи множественного доступа (МА) не только от абонентов своей соты, но и от МС, обслуживаемых сторонними базовыми станциями. Возникает естественный вопрос: насколько велик вклад в общий уровень МА компонентов, обусловленных передатчиками МС внешних сот? Для оценки интенсивности этих *межсотовых* МА обратимся к рис. 4.5, на котором две примыкающие соты C_1, C_2 аппроксимированы окружностями радиуса D_c . Имеются две БС, обозначенные как БС1 и БС2, и МС, расположенная в зоне покрытия БС2. Несмотря на то, что МС обслуживается БС2 ее сигнал также попадает на вход приемника БС1, внося вклад в меж-

отовые помехи. Обозначим расстояние от МС до БС1 и БС2 через D_1 и D_2 соответственно и напомним, что точная регулировка мощности жизненно важна для любой асинхронной CDMA системы, чтобы избежать проблемы близости–дальности. Благодаря петли регулировки мощность сигнала, принимаемого БС2 от МС, все время поддерживается постоянной и равной P . Если мощность излучения МС составляет P_t , то, согласно модели распространения, описанной в параграфе 3.5, $P = kP_t / D_2^e$. С другой стороны, сигнал, распространяющийся от МС до БС1, претерпевает ослабление, величина которого определяется расстоянием D_1 , так что мощность, принимаемая БС1, будет $P_{r1} = kP_t / D_1^e$. Используя предыдущее соотношение, выразим величину P_{r1} через мощность P полезного сигнала на входе приемника «своей» БС как

$$P_{r1}(D_2, \theta) = \left(\frac{D_2}{D_1} \right)^e P, \quad (4.14)$$

где координаты МС D_2, θ (см. рис. 4.5) подчеркивают зависимость P_{r1} от положения МС внутри соты C_2 .

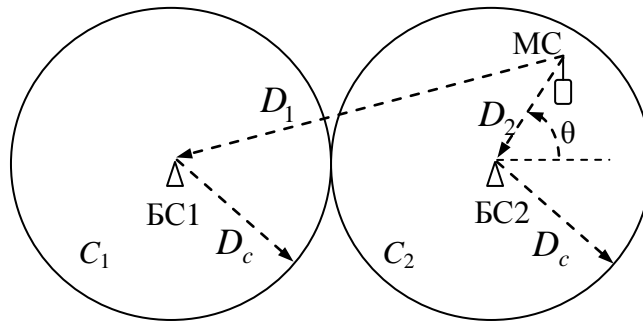


Рис.4.5. Иллюстрация к расчету межсотовых МАИ.

Усредним теперь результат (4.14) по всей площади ячейки C_2 , полагая, что любая позиция МС внутри соты равновероятна, т.е. обобщенная плотность распределения вероятности в полярных координатах D_2, θ определена как $W(D_2, \theta) = D_2 / \pi D_c^2$ внутри C_2 и нуль вне её. Тогда средняя мощность $\overline{P_{r1}}$ помехи МАИ, создаваемой одиночной сторонней МС соседней соты, составит

$$\overline{P_{r1}} = \iint_{C_2} P_{r1}(D_2, \theta) W(D_2, \theta) dD_2 d\theta = \frac{P}{\pi D_c^2} \int_0^{D_c} \int_0^{2\pi} \frac{D_2^{e+1}}{D_1^e} d\theta dD_2.$$

При значении экспоненты ослабления $e = 4$, соответствующей многим сценариям мобильной связи, интегрирование в последнем выражении может быть выполнено аналитически [20]. На основании теоремы косинусов $D_1^2 = (2D_c)^2 + D_2^2 + 2(2D_c)D_2 \cos \theta$ и

$$\overline{P_{r1}} = \frac{P}{\pi D_c^2} \int_0^{D_c} \int_0^{2\pi} \frac{D_2^5}{(4D_c^2 + D_2^2 + 4D_c D_2 \cos \theta)^2} d\theta dD_2 = \frac{2P}{\pi} \int_0^1 \int_0^\pi \frac{x^5}{(x^2 + 4x \cos \theta + 4)^2} d\theta dx.$$

Внутренний интеграл может быть оценен либо путем тригонометрической подстановки, либо найден по таблице (например, в [21]), по чего подынтегральное выражение по x становится $(4x^5 + x^7)/(4 - x^2)^3$. Прибегая вновь к таблице интегралов [21], приходим к соот-

ношению

$$\overline{P_{r1}} = P \left(16 \ln \frac{4}{3} - \frac{41}{9} \right) < 0.05P.$$

Это значение должно быть умножено на число пользователей K_c , приходящееся на одну соту, а также на число соседних ячеек, окружающих данную соту. При шестиугольном представлении ячейки последних ровно 6, и полная мощность межсотовых помех, порождаемых соседними ячейками, не превышает величины $6 \times 0.05 \times K_c P = 0.3K_c P$. Строго говоря, эту оценку в дальнейшем необходимо увеличить за счет межсотовых помех от более удаленных по сравнению с соседними сот. Однако, как легко предсказать на основании вышеприведенных вычислений, этот вклад будет пренебрежимо малым в сравнении с только что полученной оценкой [20]. Допуская некоторый запас безопасности, можно считать, что полная мощность межсотовых помех составляет $P_{I, ext} \leq 0.5K_c P$, тогда как внутренние помехи MAI, создаваемые $K_c - 1$ собственными мобильными абонентами, характеризуются, как и ранее, мощностью $P_{I, in} = (K_c - 1)P$. Теперь пороговое отношение SIR (4.10) может быть модифицировано с учетом как внутренних, так и внешних помех MAI:

$$q_{If}^2 = \frac{2E}{(K_c - 1)\frac{P}{W} + 0.5K_c \frac{P}{W}} = \frac{2WT}{1.5K_c - 1}. \quad (4.15)$$

Этот результат допускает дальнейшую ревизию для «чистой» телефонии, поскольку в диалоге любой партнер не ведет постоянного разговора и расходует некоторое время на размышление и прослушивание. Очевидно, что в течение таких пауз передатчик молчащей стороны может быть выключен или, по крайней мере, работать со значительно меньшей мощностью. Действительно, эта возможность уже использовалась в системах мобильной связи 2-го поколения (не CDMA) с целью продления срока работы батарей. Однако только в стандартах с CDMA подобная возможность допускает одновременное снижение уровня MAI и, следовательно, увеличение потенциального числа пользователей, обслуживаемых одной сотой.

Типичным значением *фактора речевой активности (voice activity factor)*, т.е. доли общего времени разговора, в течение которого участник телефонного разговора активен, составляет 3/8. Соответственно, взвешивание средней мощности MAI данным коэффициентом трансформирует вышеприведенное пороговое отношение SIR следующим образом

$$q_{If}^2 = \frac{16WT}{4.5K_c - 3}.$$

Разрешив его относительно K_c , получаем много более ободряющую оценку числа пользователей в CDMA по сравнению с изначальной (4.12), полученной безотносительно к специфике топологии системы:

$$K_c \leq \frac{32WT}{9q_{If}^2} + \frac{2}{3}. \quad (4.16)$$

Подставив в (4.16) упоминавшуюся ранее цифру в 7 дБ для требуемого отношения SIR, имеем

$$K_c \leq \frac{32WT}{45} + \frac{2}{3}. \quad (4.17)$$

Оценка, даваемая этим неравенством, примерно в пять раз больше той, которая устанавливается (4.13) при одном и том же частотно-временном ресурсе. Данный факт свидетельствует о значительно большей привлекательности сотовой системы с асинхронным CDMA в сравнении с традиционными ортогональными схемами множественного доступа на основе FDMA и TDMA.

Пример 4.6.1. Предположим, что для организации множественного доступа в полосе 5МГц, типичной для систем 3-го поколения, используется метод FDMA. При скорости передачи кодированной речи в 19.2 кбит/сек и бинарной ФМ потенциально можно организовать связь вплоть до $K_c = W_t / 7R = 37$ пользователей на соту. В то же время, как следует из (4.17), альтернативный вариант в виде асинхронной CDMA представляется значительно лучшим, поскольку позволяет обслужить до $K_c = 32W / 45R + 2/3 \approx 185$ пользователей на отдельную соту. □

Оценки вида (4.16) и (4.17) могут показаться излишне оптимистичными, поскольку они не учитывают составляющую теплового шума. Реальная ситуация, при которой они применимы в наибольшей степени, предполагает, что излучаемая каждым мобильным абонентом мощность P настолько значительна, что общая помеха MAI доминирует над АБГШ. С другой стороны, проектировщик может быть заинтересован в применении минимально возможной излучаемой мощности, например, по причине срока работоспособности батарей или электромагнитной совместимости. Прибавив спектральную плотность мощности АБГШ к знаменателю в (4.15) и снова учтя фактор речевой активности, легко показать, что, когда требуется принимать во внимание «чистое» (т.е. без MAI) отношение SNR по мощности q^2 и охватывающее MAI и шум отношение SIR q_I^2 , выражение (4.16) преобразуется к виду

$$K_c \leq \frac{32WT}{9q_I^2} \left(1 - \frac{q_I^2}{q^2} \right) + \frac{2}{3}. \quad (4.18)$$

Если, например, полная мощность MAI имела бы такой же уровень, как и мощность шума в пределах полосы сигнала, то $q^2 = 2q_I^2$, и тогда

$$K_c \leq \frac{16WT}{9q_I^2} + \frac{2}{3}, \quad (4.19)$$

что, составляет примерно половину от вычисленного значения при пороговом отношении SIR (см. (4.16)). Для установленного ранее отношения SIR (7 дБ) имеем

$$K_c \leq \frac{16WT}{45} + \frac{2}{3}. \quad (4.20)$$

Хотя с уменьшением мощности сигнала число пользователей на соту уменьшается в два раза, оно, тем не менее, более чем вдвое превышает аналогичный показатель для FDMA и TDMA.

Еще одним достоинством систем с асинхронным вариантом CDMA является специфический характер блокировки. Во всех реальных многопользовательских системах физические каналы (будь то частотные суб-полосы при FDMA, временные слоты при TDMA или кодовые сигнатуры при CDMA) не назначаются потребителям раз и навсегда. Вместо этого сеть сама управляет совокупностью каналов трафика и выделяет пользователю один из них только тогда, когда от него поступит запрос на доступ к сети. Конечно, в этом случае некоторый системный ресурс окажется зарезервированным реализации запросного канала. В системах с FDMA или TDMA число физических каналов жестко фиксировано и время от времени может произойти блокировка, т.е. ситуация, при которой сеть отвергнет пользовательский запрос, поскольку все каналы уже заняты. Вероятность блокировки, равная 2%, часто полагается допустимой и, исходя из нее, определяется необходимое число каналов. По различным причинам модель распределения абонентов в зоне охвата сетью может изменяться настолько значительно, что в некоторых сотах блокировка становится недопустимо вероятной. В этом случае оператор сети может оказаться вынужденным осуществить реконфигурацию сети, которая выльется в частотную перепланировку, затрагивающую все соты.

Сценарии, характерные для CDMA, совершенно отличаются от ранее рассмотренных. Во-первых, если число уже активных пользователей равно номинальному, опреде-

ляемому (4.16) или (4.20), и поступает еще один запрос, то он может быть удовлетворен путем назначения сигнатуры, которая отличается от всех уже используемых. Это приведет к некоторому (как правило, незначительному) уменьшению как отношения SIR, так и качества обслуживания всех активных пользователей. Следовательно, вместо прямого отказа происходит мягкая блокировка. Во-вторых, если с течением времени трафик в некоторой области увеличивается драматически, оператор может ввести в действие дополнительную базовую станцию в «горячем месте» вместо частотной перепланировки или любого другого радикального действия, затрагивающего остальные сторонние ячейки.

Основываясь на анализе главы, сделаем оправданное заключение о том, что широкополосность может оказать гибкую и эффективную помощь в обеспечении множественного доступа, а сотовые системы относятся к таким, в которых преимущества CDMA проявляются наиболее убедительно.

Задачи.

4.1. Цифровая система передачи данных с FDMA должна обслуживать не менее 100 пользователей. Оценить минимальную общую полосу, занимаемую системой, если необходимая скорость передачи данных пользователем составляет 20 кбит/сек и используется бинарная ФМ. Как изменится значение полосы, если BPSK заменить QPSK? Ответить на те же вопросы, если вместо FDMA используется TDMA.

4.2. FDMA система с QPSK предназначена для обмена цифровой информацией между летательными аппаратами и работает в диапазоне 3 ГГц. Максимальная скорость движения составляет 1800 км/час, стабильность опорного генератора – $2 \cdot 10^{-7}$, защитный интервал, обусловленный не прямоугольностью фильтра – 1 КГц. Определить максимальное число пользователей, которое может обслужить система при полосе 2.32 МГц и скорости передачи данных пользователем – 20 кбит/сек.

4.3. Цифровая многопользовательская система с TDMA должна обслуживать не менее 100 абонентов. В системе используется модуляция с помощью ФМ-8. Определить минимальную полосу, занимаемую системой, если требуемая скорость передачи на одного пользователя составляет 20 кбит/сек.

4.4. Одиночному частотному суб-каналу линии «вверх» цифровой TDMA сотовой системы с бинарной ФМ отведена полоса в 200 КГц. Временной интервал между последовательными пакетами данных произвольного пользователя должен составлять не более 5 мсек, скорость передачи данных потребителем – около 20 кбит/сек, максимальный радиус соты – 30 км. Определить максимальное число TDMA каналов, содержащихся в одном частотном суб-канале.

4.5. В синхронном варианте CDMA системы должно быть организовано 128 физических каналов. Передача данных осуществляется с помощью ФМ-8 при скорости передачи данных каждым пользователем, равной 20 кбит/сек. Оценить минимально необходимую полосу, занимаемую системой.

4.6. Некоторая синхронная CDMA система обладает 50 физическими каналами, по которым каждым пользователем данные с применением ФМ-16 передаются со скоростью 20 кбит/сек. Общая полоса, занимаемая системой, составляет 500 КГц. Какова величина выигрыша от обработки в системе? Свободна ли система от помех множественного доступа? Что произойдет при уменьшении полосы в 4 раза?

4.7. В пределах одной ячейки CDMA системы сотового телефона находятся два абонента, расположенных на расстоянии 500 м и 5 км от базовой станции. Мобильный абонент, находящийся на большем расстоянии, излучает сигнал мощностью 100 мВт. Определить мощность, излучаемую более близким пользователем, в предположении эффективной работы регуляции мощности.

4.8. В канале «вверх» сотового телефона стандарта IS-95 используется асинхронный вариант CDMA. Передача данных осуществляется посредством ортогональных сигналов, кодирующих 6-ти битовые блоки, со скоростью 28.8 кбит/сек. Полоса сигнала может быть положена равной 1.25 МГц. Каково будет число пользователей на соту, если минимально требуемое отношение SIR составляет 7 дБ, фактор речевой активности равен 3/8, внешняя помеха MAI добавляет к внутренней 50%, а тепловой шум пренебрежимо мал. Что изменится, если отношение SNR только для теплового шума составляет 9 дБ?

4.9. Оценить число необходимых сотовых позиций при условиях, задаваемых предыдущей задачей (АБГШ нельзя пренебрегать) для обслуживания зоны с 50 000 абонентов, если вероятность активного состояния абонента составляет 0.02. Сравнить результат со случаем использования FDMA системы.

4.10. Как изменится потенциальное число пользователей, если в условиях задачи 4.8 ортогональную 6-ти битовую сигнализацию заменить бинарной ФМ, причем полоса и скорость данных останутся прежними?

4.11. Основываясь на модели Окумура–Хаты, показать, что асинхронный вариант CDMA неприменим в канале «вверх» типичной сотовой системе с макросотами без эффективной регулировки мощности. Пренебречь компонентом теплового шума.

4.12. Имеется синхронный канал «вниз» CDMA сотовой системы. В пределах одной соты используется максимальное число ортогональных сигнатур. Как оценить воздействие окружающих базовых станций на приемники мобильных абонентов данной соты при отсутствии распределения частотно-временного ресурса между сотами? Какое положение вы бы выдвинули в качестве основного при выборе сигнатур для подобной системы?

5. Дискретные широкополосные сигналы.

5.1. Широкополосная модуляция.

Обратимся вновь к общей модели (2.37) полосного сигнала

$$s(t) = \operatorname{Re}[\dot{S}(t) \exp(j2\pi f_0 t)], \quad \dot{S}(t) = S(t) \exp(j\gamma(t)).$$

Совершенно очевидно, что распределение спектра сигнала осуществляется посредством соответствующего управления комплексной огибающей сигнала $\dot{S}(t)$, т.е. модуляцией мгновенных значений амплитуды $S(t)$ и начальной фазы $\gamma(t)$. Как уже было отмечено в главе 1, «чистая» амплитудная модуляция не может служить эффективным инструментом для распределения спектра, поскольку она значительно расширяет полосу только ценой концентрации энергии сигнала в пределах коротких временных интервалов. Действительно, данный тип модуляции подразумевает работу с короткими простыми сигналами. Напротив, угловая (фазовая или частотная) модуляция способна безгранично (по крайней мере, теоретически) расширить спектр без изменения распределения энергии сигнала во времени, т.е. длительности сигнала, благодаря чему ее роль в технологии распределенного спектра является фундаментальной. Амплитудная же модуляция служит только вспомогательным инструментом, который иногда проявляет себя продуктивно в комбинации с угловой модуляцией.

В зависимости от типа привлеченной модуляции все широкополосные сигналы могут быть подразделены на *непрерывные* и *дискретные*. Для первых закон модуляции, т.е. комплексная огибающая $\dot{S}(t)$, является непрерывной функцией времени, тогда как модулируемые параметры вторых (амплитуда, частота, начальная фаза) – кусочно-постоянной, скачкообразно изменяющей свои значения только в дискретные моменты времени. Пример непрерывного широкополосного сигнала будет кратко обсужден в параграфе 6.2, однако, в дальнейшем основное внимание будет сфокусировано на дискретных сигналах, учитывая их доминирующую роль в большинстве современных и проектируемых коммерческих системах.

5.2. Обобщенная модель и категории дискретных сигналов.

Дискретные сигналы, рассматриваемые в книге, могут быть охвачены следующим описанием, которое обобщает вариант уже примененный в 2.7.3. Дискретный сигнал представляет собой последовательность элементарных символов (импульсов) фиксированной формы, повторяющихся с некоторым фиксированным временным интервалом. Элементарный импульс, называемый *чипом*, комплексная огибающая $\dot{S}_0(t)$, определяющая его форму, и внутренняя угловая модуляция могут быть любыми. Как правило (но не обязательно), временной интервал Δ между последовательными чипами равен или превосходит длительность чипа Δ_c . Модуляция всего сигнала заключается в манипулировании амплитудами, фазами и, возможно, частотами отдельных чипов. Тогда формальное представление комплексной огибающей дискретного сигнала дается соотношением

$$\dot{S}(t) = \sum_{i=-\infty}^{\infty} a_i \dot{S}_0(t - i\Delta) \exp(j2\pi F_i t), \quad (5.1)$$

где, в дополнение к уже объясненным обозначениям, a_i и F_i – соответственно комплексная амплитуда и частота (в значениях сдвига относительно фиксированной центральной частоты) i -го чипа. Очевидно, что последовательность $\{|a_i|, i = \dots, -1, 0, 1, \dots\}$ определяет вещественные амплитуды чипов, т.е. их амплитудную модуляцию. Аналогично, последо-

вательности $\{\phi_i = \arg a_i, i = \dots, -1, 0, 1, \dots\}$ и $\{F_i, i = \dots, -1, 0, 1, \dots\}$ задают законы модуля-

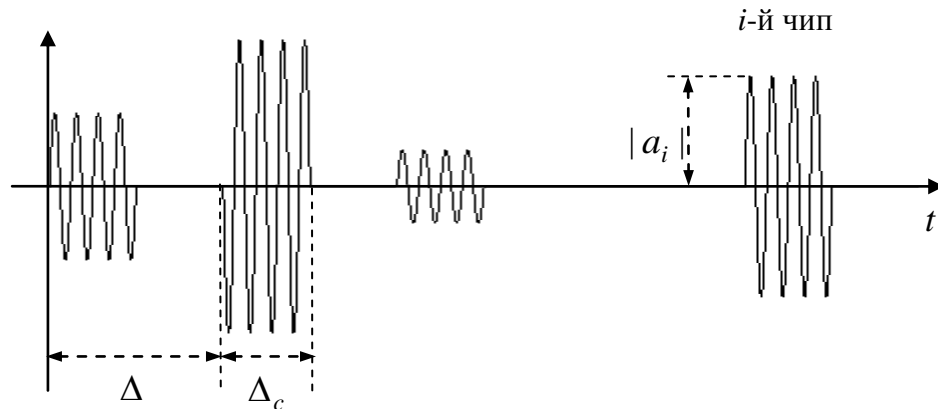


Рис. 5.1. Пример дискретного сигнала.

ции чипов по фазе и частоте. Рис. 5.1 может быть полезен для понимания некоторых вышеприведенных определений.

Предположим, что в модели (5.1) вещественные амплитуды $|a_i|$ могут принимать ненулевые значения только при попадании i в диапазон $0 \leq i \leq N-1$, а при $i < 0$ и $i > N$ значения амплитуд $|a_i| = 0$. В этом случае сигнал представляет собой пакет конечного числа N манипулированных чипов. Подобный сигнал будем называть *импульсным* или *апериодическим*. Длительность апериодического сигнала определится, как $T = (N-1)\Delta + \Delta_c$. Другой важной версией является сигнал, у которого закон модуляции повторяется с периодом N чипов: $a_i = a_{i+N}$, $F_i = F_{i+N}$, $i = \dots, -1, 0, 1, \dots$. Естественно называть дискретный сигнал подобного типа *периодическим*. Фактический период сигнала составляет $T = N\Delta$, и любой периодический сигнал представляет собой по сути повторение с периодом $N\Delta$ апериодического, который, в свою очередь, является однопериодным сегментом периодического сигнала. В обоих случаях параметр N назовем *длиной* дискретного сигнала.

В рамках описанной обобщенной модели различают несколько категорий дискретного сигнала в зависимости от конкретного способа модуляции чипа.

1. Если манипуляции подвергаются только комплексные амплитуды чипов, а все частоты остаются одинаковыми ($F_i = 0$, $i = 0, 1, \dots, N-1$), то сигнал называется *амплитудно-фазоманипулированным* (АФМ) (*amplitude-phase shift keying* (APSK)). Последовательность комплексных амплитуд чипов $\{a_i, i = 0, 1, \dots, N-1\}$ называется *кодовой последовательностью* или просто *кодом*.

2. Если манипуляция осуществляется только над фазами чипов АФМ сигнала, а амплитуды остаются неизменными ($|a_i| = 1$, $i = 0, 1, \dots, N-1$), то сигнал является ФМ (PSK). ФМ сигналы типичны для т.н. широкополосных систем с прямым расширением спектра (*direct sequence spread spectrum systems*) (см. параграф 7.1).

3. Дальнейшая типизация в рамках ФМ сигналов может производиться с учетом модуляционного алфавита. Если применяются только бинарные комплексные амплитуды ($a_i = \pm 1$, или, что эквивалентно, $|a_i| = 1, \phi_i \in \{0, \pi\}, i = 0, 1, \dots, N-1$), то сигнал называется бинарным ФМ (BPSK), при четверичном алфавите вида $a_i = \pm 1, \pm j$, или, что эквивалентно, $|a_i| = 1, \phi_i \in \{0, \pi, \pm \pi/2\}, i = 0, 1, \dots, N-1$ сигнал называется квадратурным ФМ (QPSK), и т.д.

4. Если регулированию подлежат только частоты чипов, а комплексные амплитуда остаются постоянными, то сигнал является частотно–манипулированным (ЧМ или FSK). Кодовая последовательность подобного сигнала представляет собой последовательность частот $\{F_i, i = 0, 1, \dots, N-1\}$. Сигналы этого типа используются, в частности, в системах с прыгающей частотой (frequency hopping systems) (см. 7.1).

5.3. Корреляционные функции АФМ сигналов.

Корреляционные функции, характеризующие степень подобия сдвинутых во времени копий сигналов, играют критически важную роль в задачах измерения времени и разрешения (см. 2.11–2.16). Искусство проектирования широкополосных систем, как это можно будет увидеть в дальнейшем, во многих аспектах базируется на нахождении сигналов с соответствующими корреляционными свойствами. В данном параграфе получим обобщенное выражение для корреляционных функций АФМ сигналов. На основании приведенных ранее определений комплексная огибающая АФМ сигнала имеет вид

$$\dot{S}(t) = \sum_{i=-\infty}^{\infty} a_i \dot{S}_0(t-i\Delta). \quad (5.2)$$

Обратимся к определению нормированной АКФ (2.67), учитывая, что для периодического сигнала подынтегральное выражение также будет периодическим, и, следовательно, усреднение по времени (интегрирование) может быть выполнено на одном периоде, причем нормирование производится к энергии за один период. Таким образом, при наиболее типичном для практики предположении, что $\Delta_c \leq \Delta$,¹ приходим к универсальному выражению

$$\dot{\rho}(\tau) = \frac{1}{E} \int_0^T \dot{S}(t) \dot{S}^*(t-\tau) dt \quad (5.3)$$

как для аperiodического, так и периодического сигнала, где $E = \|\mathbf{a}\|^2 E_0$ – полная энергия для первого и энергия за период для второго варианта. В свою очередь E_0 определяет энергию чипа, а $\|\mathbf{a}\|$ – длина (евклидова норма) *кодového вектора* $\mathbf{a} = (a_0, a_1, \dots, a_{N-1})$, или, другими словами, $\|\mathbf{a}\|^2 = \sum_{i=0}^{N-1} |a_i|^2$ – энергия N -элементной последовательности $\{a_0, a_1, \dots, a_{N-1}\}$.

Подстановка (5.2) в (5.3) дает

$$\dot{\rho}(\tau) = \frac{1}{E} \sum_{i=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} a_i a_k^* \int_0^T \dot{S}_0(t-i\Delta) \dot{S}_0^*(t-k\Delta-\tau) dt = \frac{1}{E} \sum_{k=-\infty}^{\infty} \sum_{i=0}^{N-1} a_i a_k^* \int_0^T \dot{S}_0(t-i\Delta) \dot{S}_0^*(t-k\Delta-\tau) dt$$

где последнее равенство следует из исчезновения интеграла, если i не принадлежит множеству $\{0, 1, \dots, N-1\}$.

Введение АКФ одиночного чипа

$$\dot{\rho}_c(\tau) = \frac{1}{E_0} \int_{-\infty}^{\infty} \dot{S}_0(t) \dot{S}_0^*(t-\tau) dt \quad (5.4)$$

приводит к выражению

$$\dot{\rho}(\tau) = \sum_{k=-\infty}^{\infty} \left(\frac{1}{\|\mathbf{a}\|^2} \sum_{i=0}^{N-1} a_i a_k^* \right) \dot{\rho}_c[\tau - (i-k)\Delta].$$

¹ В результате приходим к выражению, справедливому вне зависимости от истинности или нет данного предположения. Оно только поможет избежать некоторых второстепенных деталей.

Теперь замена индекса суммирования k на $m = i - k$ дает

$$\dot{\rho}(\tau) = \sum_{m=-\infty}^{\infty} \rho(m) \dot{\rho}_c(\tau - m\Delta), \quad (5.5)$$

где

$$\rho(m) = \frac{1}{\|\mathbf{a}\|^2} \sum_{i=0}^{N-1} a_i a_{i-m}^* \quad (5.6)$$

АКФ кодовой последовательности $\{a_0, a_1, \dots, a_{N-1}\}$, характеризующая схожесть последней с копией, сдвинутой на m позиций.

Соотношение (5.5) имеет совершенно явную трактовку. Сравнение (5.5) с моделью (5.2) позволяет увидеть, что АКФ АФМ сигнала может рассматриваться, как собственно АФМ сигнал. В качестве элементарного символа последнего выступает АКФ $\dot{\rho}_c(\tau)$ исходного чипа, тогда как кодовой последовательностью служит АКФ (5.6) кодовой последовательности $\{a_0, a_1, \dots, a_{N-1}\}$ изначального сигнала. Следовательно, при заданном элементарном символе АКФ АФМ сигнала полностью определяется АКФ $\rho(m)$ кодовой последовательности (в дальнейшем АКФ *кода*), и синтез АФМ сигналов с хорошими корреляционными свойствами состоит в отыскании последовательностей с хорошими АКФ. Отметим также, что, как и для любой АКФ, значение $\rho(m)$ при $m = 0$ равняется единице, а сама АКФ является четной функцией $\rho(m) = \rho^*(-m)$.

В многопользовательских системах с CDMA необходимы семейства дискретных сигналов с особенными взаимными корреляционными свойствами (см. параграф 4.5 и главу 7). Повторив в точности ранее приведенный вывод, но теперь для двух различных (k -го и l -го) АФМ сигналов, обладающих идентичными чипами и длиной, придем к следующему соотношению для их ВКФ

$$\dot{\rho}_{kl}(\tau) = \sum_{m=-\infty}^{\infty} \rho_{kl}(m) \dot{\rho}_c(\tau - m\Delta), \quad (5.7)$$

где

$$\rho_{kl}(m) = \frac{1}{\|\mathbf{a}_k\| \|\mathbf{a}_l\|} \sum_{i=0}^{N-1} a_{k,i} a_{l,i-m}^* \quad (5.8)$$

ВКФ кодовых последовательностей $\{a_{k,0}, a_{k,1}, \dots, a_{k,N-1}\}$ и $\{a_{l,0}, a_{l,1}, \dots, a_{l,N-1}\}$ двух сигналов, которая характеризует степень похожести первого со сдвинутой на m позиций копией второго сигнала. Очевидно, что ВКФ (5.7) снова представляет собой АФМ сигнал, у которого роль кодовой последовательности выполняет ВКФ двух исходных кодов (т.е. ВКФ кода), а синтез семейств с необходимыми взаимно корреляционными свойствами заключается в отыскании семейства последовательностей, обладающего соответствующими ВКФ. Соотношения (5.7) и (5.8) являются наиболее общими, поскольку АКФ k -го сигнала есть $\dot{\rho}_{kk}(\tau)$, что справедливо и в отношении кодовых последовательностей.

Что же касается дальнейшего, то полученные результаты найдут самое широкое применение, иногда опуская первый нормирующий множитель в (5.6) и (5.8), т.е. оперируя с ненормированными корреляционными функциями кодовых последовательностей вида:

$$R(m) = \sum_{i=0}^{N-1} a_i a_{i-m}^*, \quad R_{kl}(m) = \sum_{i=0}^{N-1} a_{k,i} a_{l,i-m}^*. \quad (5.9)$$

5.4. Вычисление корреляционных функций кодовых последовательностей.

Рассмотрим кодовую последовательность $\{a_0, a_1, \dots, a_{N-1}\}$. Если она используется для формирования импульсного сигнала, то в обобщенной модели (5.2) $a_i = 0$ при всех отрицательных i и $i \geq N$, так что, согласно (5.6), *апериодическая* или *импульсная* АКФ вычисляется как

$$\rho_a(m) = \begin{cases} \frac{1}{\|\mathbf{a}\|^2} \sum_{i=m}^{N-1} a_i a_{i-m}^*, & m \geq 0, \\ \frac{1}{\|\mathbf{a}\|^2} \sum_{i=0}^{N-1+m} a_i a_{i-m}^*, & m < 0. \end{cases} \quad (5.10)$$

В принципе, вторая строка в (5.10) является избыточной, поскольку любая АКФ обладает свойством четности и, в частности, $\rho_a(-m) = \rho_a^*(m)$. Очевидным образом также следует, что без нормирующего множителя апериодическая АКФ представляет собой скалярное произведение вектора $\mathbf{a} = (a_0, a_1, \dots, a_{N-1})$ с его *не циклически* сдвинутой на m позиций копией. Последнее означает, что при сдвиге вектор \mathbf{a} вправо ($m \geq 0$) или влево ($m < 0$) сумма в (5.10) учитывает только перекрывающиеся компоненты \mathbf{a} и его сдвинутой копии, а все остальные как бы приравниваются нулю. Например, для вычисления $\rho_a(1)$ первоначально следует записать одну под другой исходную последовательность и ее комплексно сопряженную копию, сдвинутую вправо на одну позицию, затем вычислить все компоненты в виде поэлементных произведений и сложить их.

$$\begin{array}{cccccc} a_0 & a_1 & a_2 & a_3 & \dots & a_{N-1} \\ & a_0^* & a_1^* & a_2^* & \dots & a_{N-2}^* \end{array} \quad \rho_a(1) = \frac{1}{\|\mathbf{a}\|^2} \sum_{i=1}^{N-1} a_i a_{i-1}^* .$$

Будем полагать теперь, что сигнал является периодическим, т.е. $a_{i+N} = a_i, i = \dots, -1, 0, 1, \dots$. Тогда соотношение (5.6) определяет *периодическую* АКФ $\rho_p(m)$, сумма в котором всегда содержит N слагаемых, поскольку $a_{-1} = a_{N-1}, a_{-2} = a_{N-2}$ и т.д.

$$\rho_p(m) = \frac{1}{\|\mathbf{a}\|^2} \sum_{i=0}^{N-1} a_i a_{i-m}^* . \quad (5.11)$$

В этом случае скалярное произведение вычисляется для исходной кодовой последовательности и ее *циклически* сдвинутой копии, где при $m \geq 0$ m крайних левых «пустых» позиций заполняются символами, вытолкнутыми с правой стороны. Например, схема вычисления $\rho_p(1)$ выглядит следующим образом:

$$\begin{array}{cccccc} a_0 & a_1 & a_2 & a_3 & \dots & a_{N-1} \\ a_{N-1}^* & a_0^* & a_1^* & a_2^* & \dots & a_{N-2}^* \end{array} \quad \rho_p(1) = \frac{1}{\|\mathbf{a}\|^2} \sum_{i=0}^{N-1} a_i a_{i-1}^* .$$

Поскольку $\rho_p(m)$ вычисляется в предположении о периодичности кодовой последовательности, то она сама периодична с периодом N , т.е. $\rho_p(m) = \rho_p(m+N)$, $m = \dots, -1, 0, 1, \dots$, что непосредственно следует и (5.11), и, в свою очередь, трансформирует

свойство четности к виду

$$\rho_p(-m) = \rho_p(N-m) = \rho_p^*(m-N). \quad (5.12)$$

Это соотношение показывает, что $\rho_p(m)$ полностью характеризуется своими значениями при сдвигах $m=1, 2, \dots, [N/2]$, где $[\cdot]$ символизирует взятии целой части. Другое важное свойство периодической АКФ вытекает из (5.11) после разделения суммы на две как

$$\rho_p(m) = \frac{1}{\|\mathbf{a}\|^2} \sum_{i=m}^{N-1} a_i a_{i-m}^* + \frac{1}{\|\mathbf{a}\|^2} \sum_{i=0}^{m-1} a_i a_{i-m}^*, \quad m \geq 0.$$

Первое слагаемое в последнем выражении является аperiodической АКФ $\rho_a(m)$ (см. (5.10)), тогда как второе равно $\rho_a(m-N)$, что непосредственно проверяется путем вычисления $\rho_a(m-N)$ согласно второй строки (5.10). В результате получаем соотношение, связывающее периодическую АКФ со своим аperiodическим аналогом

$$\rho_p(m) = \rho_a(m) + \rho_a(m-N), \quad m = 0, 1, \dots, N. \quad (5.13)$$

Соотношение (5.13) играет ключевую роль в синтезе импульсных сигналов с хорошими корреляционными свойствами (см. параграф 6.9).

Пример 5.3.1. Таблица 5.1 иллюстрирует процедуру вычисления аperiodической и периодической АКФ на примере бинарной последовательности длины $N=8$ $\{++++--+-\}$. В таблице символы бинарного кода $+1$ и -1 представлены только своими знаками « $+$ » и « $-$ » соответственно, что отвечает обычной практике, значения ненормированной АКФ, а символы, не участвующие в вычислении аperiodической АКФ, отмечены затенением. Результаты вычислений после нормировки отображены на рис. 5.2, где построены автокорреляционные функции АФМ сигнала с прямоугольной огибающей элементарных символов длительности $\Delta_c = \Delta$ и рассмотренной кодовой последовательностью. Сплошная, пунктирная и штрих-пунктирная линии отвечают периодической АКФ $\dot{\rho}_p(\tau)$ и сдвинутым копиям $\dot{\rho}_a(\tau)$ и $\dot{\rho}_a(\tau-T)$ аperiodической АКФ соответственно. Очевидно, что графики подтверждают справедливость соотношений (5.12) и (5.13). \square

Таблица 5.1.

m	a_0	a_1	a_2	a_3	a_4	a_5	a_6	a_7	$R_a(m)$	$R_p(m)$
0	+	+	+	-	-	+	-	-	+8	+8
1	-	+	+	+	-	-	+	-	+1	0
2	-	-	+	+	+	-	-	+	-2	-4
3	+	-	-	+	+	+	-	-	+1	0
4	-	+	-	-	+	+	+	-	0	0
5	-	-	+	-	-	+	+	+	-1	0
6	+	-	-	+	-	-	+	+	-2	-4
7	+	+	-	-	+	-	-	+	-1	0

При вычислении ВКФ двух последовательностей одинаковой длины снова можно выделить аperiodическую АКФ $\rho_{a,kl}(m)$ и периодическую $\rho_{p,kl}(m)$, определяемые как

$$\rho_{a,kl}(m) = \begin{cases} \frac{1}{\|\mathbf{a}_k\| \|\mathbf{a}_l\|} \sum_{i=m}^{N-1} a_{k,i} a_{l,i-m}^*, & m \geq 0, \\ \frac{1}{\|\mathbf{a}_k\| \|\mathbf{a}_l\|} \sum_{i=0}^{N+m-1} a_{k,i} a_{l,i-m}^*, & m < 0, \end{cases} \quad (5.14)$$

и

$$\rho_{p,kl}(m) = \frac{1}{\|\mathbf{a}_k\| \|\mathbf{a}_l\|} \sum_{i=0}^{N-1} a_{k,i} a_{l,i-m}^* . \quad (5.15)$$

Соотношение (5.13) также устанавливает для ВКФ, что

$$\rho_{p,kl}(m) = \rho_{a,kl}(m) + \rho_{a,kl}(m-N) , \quad (5.16)$$

но, что касается четности или единичного значения при $m=0$, то эти свойства не присущи произвольной ВКФ, как это было в случае АКФ.

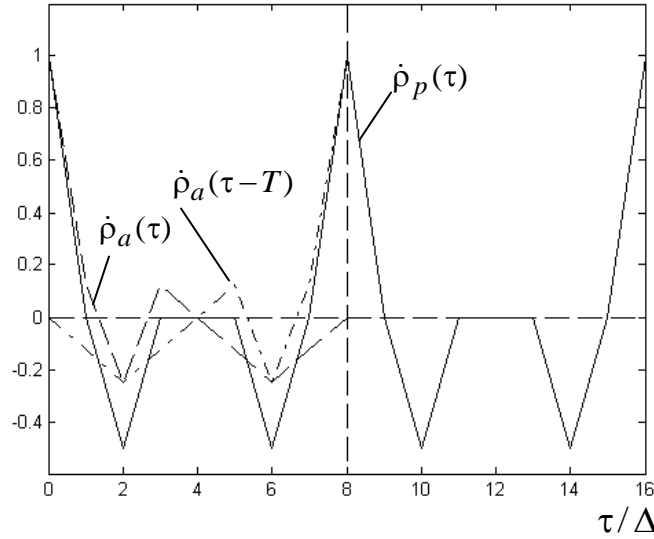


Рис. 5.2. Автокорреляционные функции бинарного сигнала длины 8.

5.5. Корреляционные функции ЧМ сигналов.

Осуществим теперь аналогичную работу, что и в предыдущих двух параграфах, но в приложении к ЧМ сигналам. На основании определения, введенного в 5.2, комплексная огибающая ЧМ сигнала может быть представлена в виде

$$\dot{S}(t) = \sum_{i=-\infty}^{\infty} a_i \dot{S}_0(t-i\Delta) \exp(j2\pi F_i t) , \quad (5.17)$$

где все a_i равняются единице в случае периодического сигнала, тогда как для импульсного сигнала длины N $a_i = 1, 0 \leq i < N$ и $a_i = 0$, если i выходит за пределы интервала $0 \leq i < N$.

Вне зависимости от периодичности или конечности сигнала на тех же основаниях, что и ранее, можно использовать универсальное выражение для АКФ вида

$$\begin{aligned} \dot{\rho}(\tau) &= \frac{1}{NE_0} \int_0^T \dot{S}(t) \dot{S}^*(t-\tau) dt = \\ &= \frac{1}{NE_0} \sum_{k=-\infty}^{\infty} \sum_{i=0}^{N-1} a_i a_k \exp(j2\pi F_k \tau) \int_0^T \dot{S}_0(t-i\Delta) \dot{S}_0^*(t-k\Delta-\tau) \exp(j2\pi(F_i - F_k)t) dt , \end{aligned}$$

где использован тот факт, что все чипы с ненулевыми амплитудами обладают одинаковыми энергиями E_0 , т.е. $\|\mathbf{a}\|^2 = N$ и $E = NE_0$.

Для ЧМ манипуляции типично использование равномерного частотного алфавита, так что $F_i \in \{0, \pm F, \pm 2F, \dots\}$, где частотный шаг F не меньше, чем полоса, занимаемая чипом. Таким образом, спектры двух элементарных символов, обладающих частотами F_i и F_k , не перекрываются, а сами чипы ортогональны независимо от их временного рассогласования (см. параграф 4.4), если только $F_i \neq F_k$. Учитывая данный факт, приходим к выражению

$$\begin{aligned} \dot{\rho}(\tau) &= \frac{1}{N} \sum_{k=-\infty}^{\infty} \sum_{i=0}^{N-1} a_i a_k \exp(j2\pi F_k \tau) \delta(F_i - F_k) \dot{\rho}_c[\tau - (i-k)\Delta] = \\ &= \frac{1}{N} \sum_{m'=-\infty}^{\infty} \sum_{i=0}^{N-1} a_i a_{i-m'} \exp(j2\pi F_i \tau) \delta(F_i - F_{i-m'}) \rho_c(\tau - m'\Delta), \end{aligned} \quad (5.18)$$

где $\dot{\rho}_c(\cdot)$ – как и ранее, АКФ чипа, а

$$\delta(x-y) = \begin{cases} 1, & x=y, \\ 0, & x \neq y. \end{cases}$$

В отличие от АФМ сигналов АКФ (5.18) в общем случае не может претерпеть дальнейшее упрощение с целью приведения к виду, подобному (5.5). Общепринятой практикой является анализ поведения АКФ ЧМ сигналов в основном при задержках, кратных длительности чипа: $\tau = m\Delta$, где m – целое. Предполагая, что на длительности чипа укладывается целое число периодов l каждой частоты ($F\Delta = l$), и, принимая во внимание, что $\dot{\rho}_c(0) = 1$, $\dot{\rho}_c(\tau) = 0$, $|\tau \geq \Delta|$, подстановка $\tau = m\Delta$ в (5.18) оставляет там только одно слагаемое в сумме по m , отвечающее $m' = m$, так что

$$\dot{\rho}(m\Delta) = \frac{1}{N} \sum_{i=0}^{N-1} a_i a_{i-m} \delta(F_i - F_{i-m}). \quad (5.19)$$

В случае финитного сигнала и $m \geq 0$ все слагаемые с индексами, не принадлежащими множеству $m, m+1, \dots, N-1$, исчезают, и аperiodическая АКФ ЧМ сигнала $\dot{\rho}(m\Delta) = \rho_a(m)$ принимает вид

$$\rho_a(m) = \frac{1}{N} \sum_{i=m}^{N-1} \delta(F_i - F_{i-m}), \quad m \geq 0, \quad \rho_a(-m) = \rho_a(m), \quad (5.20)$$

где отсутствует необходимость в комплексном сопряжении во втором соотношении (устанавливаемом свойство четности), поскольку $\dot{\rho}(m\Delta)$ всегда вещественна.

В случае периодического сигнала сумма в (5.19) не содержит нулевых произведений $a_i a_{i-m}$ и периодическая АКФ сигнала $\dot{\rho}(m\Delta) = \rho_p(m)$ имеет следующий вид

$$\rho_p(m) = \frac{1}{N} \sum_{i=0}^{N-1} \delta(F_i - F_{i-m}). \quad (5.21)$$

Суммы в соотношениях (5.20)–(5.21) аккумулируют число совпадений частот в ЧМ сигнале и его копии, сдвинутой во времени на m позиций чипов. Следовательно, для вычисления АКФ ЧМ сигнала в точке $m\Delta$ достаточно подсчитать число пар $\{F_i, F_{i-m}\}$ с одинаковыми значениями F_i и F_{i-m} , где индекс i пробегает диапазон значений $\{m, m+1, \dots, N-1\}$ в случае аperiodической АКФ и $m \geq 0$ или $\{0, 1, \dots, N-1\}$ для периодической АКФ. Ясно, что соотношение (5.13), связывающее периодическую и аperiodическую АКФ, остается справедливым и для ЧМ сигналов.

Одним из широко используемых вариантов представления ЧМ сигнала является задание его помощью решетки размерности $M \times N$, в которой горизонтальное и вертикальное направления отождествлены соответственно со временем и частотой, причем M определяет размер частотного алфавита (т.е. число частот, используемых при манипуляции).

В i -м вертикальном столбце этой решетки отмечается (например, точкой или затенением) только один элемент, который соответствует частоте i -го чипа. Тогда для вычисления аperiodической АКФ при некотором m достаточно только просуммировать по всем строкам число помеченных пар, находящихся на расстоянии m , и отнормировать результат, если это необходимо. Если же интерес представляет периодическая АКФ, указанные суммы найдутся для значений m и $N - m$, которые затем складываются.

Пример 5.5.1. На рис. 5.3 с помощью решетки задан закон модуляции ЧМ сигнала длины $N = 8$ с параметрами $N = 8, M = 5$. Его ненормированная АКФ

$$R_a(m) = \sum_{i=m}^7 \delta(F_i - F_{i-m})$$

принимает значения 8, 1, 1, 0, 1, 0, 0, 0 соответственно при $m = 0, 1, \dots, 7$, поскольку имеется только одна помеченная пара вдоль первой линии на расстоянии 1, одна подобная пара на расстояния 2 и т.д. Непосредственно из (5.13) могут быть найдены значения ненормированной периодической АКФ как 8, 1, 1, 0, 2, 0, 1, 1. \square

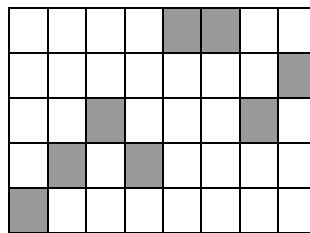


Рис. 5.3. ЧМ сигнал с параметрами $N=8, M=5$.

Обобщение результатов (5.19), (5.20) на случай ВКФ может быть осуществлено без труда посредством только согласования обозначений

$$\rho_{a,kl}(m) = \begin{cases} \frac{1}{N} \sum_{i=m}^{N-1} \delta(F_{k,i} - F_{l,i-m}), & m \geq 0, \\ \frac{1}{N} \sum_{i=0}^{N+m-1} \delta(F_{k,i} - F_{l,i-m}), & m < 0, \end{cases} \quad (5.22)$$

$$\rho_{p,kl}(m) = \frac{1}{N} \sum_{i=0}^{N-1} \delta(F_{k,i} - F_{l,i-m}), \quad (5.23)$$

где $\{F_{k,i}, i = 0, 1, \dots, N-1\}$ кодовая последовательность частот k -го сигнала. Очевидно, что вычисление ВКФ вновь сводится к подсчету числа совпадений частот в паре сигналов, сдвинутых во времени на m позиций чипов.

5.6. Выигрыш от обработки дискретных сигналов.

Обратимся вновь к обобщенной модели (5.1) с целью обсуждения вопроса выигрыша от обработки дискретного сигнала. Предположим, что все F_i принадлежат частотному алфавиту размера M , в котором соседние частоты разделены на величину полосы чипа, гарантируя тем самым их ортогональность при различных частотах. Далее, на каждой из M доступных частот имеется сигнальное подпространство размерности N , поскольку не накладывается никаких ограничений в выборе амплитудно-фазовой кодовой последовательности, т.е. N -мерного вектора $\mathbf{a} = (a_0, a_1, \dots, a_{N-1})$. Ортогональность этих подпространств означает, что размерность полного сигнального пространства, охватывающего все M частот, составляет MN . В параграфе 2.5 было показано, что размерность пространства полосных сигналов совпадает с общим частотно-временным ресурсом, содержащем сигналы. Будучи заинтересованными только в широкополосных сигналах, каж-

дый из которых занимает все доступное пространство, можно предсказать, что частотно-временное произведение дискретного сигнала, т.е. выигрыш от обработки равен MN .

Подтвердим это утверждение непосредственным расчетом, полагая, что $\Delta_c = \Delta$. Оценивая полосу, занимаемую чипом, как $1/\Delta$, и учитывая, что доступные полоса и временной ресурс определяются, как $W = M/\Delta$ и $T = N\Delta$ соответственно, приходим к результату, что $WT = MN$. Ясно, что в случае АФМ сигналов $M = 1$ и выигрыш от обработки составит $WT = N$.

Задачи.

5.1. Дискретный сигнал длины $N = 5$ имеет комплексные амплитуды вида $a_0 = 1 + j, a_1 = -1 + j, a_2 = 1 + j, a_3 = -1 - j, a_4 = 1 - j$ и частоты $F_i = 0, i = 0, 1, 2, 3, 4$. Определить значения фаз и амплитуд его элементарных символов (чипов) и классифицировать сигнал по способу модуляции.

5.2. Дискретный сигнал задан своими амплитудно-фазовым и частотным кодами вида $a_i = \exp[j\pi i(i+1)/2], F_i = 0, i = \dots, -1, 0, 1, \dots$. Вычислить значения его амплитуд и фаз. Классифицировать сигнал по способу его модуляции. Является ли данный сигнал периодическим? Если да, определить величину периода.

5.3. Доказать четность периодической и аperiodической автокорреляционных функций кодовых последовательностей АФМ сигналов.

5.4. АФМ сигнал состоит из прямоугольных чипов длительности $\Delta_c = \Delta$ и определяется кодовой последовательностью, заданной вектором $\mathbf{a} = (1, 1, 0, 1, 0, 0, -1)$. Вычислить и построить его аperiodическую и периодическую АКФ. Повторить операции для случая $\Delta_c = \Delta/2$.

5.5. Что произойдет с периодической и аperiodической АКФ АФМ сигнала при следующих трансформациях кодовой последовательности:

- циклическом сдвиге элементов;
- изменении знака всех элементов;
- изменении знака только у элементов с четными позициями;
- умножении всех элементов на одну и ту же константу;
- зеркальном отображении последовательности (т.е. считывании справа налево)?

5.6. Возможна ли для ФМ кода комбинация $|R_a(1)| = 3, |R_p(1)| = 1$? Что можно сказать о комбинациях $R_a(1) = -2.1, R_p(1) = 0.8 - j0.6$; $R_a(1) = 0.6 + j0.8, R_p(1) = 1.1 + j$? Каково возможное значение $|R_p(1) - R_a(1)|$ для ФМ кода?

5.7. Расстояния между частотами ЧМ сигнала кратны величине $F = 1/\Delta$. Доказать, что последовательность согласованных по времени прямоугольных импульсов с различными частотами ортогональна.

5.8. ЧМ сигнал длины N содержит $M < N$ частот. Чипы с различными частотами ортогональны. Возможно ли, чтобы АКФ данного сигнала имела нулевые значения при всех ненулевых сдвигах вида $\tau = m\Delta$?

6. Широкополосные сигналы для измерения времени, синхронизации и разрешения во времени.

6.1. Дополнительные требования, предъявляемые к АКФ.

Вернемся к задачам оценивания временной задержки и разрешения во времени, рассмотренным в 2.12, 2.15 и вспомним требования, предъявляемые к сигналу в случае, когда требуются высокая точность измерения и разрешающая способность. В обеих этих задачах необходимо выполнение следующего принципиального условия, заключающегося в высокой концентрации во времени отклика согласованного фильтра на входной сигнал или, что эквивалентно, «острой» АКФ сигнала, и тесно связанной с этим широкой полосы сигнала. Привлекательность распределенности спектра в противовес простому укорачиванию сигнала состоит в том, что при значительном выигрыше от обработки $WT \gg 1$ возможно вложение в сигнал такого количества энергии, которое диктуется требуемой величиной отношения сигнал–шум (SNR), управляя только длительностью, а не пиковой мощностью, которая, как правило, жестко ограничена сверху. Тогда привлечение соответствующей угловой модуляции позволяет расширить полосу сигнала, что обеспечит временную компрессию сигнала согласованным фильтром, так что длительность отклика фильтра (время корреляции $\tau_c \approx 1/W$) окажется во много (примерно WT) раз меньше длительности T сигнала.

Уточним, какого вида АКФ можно считать «острой» или «хорошей», ориентируясь на рассматриваемые задачи приема. Действительно, АКФ (см. определения (2.66) или (2.67)) любого физически реализуемого сигнала не может строго равняться нулю при всех значениях τ , не принадлежащих интервалу $[-\tau_c, \tau_c]$, если время корреляции τ_c меньше длительности сигнала T . Тогда наряду с т. н. *основным лепестком (main lobe)*, или *центральной пику (central peak)*, сосредоточенным в пределах интервала $[-\tau_c, \tau_c]$, АКФ будет обладать также *боковыми лепестками (side lobe)*, находящимися вне указанного интервала (см. рис. 6.1). Наличие боковых лепестков носит преобладающе вредный харак-

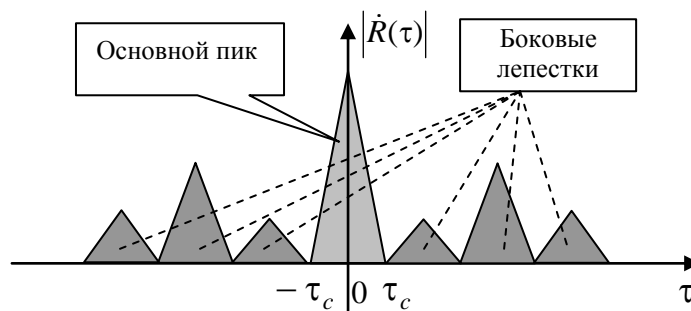


Рис. 6.1. Основной пик и боковые лепестки АКФ.

тер, как при измерении запаздывания, так и разрешении во времени. Действительно, оптимальное (по правилу максимального правдоподобия) измерение задержки сигнала требует фиксации временной позиции максимума огибающей выходного сигнала согласованного фильтра $r_d(t)$ (см. параграф 2.12), а в отсутствии шума отклик согласованного фильтра–детектора на входной сигнал в точности представляет собой АКФ огибающей. В реальной ситуации, когда сигнал наблюдается на фоне шума, всегда имеется вероятность возникновения ложного максимума вне «тела» основного пика АКФ, превышающего по величине истинный (т.е. расположенный в пределах «тела»), что изображено пунктирной линией на рис.6.2. В подобной ситуации возникает аномальная ошибка оценивания, при

которой отклонение ε оценки $\hat{\tau}$ от истинного значения τ превосходит τ_c . Ясно, что перепутывание основного лепестка с ложным пиком, возникающим поблизости от значительного бокового лепестка, более вероятно, чем с ложным пиком, возникающим «на пустом месте», поскольку для гауссовского шума «легче» поднять боковой лепесток до высоты основного, чем создать его с нулевого уровня.

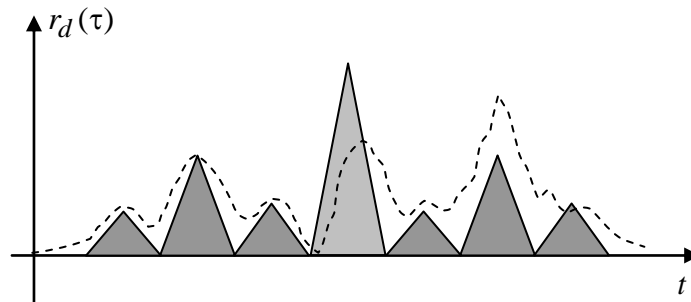


Рис. 6.2. Иллюстрация возникновения аномальной ошибки.

Для иллюстрации вредного влияния боковых лепестков на процедуру временного разрешения рассмотрим суперпозицию двух копий полосного сигнала, которые характеризуется разными амплитудами и временными сдвигами, как это показано на рис. 6.3, *a*. После обработки согласованным фильтром основной лепесток более слабой копии оказывается полностью замаскированным боковым лепестком более сильной копии (см. рис. 6.3, *b*). В этих условиях наблюдатель не сможет надежно извлечь необходимую информацию из обеих сигнальных копий или даже определить число принятых копий. Ситуация подобного рода типична для случая неразрешимых сигналов, несмотря на то, что основной лепесток АКФ значительно уже длительности сигнала.

В наиболее общих терминах требования, предъявляемые широкополосным сигналам при решении задач измерения запаздывания и разрешения во времени, теперь могут быть сформулированы следующим образом: *АКФ сигнала должна иметь достаточно острый центральный пик и по возможности наиболее низкий уровень боковых лепестков*. В следующих параграфах этой главы будут рассмотрены пути и инструменты достижения этой поставленной фундаментальной задачи.

6.2. Сигналы с непрерывной частотной модуляцией.

Исторически к числу первых открытых сигналов, обладающих эффектом временной компрессии согласованным фильтром, относится импульсный сигнал с линейной частотной модуляцией (ЛЧМ). Как следует из названия, мгновенная частота этого сигнала линейно изменяется в течение его длительности. Рассмотрим полосный сигнал, мгновенная частота $f(t)$ которого возрастает во времени по закону

$$f(t) = f_0 + \frac{W_d t}{T}, \quad |t| \leq \frac{T}{2},$$

где W_d — девиация частоты, т.е. полный диапазон изменения частоты, а f_0 — как обычно, центральная частота. Полная мгновенная фаза $\Phi(t)$ сигнала определяется как интеграл от мгновенной частоты и, следовательно, фаза ЛЧМ импульса подчиняется квадратичному закону

$$\Phi(t) = 2\pi \int_0^t f(u) du = 2\pi f_0 t + \frac{\pi W_d t^2}{T}, \quad |t| \leq \frac{T}{2}.$$

Полагая вещественную огибающую прямоугольной, комплексная огибающая ЛЧМ сигнала принимает вид

$$\dot{s}(t) = \begin{cases} \exp\left(\frac{j\pi W_d t^2}{T}\right), & |t| \leq \frac{T}{2}, \\ 0, & |t| \geq \frac{T}{2}. \end{cases}$$

Подставляя это соотношение в общую формулу (2.66) для АКФ, последняя может быть вычислена формально без особых трудностей. Однако существует менее формальный и более прозрачный с физической точки зрения способ, приводящий к тому же результату.

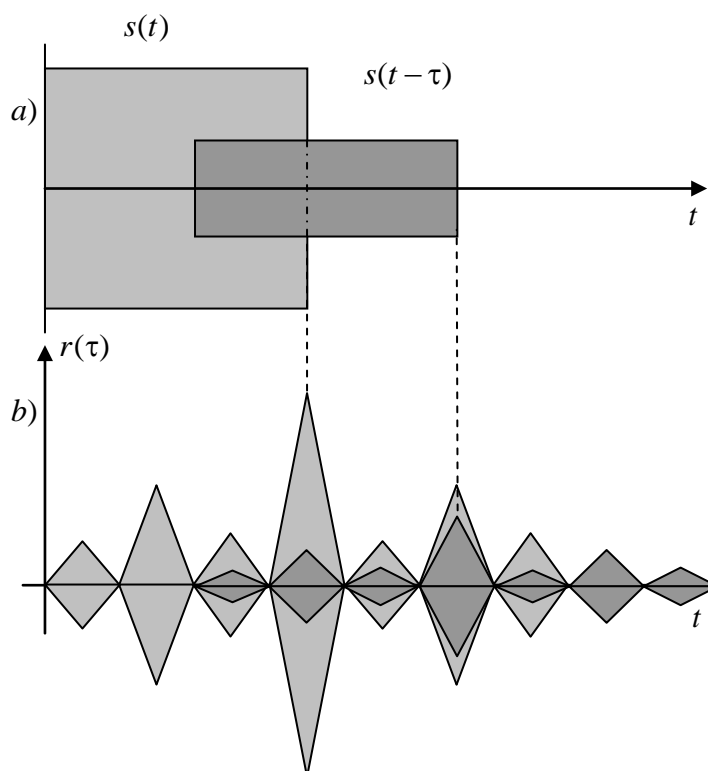


Рис. 6.3. Иллюстрация отсутствия разрешения сигналов.

Из теории частотной модуляции хорошо известно, что если индекс модуляции $\beta = W_d T$ значителен ($W_d t \gg 1$), то спектр частотно-модулированного колебания содержит компоненты всех мгновенных частот, причем форма спектра описывается вещественной огибающей сигнала. Таким образом, в рассматриваемом случае спектр располагается в диапазоне $[f_0 - \frac{W_d}{2}, f_0 + \frac{W_d}{2}]$ и имеет форму, близкую к прямоугольной (см. рис. 6.4, *a*). Теперь АКФ (2.66) может быть найдена в результате обратного преобразования Фурье, подобно тому, как это уже использовалось в 2.12.2. Поскольку энергетический спектр оказывается прямоугольным, то его обратное преобразование Фурье даст функцию вида $\sin x/x$, так что нормированная АКФ ЛЧМ сигнала представима как

$$\dot{\rho}(\tau) \approx \frac{\sin(\pi W_d \tau)}{\pi W_d \tau}, \quad (6.1)$$

что и демонстрирует рис. 6.4, *b*.

Как следует из рис. 6.4, полная (т.е. измеренная между двумя ближайшими к нача-

лу координат нулями) ширина основного лепестка АКФ составляет $2\tau_c = 2/W_d$. Тогда, полагая ширину основного лепестка на некотором ненулевом уровне равной $\tau_c = 1/W_d$, получаем, что согласованный фильтр осуществляет временную компрессию ЛЧМ сигнала в $T/\tau_c \approx W_d T \approx WT$ раз.

Существенным недостатком ЛЧМ сигнала является высокий уровень боковых лепестков АКФ. Ближайший к началу координат боковой лепесток имеет уровень, равный $2/3\pi \approx 0.21(-13\text{дБ})$ по отношению к основному, и не зависит от выигрыша от обработки

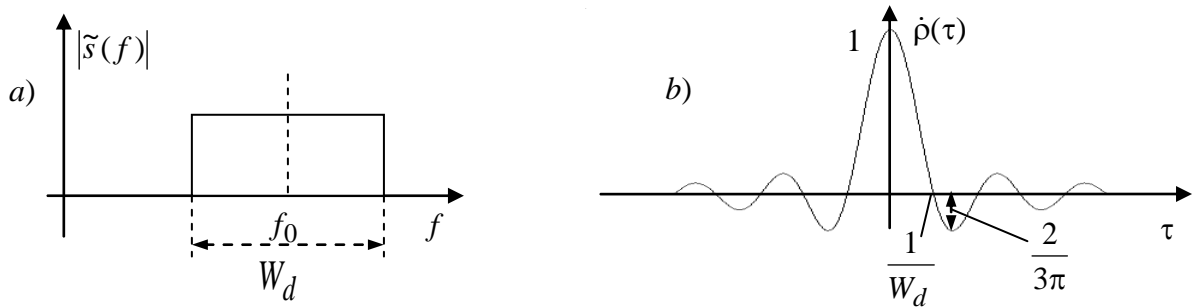


Рис. 6.4. Аппроксимация спектра и АКФ ЛЧМ импульса.

WT , т.е. он не может быть уменьшен за счет увеличения девиации W_d . Известны эффективные методы снижения уровня боковых лепестков путем сглаживания огибающей сигнала либо использованием взвешивания, либо рассогласованной обработкой в приемнике. Однако выигрыш в уровне боковых лепестков, достигаемый данными методами, обеспечивается или за счет расширения основного пика, и/или потерями в отношении сигнал-шум.

Пример 6.2.1. Рассмотрим прямоугольный ЛЧМ импульс с девиацией $W_d = 20/T$. На рис. 6.5 представлены собственно сигнал и отклик на него согласованного фильтра, полученные с использованием программы, разработанной в среде Matlab. Сравнение ве-

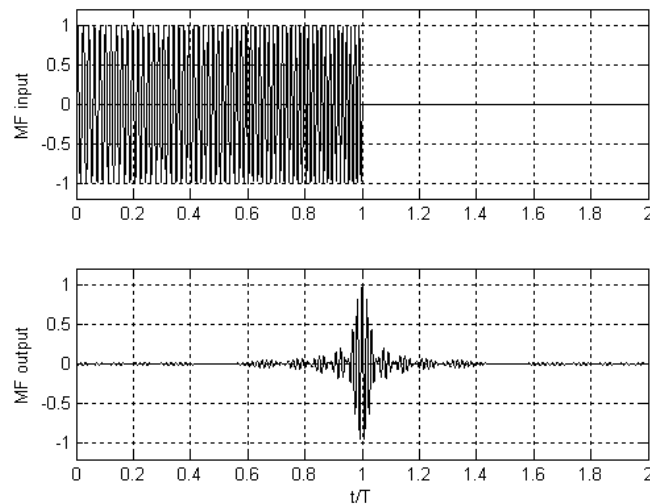


Рис. 6.5 Временная компрессия прямоугольного ЛЧМ импульса.

личины временной компрессии и уровня первого бокового лепестка с теоретически предсказанными значениями показывают их практическое совпадение. \square

Другим недостатком ЛЧМ сигнала является гребенчатая форма функции неопределенности $\rho_0(\tau, F)$. Из материала, изложенного в 2.14-2.15, можно сделать вывод о том,

что для одновременного измерения временной задержки и частоты точно также, как и для частотно–временного разрешения наилучшей является игольчатая функция неопределенности, обладающая одним центральным пиком в начале координат и резко спадающая во всех направлениях частотно-временной плоскости. Как следует из рис. 6.6, на котором изображены функция неопределенности (а) и диаграмма неопределенности (б) ЛЧМ сигнала, последний не может служить эффективным инструментом для решения упомянутых выше задач. Если пара значений τ, F попадает в эллипс, изображенный на рис. 6.6, б, точность оценивания этих параметров становится очень низкой, поскольку копии сигнала со всеми подобными парами значений сильно коррелированы, т.е. с трудом поддаются различению. Аналогичный вывод можно сделать о разрешении таких сдвинутых по времени и частоте копий сигнала: высокая схожесть делает задачу их разделения весьма проблематичной.

Известны несколько способов улучшения формы функции неопределенности частотно–модулированных сигналов, например, применение симметричной (V-типа) линейной модуляции.

До сих пор непрерывные ЛЧМ сигнала и их модификации находят широкое приме-

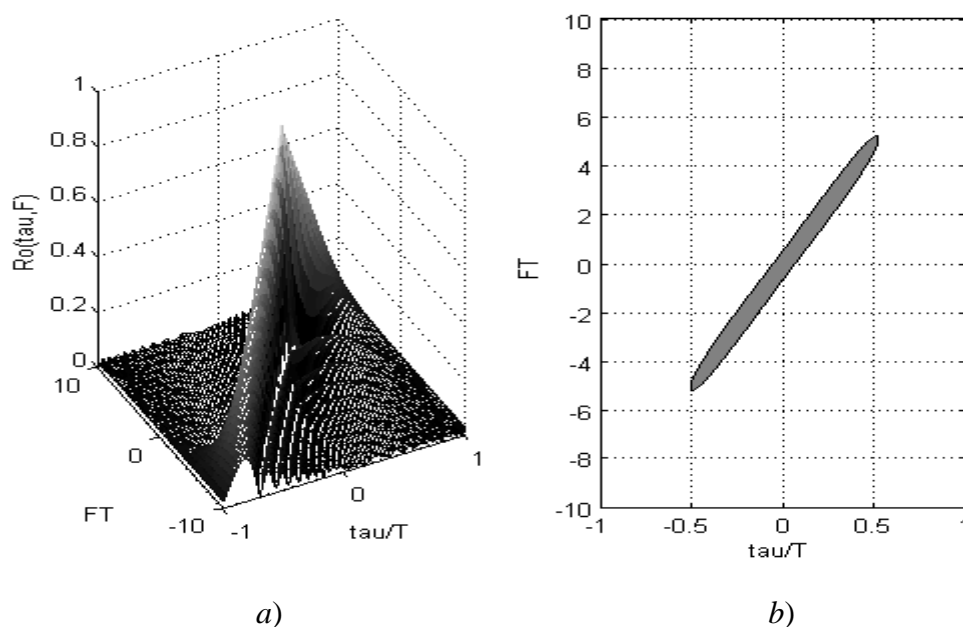


Рис. 6.6 Функция неопределенности (а) и ее горизонтальное сечение (б) ЛЧМ сигнала.

нение в разнообразных широкополосных радарных и сонарных системах, однако в современных коммерческих коммуникационных системах или общедоступных системах дальней навигации они не нашли широкого применения, уступая указанное пространство дискретным сигналам. Одной из причин подобного предпочтения в дополнении к тому, о чем уже упоминалось, может служить лучшая совместимость дискретных сигналов с современными цифровыми технологиями, микроэлектроникой и программно ориентированной философией современных радиосистем.

6.3. Критерий выбора АФМ сигналов с хорошей аперидической АКФ.

Обратимся вновь к соотношению (5.5) и вспомним, что АКФ $\dot{\rho}(\tau)$ АФМ сигнала представляет собой АФМ сигнал, у которого в качестве чипов выступает АКФ $\dot{\rho}_c(\tau)$ исходного элементарного импульса, а роль кодовой последовательности выполняет АКФ

$\rho(m)$ исходного кода a_0, a_1, \dots, a_{N-1} . Подобная конструкция явным образом свидетельствует, что при заданном чипе профиль полной АКФ полностью определяется АКФ $\rho(m)$ кода. В частности, если длительность чипа не превышает периода их повторения ($\Delta_c \leq \Delta$), то «высота» $|\dot{\rho}(m\Delta)|$ любого бокового лепестка в точке $\tau = m\Delta$ просто повторяет величину АКФ кода $|\rho(m)|$ при сдвиге на m позиций. Из установленного в 6.1 факта следует, что минимизация уровня боковых лепестков АКФ имеет наивысший приоритет при конструировании сигнала всякий раз, когда одним из предназначений системы является измерение времени запаздывания или временное разрешение. Конечно, было бы замечательно иметь все боковые лепестки равными нулю, однако это абсолютно невозможно для финитных или аперiodических АФМ сигналов. Действительно, рассмотрим некоторый сигнал конечной длины N , что влечет выполнение неравенств $a_0 \neq 0$ и $a_{N-1} \neq 0$, поскольку в противном случае длина сигнала была бы меньше N . Тогда крайний правый боковой лепесток нормированной аперiodической АКФ кода (5.10) подобного сигнала будет

$$\rho_a(N-1) = \frac{a_0 a_{N-1}^*}{\|\mathbf{a}\|^2} \neq 0. \quad (6.2)$$

Последнее соотношение немедленно ведет к применению *минимаксного критерия* при синтезе сигналов, который требует достижения минимально возможной величины максимального бокового лепестка АКФ аперiodического кода. Формальная запись данного критерия имеет следующий вид

$$\rho_{a, \max} = \max_{m \neq 0} \{|\rho_a(m)|\} = \min. \quad (6.3)$$

В свете критерия (6.3) предпочтительными являются кодовые последовательности с наименьшим значением максимального бокового лепестка, однако это требование всегда сопровождается ограничением на метод модуляции или, более конкретно, на алфавит, которому принадлежат символы кодовой последовательности. Это ограничение отражает технологические аспекты, касающиеся сложности формирования и обработки сигнала, и, как вскоре станет явным, может оказаться очень обязывающим. Таким образом требования, предъявляемые к наилучшему сигналу, могут быть сформулированы в виде следующей оптимизационной задачи: *на множестве всех возможных последовательностей длины N с символами из заранее выбранного алфавита найти последовательность или последовательности с минимальной величиной максимального бокового лепестка аперiodической АКФ.*

6.4. Об оптимизации аперiodических АФМ сигналов.

Сформулированная выше оптимизационная задача, как и многие другие задачи дискретной оптимизации, не имеют некоторого общего аналитического решения, и типичной процедурой ее выполнения является осуществление исчерпывающего поиска. Ограничимся классом ФМ сигналов, общепризнанных наиболее привлекательными. В качестве объяснения причин подобного утверждения достаточно напомнить, что в задачах временного измерения и разрешения во времени основное достоинство широкополосности состоит в возможности распределения энергии сигнала на значительном временном интервале, тем самым, снижая пиковую мощность. Сигналы с ФМ, свободные от амплитудной модуляции, представляют собой предельную версию такого расширения, позволяющую получить пик-фактор ν сигнала (отношение пиковой к средней мощности), равный единице.

Для любого ФМ сигнала $|a_i| = 1, i = 0, 1, \dots, N-1$, так что $|a_0 a_{N-1}| = 1$, и крайний пра-

вый боковой лепесток аperiodической АКФ (6.2) $|\rho_a(N-1)| = 1/N$. Следовательно, максимальный боковой лепесток ФМ сигнала ограничен снизу величиной

$$\rho_{a,\max} \geq 1/N. \quad (6.4)$$

Естественно, что ФМ сигналы, удовлетворяющие данной границе, будут оптимальными. В честь ученого, первым предпринявшего поиск подобных сигналов, они получили название *кодов Баркера*. Действительно, Баркеру удалось найти оптимальные бинарные коды, лежащие на границе (6.4). Традиционно бинарные последовательности с символами ± 1 считаются наиболее привлекательными, поскольку в наибольшей степени отвечают цифровой схемотехнике и гарантируют наименьшую сложность в их формировании и обработки. В таблице 6.1 представлены все бинарные коды Баркера. Некоторые из кодов не являются уникальными в том смысле, что существуют другие последовательности той же длины, удовлетворяющие нижней границе (6.4).

Таблица 6.1. Бинарные коды Баркера.

N	Код												
2	+	-											
3	+	+	-										
4	+	+	-	+									
5	+	+	+	-	+								
7	+	+	+	-	-	+	-						
11	+	-	+	+	-	+	+	+	-	-	-		
13	+	+	+	+	+	-	-	+	+	-	+	-	+

Пример 6.4.1. Таблица 6.2 иллюстрирует вычисление аperiodической и периодической АКФ кода Баркера длины $N = 7$. Структура таблицы полностью аналогична таблице 5.1. Как видно, не только аperiodическая, но и нормированная периодическая АКФ характеризуется максимальным боковым лепестком, равным $1/N$. \square

Таблица 6.2. Вычисление АКФ бинарного кода Баркера длины 7.

m	a_0	a_1	a_2	a_3	a_4	a_5	a_6	$R_a(m)$	$R_p(m)$
0	+	+	+	-	-	+	-	+7	+7
1	-	+	+	+	-	-	+	0	-1
2	+	-	+	+	+	-	-	-1	-1
3	-	+	-	+	+	+	-	0	-1
4	-	-	+	-	+	+	+	-1	-1
5	+	-	-	+	-	+	+	-0	-1
6	+	+	-	-	+	-	+	-1	-1

Рассмотрим кратко процедуру согласованной фильтрации сигнала Баркера на некотором конкретном примере.

Пример 6.4.2. На рис. 6.7 изображена структура согласованного фильтра для сигнала Баркера длины $N = 7$. Первым блоком данной структуры является линия задержки с отводами, разделенными во времени на величину, равную периоду следования чипов Δ . Выходные сигналы отводов линии задержки подаются на сумматор с весами, определяемыми как зеркальное отражение символов кода. Второй частью структуры служит фильтр, согласованный с одиночным чипом (СФОИ). Чтобы убедиться в том, что указанная схема является искомым согласованным фильтром, достаточно осуществить проверку, подав на

ее вход дельта-импульс. Выходной отклик представит собой зеркальное отображение сигнала, что в точности соответствует реакции согласованного фильтра.

На рис.6.8 представлены диаграммы, пронумерованные в соответствии с точками

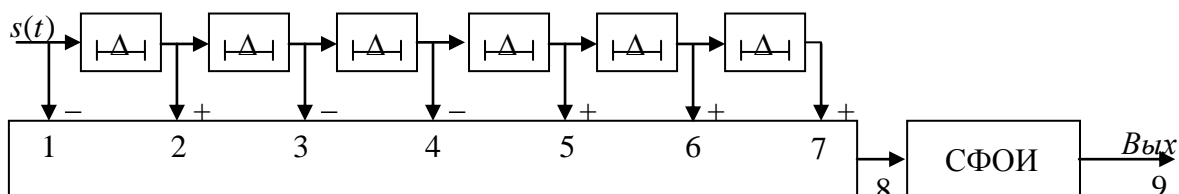


Рис. 6.7. Согласованный фильтр для бинарного сигнала Баркера длины $N=7$.

схемы, изображенной на рис. 6.7, которые иллюстрируют в деталях согласованную фильтрацию видеосигнала Баркера, составленного из чипов с прямоугольной огибающей. Когда последний чип сигнала поступит на вход фильтра, все предшествующие чипы появятся на входах сумматора с соответственно измененными полярностями и суммирование осуществляется в фазе, обеспечивая получение основного пика АКФ. До этого момента и после получаем боковые лепестки, имеющие полярность и уровень относительно основного в соответствии с результатами таблицы 6.2. □

К сожалению, бинарные коды Баркера существуют только для длин, приведенных в таблице 6.1. Еще в 60-х годах Турин и Сторер доказали их не существование для любых других нечетных длин и для четных, по крайней мере из диапазона $4 < N < 12100$ ¹. Были предприняты достаточно обширные исследования по поиску небинарных ФМ кодов Баркера с эквидистантным фазовым алфавитом (многофазные или M -ичные ФМ коды), однако достигнутые результаты далеки от обнадеживающих. Было обнаружено, что даже скромные успехи в увеличении длины достигаются только за счет значительного увеличения размера M фазового алфавита. Возможно, что известные к настоящему времени многофазные коды Баркера наибольшей длины и найденные путем стохастического поиска характеризуются объемами фазового алфавита от сотен до десятков тысяч ($N = 32 \dots 36$) [23], или 60, 90, 120 ($N = 37 \dots 45$) [24]. Большой размер алфавита неизбежно влечет серьезные затруднения при аппаратной реализации, а также значительно более жесткие и трудновыполнимые требования к допустимым ошибкам воспроизведения, дрейфу параметров и т.п.

Как видно, длины существующих кодов Баркера достаточно малы для удовлетворения многочисленных практических нужд, что в значительной степени стимулирует поиск бинарных последовательностей большей длины с уровнем боковых лепестков, превышающих нижнюю границу (6.4). Поскольку ненормированная АКФ (5.9) любой бинарной последовательности всегда определяется суммой ± 1 , то возможные значения $\rho_{a, \max}$ для не баркерских кодов принадлежат множеству $2/N, 3/N, \dots$. Гарантированное нахождение глобально оптимального (т.е. имеющего минимально возможное $\rho_{a, \max} > 2/N$ при заданном N) бинарного кода может быть осуществлено только путем полного перебора возможных комбинаций. К сожалению, вычислительный объем, необходимый для подобной оптимизации, экспоненциально возрастает с увеличением длины N и становится не реализуемым при величинах N , превышающих 50. По крайней мере, как известно автору, глобально оптимальные бинарные коды, открытые к настоящему времени, охватывают диапазон вплоть до 50 [25, 26].

¹ В соответствии с результатами [22] этот диапазон увеличен теперь до 1 898 884 и маловероятно, что бинарные коды Баркера четной длины существуют вне указанного диапазона.

Если основываться на факте, что нахождение глобально оптимальных бинарных последовательностей большой длины практически не реализуемо, задача (6.3) может быть сформулирована в более ограниченном виде: *найти бинарный код с удовлетворительно*

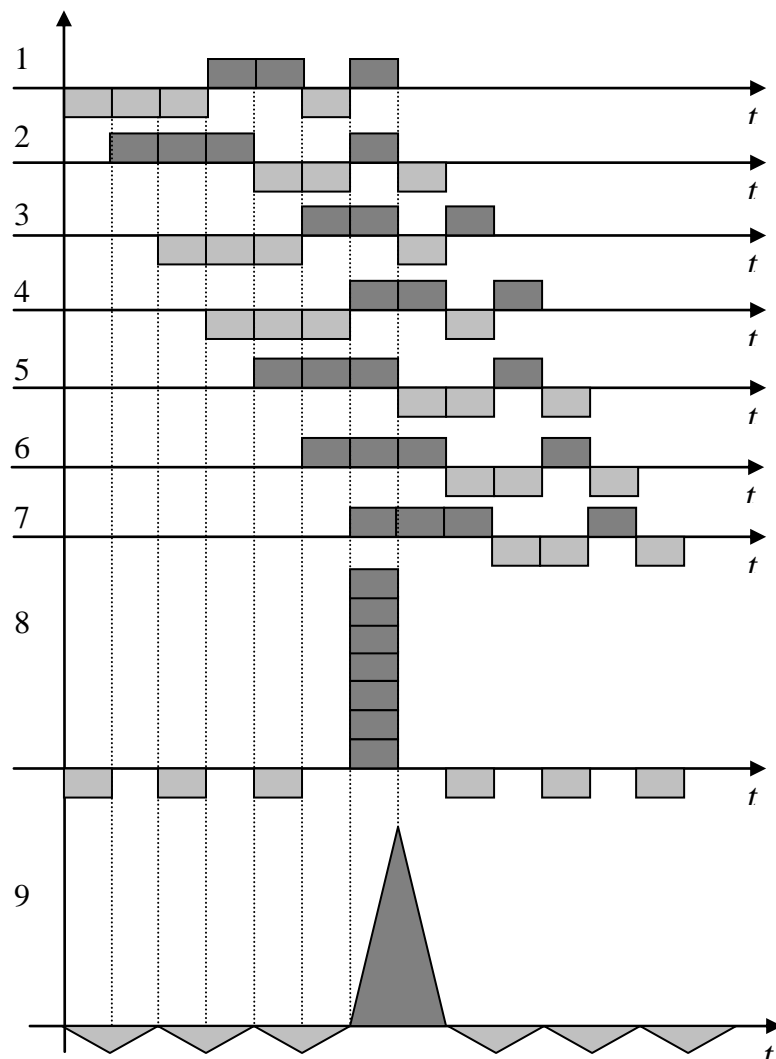


Рис.6.8 Согласованная фильтрация бинарного сигнала Баркера длины $N=7$.

малым уровнем аperiodического бокового лепестка $\rho_{a, \max}$ без гарантии глобальной оптимальности. Общая идея алгоритмов, направленных на решение этой задачи, состоит в предварительном отборе некоторого ограниченного множества последовательностей, которое кажется многообещающим в плане корреляционных свойств, и последующем скрупулезном поиске кода с минимальным значением $\rho_{a, \max}$ только среди последовательностей, вошедших в указанное множество. Одним из примеров подобной стратегии является эволюционный алгоритм [27], посредством которого были найдены бинарные коды длин до ста включительно, некоторые из которых незначительно лучше кодов Баркера. Так, для кода Баркера длины 13 $\rho_{a, \max} = 1/13 \approx 0.077$, а для лучшего в смысле $\rho_{a, \max}$ кода из [27] $\rho_{a, \max} = 6/88 \approx 0.068$. Другой продуктивный подход основан на соотношении (5.13), связывающем аperiodическую АКФ со своим периодическим аналогом. Обозначая через $\rho_{p, \max}$, максимальный боковой лепесток периодической АКФ

$$\rho_{p, \max} = \max_{m=1,2,\dots,N-1} \{|\rho_p(m)|\},$$

и используя неравенство $\max\{|x + y|\} \leq \max\{|x| + |y|\} \leq \max\{|x|\} + \max\{|y|\}$, приходим к оценке $\rho_{p, \max} \leq 2\rho_{a, \max}$ или

$$\rho_{a, \max} \geq \frac{1}{2} \rho_{p, \max} . \quad (6.5)$$

Значение этого соотношения совершенно прозрачно: необходимым условием «хорошей» аперiodической АКФ является существование хорошей (имеющей низкий уровень максимального бокового лепестка $\rho_{p, \max}$) периодической АКФ. Другими словами, последовательности с хорошей аперiodической АКФ могут быть найдены только среди последовательностей с хорошей периодической АКФ. Как будет показано позднее, существуют достаточно эффективные аналитические инструменты построения последовательностей с хорошими периодическими АКФ. Следовательно, можно построить некоторое множество последовательностей с хорошими периодическими АКФ, которое послужит исходным материалом для поиска в его рамках последовательностей с хорошей аперiodической АКФ. Эта возможность подчеркивает ведущую роль периодической АКФ в задаче синтеза последовательностей с требуемыми корреляционными свойствами и объясняет причину, по которой в следующем параграфе основное внимание уделяется исследованию периодических АКФ.

6.5. Идеальная периодическая АКФ. Бинарные минимаксные последовательности.

Мотивация интереса к последовательностям с хорошей периодической АКФ не ограничивается только их ролью исходного материала для построения хороших аperiodических последовательностей. Существует множество приложений, основанных на использовании периодических дискретных сигналов (CW-локация, навигация, пилотный канал и канал синхронизации в мобильных системах радиосвязи и т.п.), что предопределяет чрезвычайную важность периодической АКФ в отношении системных характеристик. Будем считать «идеальной» такую периодическую АКФ, которая обладает нулевыми боковыми лепестками, т.е. нулевыми значениями между периодическими основными лепестками, повторяющимися с периодом N . Используя нормированную версию АКФ, запишем указанное условие в виде

$$\rho_p(m) = \frac{1}{E} \sum_{i=0}^{N-1} a_i a_{i-m}^* = \begin{cases} 1, & m = 0 \bmod N, \\ 0, & m \neq 0 \bmod N, \end{cases} \quad (6.6)$$

где запись $m = 0 \bmod N$ читается, как m делится на N (или кратно N). Очевидно, что для идеальной АКФ $\rho_{p,\max} = 0$. На рис. 6.9 изображена АКФ дискретного видеосигнала с прямоугольными чипами, манипулированного кодом с идеальной АКФ. Практические

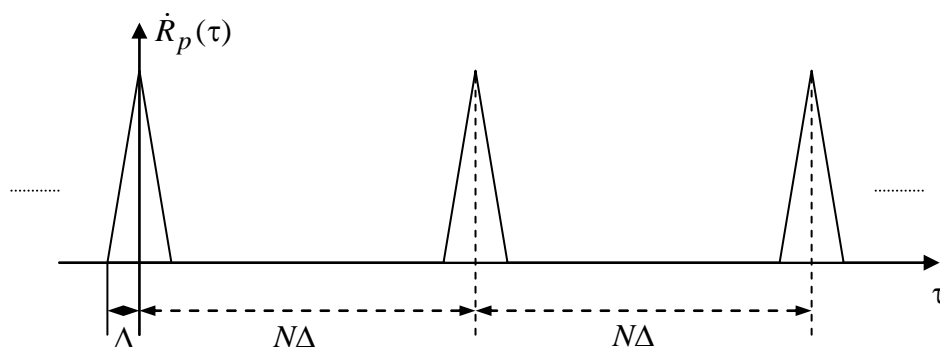


Рис. 6.9. Идеальная периодическая АКФ.

достоинства идеальной АКФ явным образом иллюстрирует рис. 6.10, на котором представлены колебания a) и b) двух сдвинутых во времени копий одного и того же полосного периодического сигнала, АКФ кода которого удовлетворяет соотношению (6.6). При появлении на входе фильтра, согласованного с однопериодным сегментом сигнала, суперпозиции двух сдвинутых во времени копий на его выходе будет наблюдаться две сдвинутые во времени копии АКФ сигнала. Если временная задержка между сигнальными копиями больше длительности основного лепестка АКФ 2Δ (но меньше $(N-2)\Delta$), то отклики фильтра на оба сигнала полностью разрешимы без взаимных помех (рис. 6.10, c).

Проанализируем ненормированную периодическую АКФ бинарной последовательности

$$R_p(m) = \sum_{i=0}^{N-1} a_i a_{i-m}, \quad (6.7)$$

в которой знак комплексного сопряжения опущен за ненадобностью, поскольку все $a_i = \pm 1$. Просуммировав обе части соотношения (6.7) по возможным значениям $m = 0, 1, 2, \dots, N-1$, получаем

$$\sum_{m=0}^{N-1} R_p(m) = \sum_{m=0}^{N-1} \sum_{i=0}^{N-1} a_i a_{i-m} = \sum_{i=0}^{N-1} a_i \sum_{m=0}^{N-1} a_{i-m} = |\tilde{a}_0|^2, \quad (6.8)$$

где

$$\tilde{a}_0 = \sum_{i=0}^{N-1} a_i -$$

постоянная составляющая кодовой последовательности $\{a_0, a_1, \dots, a_{N-1}\}$. Поскольку постоянная составляющая бинарной последовательности может принимать только целые значения, то сумма в (6.8) есть квадрат целого. Предположим теперь, что бинарный код

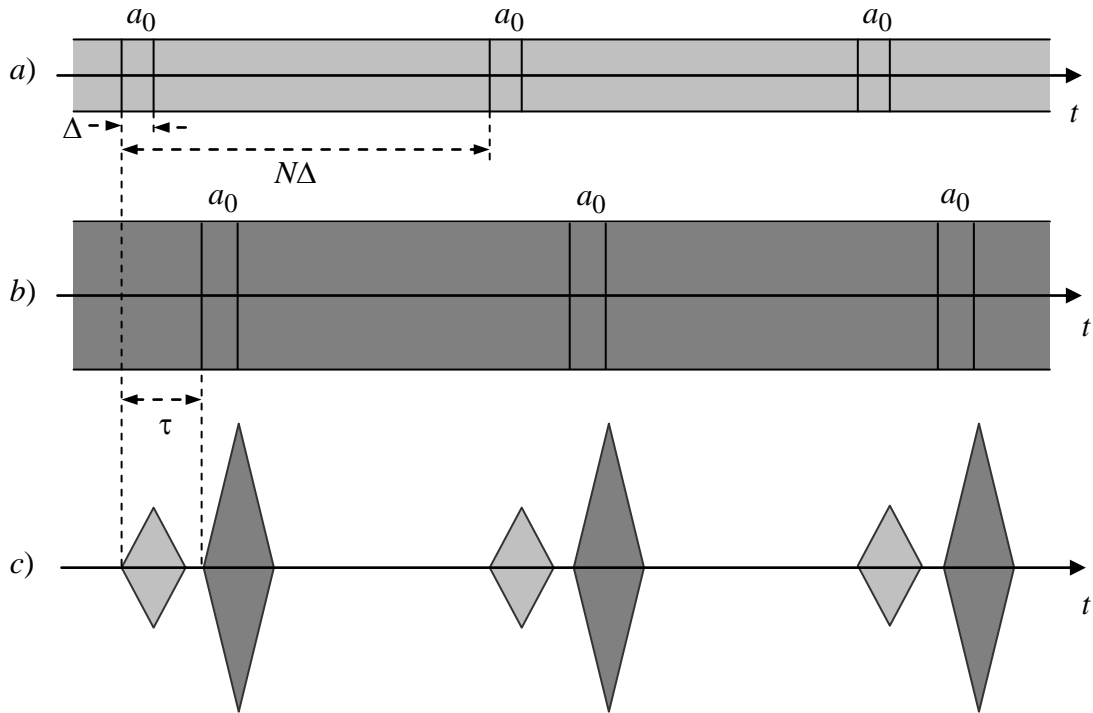


Рис. 6.10. Разрешение копий сигнала с идеальной периодической АКФ.

обладает идеальной периодической АКФ. Тогда $R_p(0) = \sum_{i=0}^{N-1} (a_i)^2 = N$ и

$R_p(m) = 0, 1, \dots, N-1$, что дает

$$\sum_{m=0}^{N-1} R_p(m) = N = |\tilde{a}_0|^2. \quad (6.9)$$

Полагая $m \neq 0 \pmod N$, пусть N_e и N_d число произведений $a_i a_{i-m}$ в сумме (6.7), равных +1 и -1 соответственно. Тогда $R_p(m) = N_e - N_d = 0$ означает равенство $N_e = N_d$ и $N = N_e + N_d = 2N_e$. Таким образом, согласно (6.9) и последнему результату, длина N представляет собой четный квадрат целого числа, т.е. необходимое условие получения идеальной АКФ для бинарной последовательности может быть записано как $N = 4h^2$, где h – целое. В начале 60-х годов все подобные длины (4, 16, 36, 64, ... и т.д.) были проанализированы в работе Турина, который показал, что до длин $N \leq 12100$ единственным би-

нарным кодом¹ с идеальной периодической АКФ (ПАКФ) является тривиальный код длины 4 вида: +1 +1 +1 -1 [28]. Позднее не существование подобных последовательностей было доказано до длин $N < 4 \times 165^2 = 108900$ [29]. Представляется совершенно невероятным их существование за пределами указанного диапазона.

В свете приведенных данных вызывает несомненный интерес определение потенциала минимизации максимального бокового лепестка периодической АКФ бинарных кодов. Пусть N_{++} , N_{+-} , N_{-+} обозначают число пар $a_i a_{i-m}$ в соотношении (6.7), значения которых совпадают с подстрочными индексами, например N_{+-} – число пар, в которых $a_i = +1$, $a_{i-m} = -1$. Поскольку как $N_{++} + N_{+-}$, так и $N_{++} + N_{-+}$ дают один и тот же результат – общее число положительных элементов на периоде кода, то $N_{+-} = N_{-+}$. Тогда разность

$$N - R_p(m) = \sum_{i=0}^{N-1} (1 - a_i a_{i-m}) = 2(N_{+-} + N_{-+}) = 4N_{+-}$$

всегда делится без остатка на четыре. Поэтому очевидно, что для любого бинарного кода ненормированная периодическая АКФ всегда отличается от длины N некоторым множителем на четыре, т.е.

$$R_p(m) = N - 4h, \quad (6.10)$$

где h – целое.

Очевидно, что в отсутствие бинарных кодов с идеальной периодической АКФ следующими по привлекательности являются бинарные последовательности, для которых $R_p(m)$ принимает значения ± 1 при $m = 1, 2, \dots, N-1$, т.е. обладают $\rho_{p, \max} = 1/N$. Как следует из (6.10), значение $R_p(m) = +1$ возможно только при длине $N = 4h + 1$, тогда как $R_p(m) = -1$ может иметь место только при длине $N = 4h - 1$, где h – целое. Последнее свидетельствует, что бинарные последовательности с $\rho_{p, \max} = 1/N$ могут иметь только два возможных значения ненормированной периодической АКФ: либо

$$R_p(m) = \begin{cases} N, & m = 0 \bmod N, \\ +1, & m \neq 0 \bmod N, \end{cases} \quad (6.11)$$

при длине $N = 4h + 1$, либо

$$R_p(m) = \begin{cases} N, & m = 0 \bmod N, \\ -1, & m \neq 0 \bmod N, \end{cases} \quad (6.12)$$

при длине $N = 4h - 1$.

Последовательности, удовлетворяющие соотношениям (6.11)–(6.12) и, следовательно, обладающие теоретически минимальным уровнем боковых лепестков периодической АКФ ($\rho_{p, \max} = 1/N$) для бинарных кодов нечетной длины, называются *минимаксными*. Известны только два примера ($N = 5$ и $N = 13$) последовательностей, подчиняющихся соотношению (6.11), тогда как существуют чрезвычайно мощные регулярные правила формирования минимаксных последовательностей, удовлетворяющие (6.12). Рассмотрению двух наиболее популярных типов посвящены параграфы 6.6–6.9, хотя, по меньшей мере, известны еще три варианта минимаксных последовательностей.

¹ Здесь не считаются новыми (и это принято повсеместно) последовательности, получаемые из исходных циклическим сдвигом, зеркальным отображением или изменением знака всех элементов. Указанные преобразования не изменяют свойств периодической АКФ (см. задачу 5.5) и последовательности, получаемые одна из другой подобным путем, рассматриваются как тривиально различные или эквивалентные.

6.6. Введение в теорию конечных полей и линейных последовательностей.

6.6.1. Определение конечного поля.

Для осознанного восприятия способов конструирования бинарных минимаксных последовательностей необходимо хотя бы кратко ознакомиться с основными понятиями конечных полей. Предлагаемое описание будет менее формальным, чем это дается в истинно математических источниках. Присвоим наименование *поля* множеству элементов, на котором определены две операции, называемые *сложением* и *умножением* и обозначаемые обычными символами «+» и « \cdot » (или « \times », или обычной записью одного элемента за другим). Термин «определены» означает, что обе эти операции являются замкнутыми, т.е., если x, y являются элементами поля F , то их сумма и произведение также принадлежит F : $x + y \in F$, $xy \in F$. В любом поле имеются нулевой «0» и единичный «1» элементы, которые не изменяют значения произвольного элемента $x \in F$ соответственно в операциях сложения и умножения: $x + 0 = x$, $x \cdot 1 = x$. Таблицы сложения и умножения строятся таким образом, чтобы указанные операции были коммутативны ($x + y = y + x$; $xy = yx$), ассоциативны ($(x + y) + z = x + (y + z)$; $x(yz) = (xy)z$) и обратимыми, т.е. были бы также определены операции вычитания и деления на ненулевой элемент: $x + y = z \Rightarrow x = z - y$; $xy = z$, $y \neq 0 \Rightarrow x = z / y$. Последнее, в частности, влечет существование элементов *противоположного* x , обозначаемого как $-x = 0 - x$, и *обратного* ненулевому x , обозначаемого как $x^{-1} = 1/x$. И, наконец, таблицы операций подчиняются дистрибутивному закону: $(x + y)z = xz + yz$.

Очевидно, что поле представляет собой множество, в рамках которого операции с элементами осуществляются аналогично тому, как это производится с вещественными числами в обычной арифметике. Следовательно, поле есть просто абстрактное обобщение множества вещественных чисел или, другими словами, множество вещественных чисел есть тривиальный пример поля. Другими примерами поля служат множества рациональных чисел, комплексных чисел и т.п. Все упомянутые поля обладают бесконечным *порядком*, где последний термин означает число элементов поля. В противоположность этому, все нижеприведенные конструкции оперируют с *конечными* или Галуа полями, порядки которых конечны. В современной алгебре доказано (см., например [30]), что существуют конечные поля любого (и только) порядка вида p^m , где p – простое, m – натуральное числа. Стандартным обозначением поля Галуа порядка p^m служит $GF(p^m)$. Наше рассмотрение ограничится только *простыми* полями $GF(p)$, порядки которых есть простые числа ($m = 1$). Простейшим способом трактовки простого поля $GF(p)$ служит представление всех его элементов как p целых чисел вида $0, 1, \dots, p-1$, сложение и умножение которых осуществляется по модулю p . На рис. 6.11 приведены таблицы сложения и умножения для трех простейших простых полей $GF(2)$, $GF(3)$, $GF(5)$. Отметим, что противоположный любому из двух элементов поля $GF(2)$ есть тот же самый элемент, поскольку $0 + 0 = 1 + 1 = 0$, а единственный ненулевой элемент 1 является обратным самому себе. Операции во всех остальных простых полях не являются настолько вырожденными, например в $GF(5)$, $2 + 3 = 0 \Rightarrow -2 = 3$, и $3 \cdot 2 = 1 \Rightarrow 3^{-1} = 2$.

6.6.2. Линейные последовательности над конечными полями.

Введем теперь в рассмотрение последовательность d_0, d_1, \dots с элементами (симво-

лами) из заданного конечного поля $GF(p)$, которые подчиняются линейной рекурсии вида

$$d_i = -f_{n-1}d_{i-1} - f_{n-2}d_{i-2} - \dots - f_0d_{i-n}, \quad i = n, n+1, \dots, \quad (6.13)$$

где коэффициенты f_0, f_1, \dots, f_{n-1} – фиксированные константы, принадлежащие $GF(p)$. Подобная последовательность называется *линейной рекуррентной последовательностью*

$GF(2)$			$GF(3)$				$GF(5)$					
+	0	1	+	0	1	2	+	0	1	2	3	4
0	0	1	0	0	1	2	0	0	1	2	3	4
1	1	0	1	1	2	0	1	1	2	3	4	0
			2	2	0	1	2	2	3	4	0	1
							3	3	4	0	1	2
							4	4	0	1	2	3
×	0	1	×	0	1	2	×	0	1	2	3	4
0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	1	1	0	1	2	1	0	1	2	3	4
			2	0	2	1	2	0	2	4	1	3
							3	0	3	1	4	2
							4	0	4	3	2	1

Рис. 6.11. Таблицы сложения и умножения для простейших полей.

над полем $GF(p)$ памяти n . Элементы линейной рекуррентной последовательности вычисляются один за другим, причем каждый последующий определяется n предшествующими, так что задавав n начальных элементов d_0, d_1, \dots, d_{n-1} , можно сформировать всю последовательность.

Пример 6.6.1. Построим линейную рекуррентную последовательность памяти $n=3$ над полем $GF(2)$ (последовательности над $GF(2)$ называются также *бинарными*), начиная с исходных элементов $d_0=1, d_1=0, d_2=0$, при условии, что коэффициентами рекурсии являются $f_2=0, f_1=1, f_0=1$. Отмечая, что в двоичном поле противоположным любому элементу является сам элемент, рекурсия (6.13) принимает вид $d_i = d_{i-2} + d_{i-3}$, $i \geq 3$, так что получаем последовательность вида $1, 0, 0, 1, 0, 1, 1, 0, 1, 1, 0, 0, 1, 0, 1, 1, \dots$. Данная последовательность имеет период, равный 7. \square

Пример 6.6.2. Задавав начальные символы, как $d_0=1, d_1=0, d_2=0$, а коэффициенты рекурсии в (6.13) как $f_2=0, f_1=2, f_0=1$, построим линейную последовательность над полем $GF(3)$ (иначе *троичную* последовательность) памяти $n=3$. Поскольку в $GF(3)$ $-2=1, -1=2$, то рекурсия (6.13) принимает вид $d_i = d_{i-2} + 2d_{i-3}$, $i \geq 3$, в результате чего получаем следующую последовательность: $1, 0, 0, 2, 0, 2, 1, 2, 2, 1, 0, 2, 2, 2, 0, 0, 1, 0, 1, 2, 1, 1, 2, 0, 1, 1, \dots$. Отметим снова, что полученная последовательность является периодической с периодом 26, содержащей два блока длиной 13, причем второй блок представляет собой повторение первого, поэлементно умноженного на 2. \square

Обратимся к рис. 6.12, демонстрирующему типичный генератор линейной рекуррентной последовательности. В соответствии со своей структурой данная схема называется *регистром сдвига с линейной обратной связью (linear feedback shift register (LFSR))*. Регистр состоит из n p -ичных элементов задержки или разрядов (изображенных в виде

квадратов), имеющих p возможных состояний и хранящих некоторый элемент поля $GF(p)$ в течение тактового интервала. Схема тактовой синхронизации (не показана на рисунке) управляет работой регистра таким образом, что под действием каждого тактового импульса состояние любого разряда передается следующему слева направо. Схема обратной связи включает умножители элементов (состояний), хранящихся в разрядах, на константы f_i и сумматоров. Ясно, что обе арифметические операции выполняются по правилам конечного поля $GF(p)$.

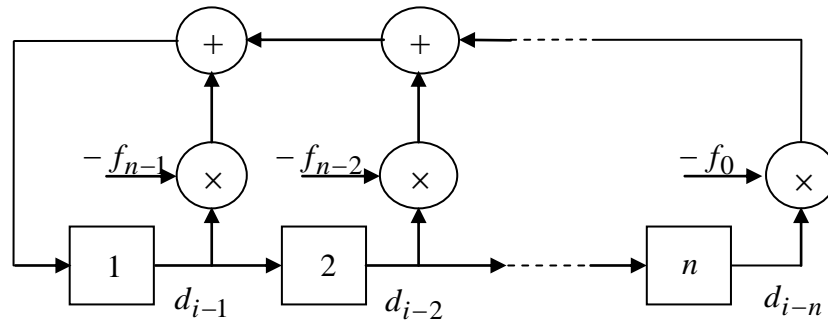


Рис. 6.12. Генератор линейной рекуррентной последовательности.

Предположим, что начальные состояния (т.е. начальные символы последовательности) $d_{n-1}, d_{n-2}, \dots, d_0$ записываются в разряды регистра слева направо, как это показано на рис. 6.12. Тогда состояние выхода схемы обратной связи будет $-f_{n-1}d_{n-1} - f_{n-2}d_{n-2} - \dots - f_0d_0 = d_n$ и после подачи тактового импульса содержимое регистра сдвига станет d_n, d_{n-1}, \dots, d_1 , формируя следующее состояние обратной связи в виде $-f_{n-1}d_n - f_{n-2}d_{n-1} - \dots - f_0d_1 = d_{n+1}$. После очередного такта состояние регистра станет d_{n+1}, d_n, \dots, d_2 и т. д. В общем случае, текущее содержимое регистра $d_{i-1}, d_{i-2}, \dots, d_{i-n}$ формирует d_i состояние обратной связи. Таким образом, полностью линейная рекуррентная последовательность может быть непосредственно считана с крайнего правого разряда, начиная с самого первого символа d_0 , или с любого другого разряда, но с соответствующим сдвигом.

Очевидно, что тот или иной разряд соединяется с сумматором посредством перемножителя только в том случае, если соответствующий коэффициент f_i в обратной связи ненулевой, в противном случае необходимость в соединении вообще отсутствует.

Пример 6.6.3. На рис. 6.13, а, представлена реализация генератора на основе регистра сдвига с линейной обратной связью, формирующего бинарную последовательность из примера 6.6.1. Отметим, что в случае бинарной последовательности умножение на единицу эквивалентно простому соединению выхода разряда с сумматором. Рис. 6.13, б, иллюстрирует следующие друг за другом содержания регистра (состояния разрядов), а также состояния выхода обратной связи (точка ОС на схеме) при подаче тактовых импульсов. Последовательность считывается в виде последовательных состояний крайнего правого разряда. Считывание состояний других разрядов приводит к копиям той же самой последовательности, сдвинутой на один или два такта. \square

Поскольку число различных состояний регистра конечно (не более p^n) неизбежна ситуация, когда после некоторого числа тактов состояние повторится в виде одного из ранее случившихся. Однако, стартуя с некоторой начальной загрузкой, т.е. фиксированного состояния, схема на рис. 6.12 сформирует только единственную последовательность, определяемую (6.13). Следовательно, повторение состояния регистра ведет к повторению всех последующих генерируемых символов, означаящее, что любая линейная рекуррент-

ная последовательность периодична. Более того, как следует непосредственно из (6.13), в случае нулевого состояния регистра (наличия нулей во всех разрядах) всегда будет формироваться бесконечная вырожденная последовательность, состоящая только из одних нулей. Очевидно, что подобный случай абсолютно бесперспективен, так что нулевое состояние регистра должно быть исключено. В результате остается не более $p^n - 1$ допустимых состояний регистра, что ограничивает максимально возможный период последовательности величиной, не большей $p^n - 1$.

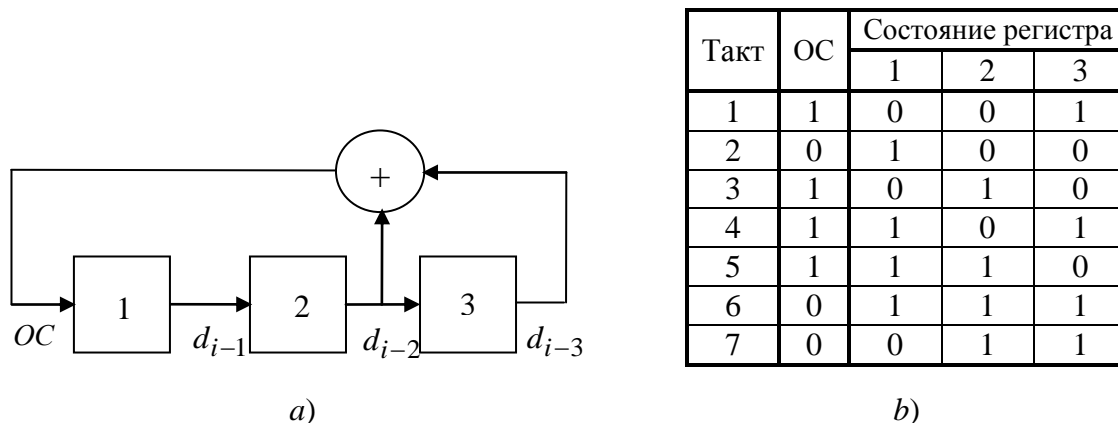


Рис. 6.13. Генератор бинарной последовательности (а) и таблица состояний (b).

6.6.3. m -последовательности.

Линейные рекуррентные последовательности, обладающие наибольшим периодом $L = p^n - 1$, представляют особый интерес в современной информационной технологии и называются последовательностями *максимальной длины* или просто m – последовательностями. Будучи полностью детерминированными, они обладают многими свойствами, присущими случайным последовательностям, например последовательности выпадания орлов и решеток при подбрасывании честной монеты. Следующие отличительные свойства делают m – последовательности чрезвычайно ценными в плане построения кодов с хорошей автокорреляцией.

1. *Свойство сбалансированности.* На одном периоде p – ичной m – последовательности любой ненулевой элемент $GF(p)$ встречается p^{n-1} раз, тогда как нулевой – $p^{n-1} - 1$ раз. Для того чтобы убедиться в этом, достаточно заметить, что все возможные $p^n - 1$ ненулевые состояния регистра должны произойти в том или ином порядке во время генерирования одного периода m – последовательности, иначе период не будет максимальным. Все эти состояния есть ничто иное, как различные n – разрядные p – ичные числа из диапазона $1, 2, \dots, p^n - 1$, а m – последовательность, считываемая с крайней правой ячейки регистра, может трактоваться как последовательность крайних правых разрядов этих чисел. При прохождении по диапазону $0, 1, 2, \dots, p^n - 1$ любое p – ичное значение на любой конкретной позиции (например, крайней правой) n – разрядного числа возникнет ровно p^{n-1} раз. Отбрасывание n – разрядного числа, состоящего из одних нулей, уменьшит на единицу только число появлений нуля среди крайних правых (или любых других) цифр. В частности, период двоичной m – последовательности памяти n составля-

ет $L = 2^n - 1$ и содержит $L_0 = 2^{n-1} - 1$ нулей и $L_1 = 2^{n-1}$ единиц. Например, легко увидеть, что последовательность, полученная в примере 6.6.1, является двоичной m – последовательностью периода $L = 2^3 - 1 = 7$, содержащей на данном интервале $L_0 = 2^2 - 1 = 3$ нуля и $L_1 = 2^2 = 4$ единицы. Последовательность из примера 6.6.2 является троичной m – последовательностью длины $L = 3^3 - 1 = 26$, содержащей на периоде $L_0 = 3^2 - 1 = 8$ нулей и $L_1 = L_2 = 3^2 = 9$ повторений каждого из элементов 1 и 2.

2. Любые две m – последовательности, сформированные на основе одной и той же рекурсии (6.13), отличаются друг от друга не более чем циклическим сдвигом. Действительно, поскольку при фиксированных n начальных элементах соотношение (6.13) полностью определяет последовательность, то две несовпадающие m – последовательности, формируемые заданной рекурсией, не могут иметь абсолютно идентичных n начальных элементов. С другой стороны, на одном периоде m – последовательности содержатся все состояния генератора LFSR кроме нулевого, и после того, как состояние, воспроизводящее начальную установку первого генератора, произойдет во втором, вторая последовательность полностью повторит первую, т.е. она представляет собой только некоторую задержанную копию первой m – последовательности.

3. *Свойство сдвига и вычитания.* Возьмем некоторую m – последовательность, определяемую (6.13), и вычтем поэлементно (очевидно, что по модулю p) ее копию, циклически сдвинутую на m позиций

$$d_{i+m} = -f_{n-1}d_{i+m-1} - f_{n-2}d_{i+m-2} - \dots - f_0d_{i+m-n},$$

где m – произвольное целое. В результате получим

$$d_i - d_{i+m} = -f_{n-1}(d_{i-1} - d_{i+m-1}) - f_{n-2}(d_{i-2} - d_{i+m-2}) - \dots - f_0(d_{i-n} - d_{i+m-n}).$$

Введя обозначение $d'_i = d_i - d_{i+m}$, приходим к линейной рекуррентной последовательности, элементы d'_i которой определяются первоначальной рекурсией

$$d'_i = -f_{n-1}d'_{i-1} - f_{n-2}d'_{i-2} - \dots - f_0d'_{i-n}. \quad (6.14)$$

Теперь возможны только два варианта. Предположим вначале, что сдвиг m равен целому числу периодов L . Тогда $d_i = d_{i+m}$ и $d'_i = 0$ для всех i , т.е. последовательность, описываемая (6.14), состоит из одних нулей.

Пусть теперь m не кратно L . Тогда d_i и d_{i+m} не могут быть идентичными при всех $i = 0, 1, \dots, L-1$, поскольку в противном случае совпадали бы начальные состояния генератора, как для исходной последовательности, так и ее сдвинутой копии. Это повлекло бы полное совпадение исходной последовательности и ее сдвинутой копии, поскольку начальная загрузка определяет единственную последовательность, формируемую рекурсией (6.13). Однако сдвинутая копия повторяет исходную последовательность только в случае сдвига, кратного периоду, что противоречит сделанному ранее предположению. Тогда очевидно, что рекурсия (6.14), полностью повторяющая (6.13), генерирует некоторую ненулевую последовательность. Однако, согласно предыдущему свойству, когда рекурсия (6.13) или, что эквивалентно, соответствующая схема на рис. 6.12, генерирует m – последовательность, стартуя с некоторого определенного начального состояния, то другую сдвинутую копию этой m – последовательности можно сформировать лишь в случае отличной ненулевой первоначальной загрузки. В результате приходим к следующему выводу: посимвольное вычитание двух сдвинутых копий одной и той же m – последовательности приводит либо к нулевой последовательности, если сдвиг кратен периоду, либо к некоторому другому сдвигу исходной m – последовательности в противном случае.

В случае двоичных последовательностей вычитание аналогично сложению, что объясняет альтернативное название данного свойства – свойство *сдвига и сложения*. Обратившись вновь к примеру 6.6.1, можно увидеть, например, что сложение полученной в нем последовательности с ее сдвинутой влево на две позиции копией $0,1,0,1,1,1,0,0,1,0,1,1,1,0,\dots$ дает последовательность $1,1,0,0,1,0,1,1,0,0,1,0,1,\dots$, которая является циклически сдвинутой вправо на одну позицию копией исходной последовательности.

Читателю предоставляется возможность самостоятельно убедиться в справедливости этого же свойства для троичной m – последовательности из примера 6.6.2.

Очевидно, что формирование p – ичной m – последовательности, т. е. последовательности с максимальной длиной, допускаемой данным значением памяти n , а не некоторой другой последовательности меньшего периода, полностью определяется адекватным выбором коэффициентов f_i в рекурсии (6.13) (или в схеме обратной связи генератора). Необходимое и достаточное условие формирования линейной рекуррентной последовательности максимальной длины состоит в том, что в качестве f_i должны использоваться коэффициенты $f_i, i = 0, 1, \dots, n-1$ примитивного над полем $GF(p)$ полинома степени n вида $f(x) = x^n + f_{n-1}x^{n-1} + f_{n-2}x^{n-2} + \dots + f_0$. Примитивные полиномы являются подклассом неприводимых полиномов. Полином $f(x)$ степени n над полем $GF(p)$ (т.е. с коэффициентами, принадлежащими $GF(p)$) называется неприводимым над $GF(p)$, если он не может быть представлен в виде произведения двух полиномов меньшей, чем n , степени. Среди множества всех полиномов указанные полиномы играют роль, аналогичную простым числам во множестве целых чисел. Не любой неприводимый полином является примитивным, хотя в случае $p = 2$ и при простом $2^n - 1$ все неприводимые полиномы оказываются примитивными. Доказательство необходимости и достаточности выбора обратной связи указанным выше способом требует привлечения несколько большего знания современной алгебры, что может увести достаточно далеко от основной линии изложения материала. Заинтересованный читатель может найти более подробное изложение в многочисленных источниках (например, [31-32]).

Примитивные полиномы широко представлены в виде таблиц в книгах по современной алгебре и теории кодирования или (в основном для $p = 2$) широкополосной связи [5, 6, 18, 32]. Альтернативным путем отыскания примитивных полиномов служит компьютерный поиск, который не представляет собой особо трудную задачу (например, см. задачу 6.47). В частности, функции поиска примитивных полиномов содержатся в Matlab Communications Toolbox.

Построение генератора m – последовательности осуществляется следующим непосредственным образом. При заданном значении p величина памяти n определяется необходимой длиной L , а отыскание соответствующего примитивного полинома исчерпывает искомую задачу.

В качестве комментария к примерам 6.6.1 и 6.6.2 можно отметить, что бинарная m – последовательность длины 7 построена на основе примитивного над $GF(2)$ полинома $f(x) = x^3 + x + 1$, а в качестве примитивного над $GF(3)$ полинома, привлеченного для построения троичной последовательности длины 26, использовался полином $f(x) = x^3 + 2x + 1$.

6.7. Периодическая АКФ m –последовательностей.

Результаты предыдущего параграфа легко приводят к минимаксным бинарным последовательностям с АКФ, удовлетворяющим (6.12). Рассмотрим двоичную m – последо-

вательность $\{d_i\}$ памяти n , т.е. длины $L = 2^n - 1$. Отобразим ее символы 0 и 1 в бинарный алфавит ± 1 согласно правилу

$$a_i = (-1)^{d_i} = \begin{cases} +1, & d_i = 0, \\ -1, & d_i = 1, \end{cases} \quad (6.15)$$

где при возведении (-1) в степень d_i последнее трактуется как вещественное число 0 или 1. Полученная в результате подобного преобразования последовательность $\{a_i\}$ вещественных бинарных символов ± 1 обладает периодом $N = L = 2^n - 1$ и является однозначным отображением исходной двоичной m -последовательности $\{d_i\}$. Было бы вполне естественным сохранить за ней то же самое название бинарной m -последовательности. Для исключения перепутывания можно было бы использовать дополнительное обозначение вида бинарная $\{\pm 1\}$ и бинарная $\{0,1\}$ последовательности. Определим теперь ненормированную периодическую АКФ (6.7) последовательности $\{a_i\}$:

$$R_p(m) = \sum_{i=0}^{N-1} a_i a_{i-m} = \sum_{i=0}^{L-1} (-1)^{d_i} (-1)^{d_{i-m}} = \sum_{i=0}^{L-1} (-1)^{d_i + d_{i-m}}. \quad (6.16)$$

Теперь можно привлечь свойство сдвига и сложения бинарных $\{0,1\}$ m -последовательностей. Сложение показателей степеней здесь осуществляется по модулю 2, поскольку получается тот же результат от возведения в степень, что и при обыкновенном арифметическом сложении. Однако $\{d'_i\} = \{d_i + d_{i-m}\}$ представляет собой бинарную $\{0,1\}$ m -последовательность периода L при $m \neq 0 \pmod{L}$, или последовательность из одних нулей в противном случае. Вследствие свойства сбалансированности на одном периоде $\{d'_i\} = \{d_i + d_{i-m}\}$ содержится $L_0 = 2^{n-1} - 1$ нулей и $L_1 = 2^{n-1}$ единиц, поэтому сумма (6.16) состоит из L_0 «+1» и L_1 «-1» при $m \neq 0 \pmod{N}$, так что

$$R_p(m) = L_0 - L_1 = \begin{cases} N, & m = 0 \pmod{N}, \\ -1, & m \neq 0 \pmod{N}. \end{cases}$$

Как можно видеть, данное выражение в точности совпадает с (6.12), подтверждая тем самым, что бинарные m -последовательности являются минимаксными.¹

Таблица 6.3. Вычисление периодической АКФ бинарной m -последовательности.

m	a_0	a_1	a_2	a_3	a_4	a_5	a_6	$R_p(m)$
0	-	+	+	-	+	-	-	+7
1	-	-	+	+	-	+	-	-1
2	-	-	-	+	+	-	+	-1
3	+	-	-	-	+	+	-	-1

Пример 6.7.1. Рассмотрим снова последовательность из примера 6.6.1. Ее отображение на алфавит $\{\pm 1\}$ в соответствии с (6.15) дает бинарную $\{\pm 1\}$ m -последовательность $-1, +1, +1, -1, +1, -1, -1, \dots$. Табл. 6.3, содержащая только минимально необходимое

¹ Обобщение отображения (6.15) на случай последовательностей над $GF(p)$, $p \geq 2$ представимо как $a_i = \exp(j2\pi d_i / p)$, приводя к многофазным (p -фазным) кодам, периодическая АКФ которых также удовлетворяет (6.12). Однако многофазные коды этого типа с $p > 2$ менее интересны с практической точки зрения, чем минимаксные бинарные последовательности.

число записей, иллюстрирует вычисление периодической АКФ последовательности. Представляет также интерес рассмотрение процесса согласованной фильтрации дискретного сигнала, модулированного данной бинарной последовательностью. На рис. 6.14 изображен подобный фильтр, согласованный с одним периодом периодического видеосигнала с прямоугольными чипами. Все блоки данной структуры абсолютно идентичны схеме, представленной на рис. 6.7. Диаграммы напряжений в характерных точках фильтра показаны на рис. 6.15. Как видно, выходной сигнал обладает основными лепестками, повторяющимися с периодом $N\Delta$ и равномерным боковым лепестком отрицательной полярности, уровень которого в семь раз меньше основного. □

Бинарные m – последовательности относятся к числу наиболее популярных дис-

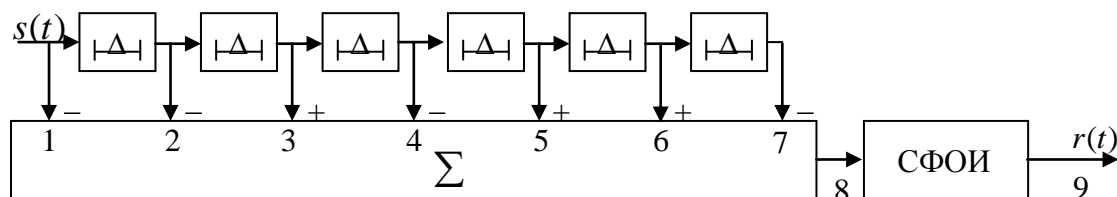


Рис.6.14. Согласованный фильтр для бинарной m - последовательности.

кретных сигналов, применяемых в современной информационной технологии, благодаря своим оптимальным периодическим корреляционным свойствам и простоте формирования и обработки. Возможно, одним из наиболее ярких примеров их практического использования является система мобильной связи 2-го поколения cdmaOne (IS-95), в которой в качестве пилот-сигналов для начальной синхронизации, мультиплексирования сигналов базовых станций и скремблирования данных применяются m – последовательности различной длины.

В дополнение к этому, m – последовательности образуют базис для формирования других важных семейств сигналов (Касами, Голда и др., см. главу 7).

В то же время, множество длин $N = 2^n - 1 = 3, 7, 15, 31, 63, 127, 255, 511, 1023, \dots$, при которых существуют данные последовательности, достаточно разрежено, что в некоторых случаях может оказаться ограничительным фактором к практическому использованию. В связи с этим представляет интерес рассмотреть еще один класс бинарных минимаксных последовательностей, что, однако, потребует дополнительных сведений о теории конечных полей.

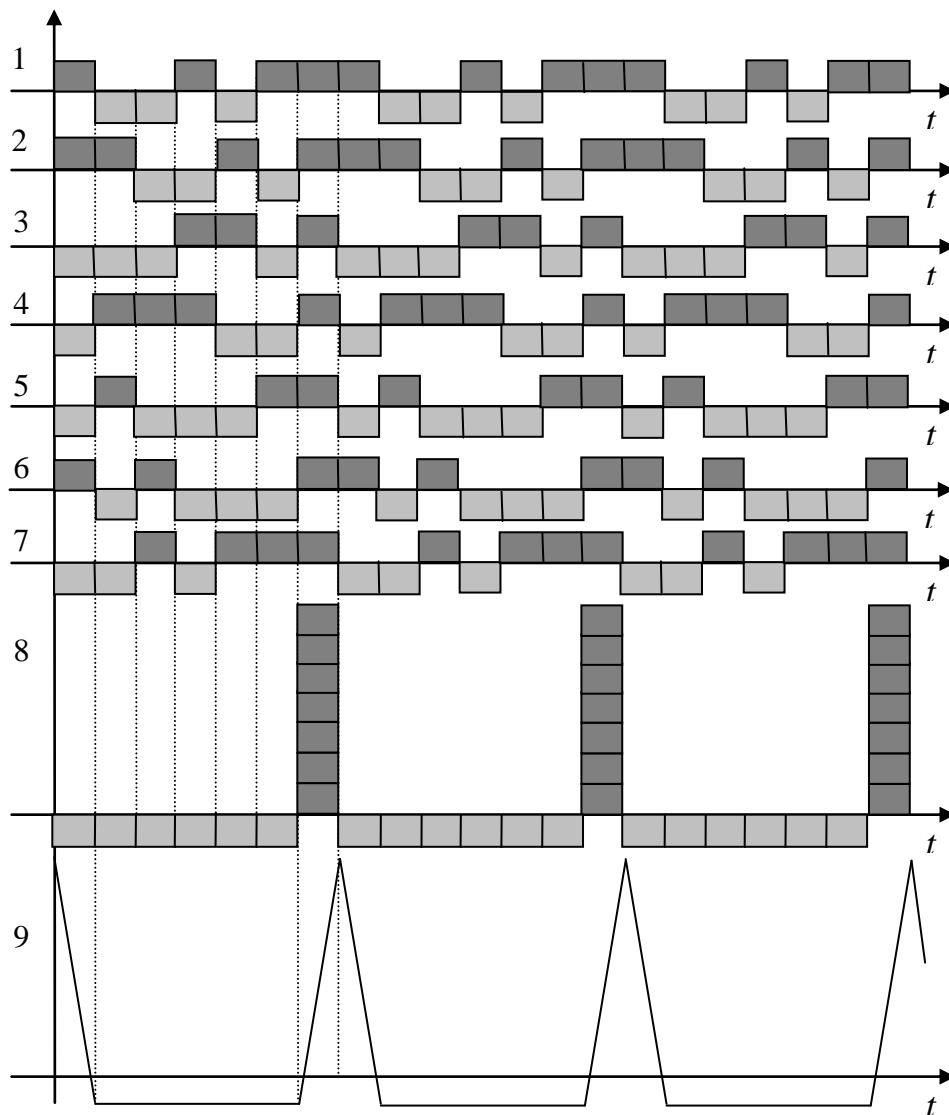


Рис.6.15. Процедура согласованной фильтрации.

6.8. Дополнительные сведения о конечных полях.

Возьмем некоторый элемент x конечного поля $GF(p)$ и умножим его сам на себя m раз, присвоив результату наименование m -й степени элемента x :

$$\underbrace{x \cdot x \cdot \dots \cdot x}_{m \text{ раз}} = x^m.$$

Стандартные правила обращения со степенями обычной алгебры сохраняют свою справедливость и в полях, включая и конечные. В частности

$$x^m x^n = \underbrace{x \cdot x \cdot \dots \cdot x}_{m \text{ раз}} \cdot \underbrace{x \cdot x \cdot \dots \cdot x}_{n \text{ раз}} = \underbrace{x \cdot x \cdot \dots \cdot x}_{m+n \text{ раз}} = x^{m+n}, \quad (x^m)^n = \underbrace{x^m \cdot x^m \cdot \dots \cdot x^m}_{n \text{ раз}} = x^{mn}.$$

Более того, обозначая n -ю степень x^{-1} ($x \neq 0$) как x^{-n} , имеем

$$x^m x^{-n} = \underbrace{x \cdot x \cdot \dots \cdot x}_{m \text{ раз}} \cdot \underbrace{x^{-1} \cdot x^{-1} \cdot \dots \cdot x^{-1}}_{n \text{ раз}}.$$

Тогда неоднократное использование определение обратного элемента $x^{-1}x=1$ приводит к соотношению

$$x^m x^{-n} = \begin{cases} x^{m-n}, & m \geq n, \\ (x^{-1})^{n-m}, & m < n, \end{cases} = x^{m-n}.$$

Аналогичным образом из

$$\underbrace{x \cdot x \cdot \dots \cdot x}_{m \text{ раз}} \cdot \underbrace{x^{-1} \cdot x^{-1} \cdot \dots \cdot x^{-1}}_{m \text{ раз}} = 1 = x^m x^{-m} = x^{m-m}$$

и единственности обратного элемента любому ненулевому следует, что

$$x^0 = 1 \text{ и } (x^m)^{-1} = x^{-m}.$$

Пример 6.8.1. В поле $GF(5)$ (см. таблицы на рис. 6.11)

$$2^0 = 1, 2^1 = 2, 2^2 = 2 \cdot 2 = 4, 2^3 = 2^2 \cdot 2 = 4 \cdot 2 = 3, 2^4 = 2^3 \cdot 2 = 3 \cdot 2 = 1;$$

$$2^{-1} = 3, 2^{-2} = (2^{-1})^2 = 3^2 = 4, 2^{-3} = (2^{-1})^3 = 3^3 = 4 \cdot 3 = 2, 2^{-4} = (2^{-1})^4 = 3^4 = 2 \cdot 3 = 1.$$

□

Рассмотрим теперь последовательные степени элемента $x \neq 0$ поля $GF(p)$: $x^0 = 1, x^1, x^2, \dots$. Поскольку все члены этой серии принадлежат $GF(p)$, то они не могут быть все различными и, следовательно, для некоторых $i > k$ будет выполняться равенство $x^i = x^k \Rightarrow x^{i-k} = 1$. Предположим, что существует элемент α , у которого первые $p-1$ степеней $\alpha^0 = 1, \alpha^1, \alpha^2, \dots, \alpha^{p-2}$ все различны. Поскольку $p-1$ в точности равно числу ненулевых элементов $GF(p)$, то вышеприведенные степени в точности исчерпывают все ненулевые элементы $GF(p)$. Таким образом, элемент α , если он действительно существует, позволяет построить все поле $GF(p)$, за исключением нулевого элемента, посредством возведения α в степени $0, 1, \dots, p-2$. Подобный элемент получил название *примитивного* элемента.

Одним из наиболее важных положений о конечных полях является то, что в любом из них содержится примитивный элемент. Доказательство данного положения может быть найдено во многих книгах по современной алгебре или теории кодирования (например, [30, 32, 33]). Примитивный элемент не является единственным: в любом конечном поле, порядок которого превышает 3, имеется более одного примитивного элемента. Например, как видно из примера 6.8.1, как элемент 2, так и 3 являются примитивными в поле $GF(5)$.

Поскольку для примитивного элемента степени $\alpha^0 = 1, \alpha^1, \alpha^2, \dots, \alpha^{p-2}$ исчерпывают все ненулевые элементы $GF(p)$, то α^{p-1} должен быть равен какому-то из них. В действительности он не может быть равен никакому другому, кроме 1, поскольку равенство $\alpha^{p-1} = \alpha^l$ при $0 < l \leq p-2$ означает, что $\alpha^{p-1-l} = 1$. Однако это не возможно, поскольку $1 < p-1-l < p-1$ и среди элементов $\alpha^1, \alpha^2, \dots, \alpha^{p-2}$ не может быть равного единице. Следовательно, $\alpha^{p-1} = 1$. Теперь легко увидеть, что подобное утверждение верно для любого ненулевого элемента конечного поля, а не только для примитивного. Действительно, любой ненулевой элемент $x \in GF(p)$ представим как l -я степень примитивного элемента для соответствующего целого l : $x = \alpha^l$, так что (малая теорема Ферма)

$$x^{p-1} = (\alpha^l)^{p-1} = \alpha^{l(p-1)} = (\alpha^{p-1})^l = 1. \quad (6.17)$$

Следующее определение имеет совершенно естественное наименование, полностью совпадающее с категорией обычной алгебры. Целое значение показателя l , в которое возводится α для получения $x = \alpha^l$, называется *логарифмом x по основанию α* с использованием обозначения $\log_\alpha x$. Следовательно, $\alpha^{\log_\alpha x} = x$.

Рассмотрим теперь только простые поля нечетного порядка ($p > 2$) и введем понятие *двузначного характера* $\psi(x)$ ненулевого элемента x , определяемое как

$$\psi(x) = \begin{cases} 1, & \log_\alpha x \equiv 0 \pmod{2}, \\ -1, & \log_\alpha x \not\equiv 0 \pmod{2}, \end{cases} = (-1)^{\log_\alpha x}. \quad (6.18)$$

Ясно, что двузначный характер представляет собой просто отображение поля $GF(p)$ на пару вещественных чисел $\{+1, -1\}$, превращая ненулевой элемент x в $+1$, если его логарифм четен, и в -1 в противном случае. Очевидно, что подобное отображение зависит от конкретного выбора примитивного элемента (задача 6.24). В дальнейшем будут использованы следующие свойства двузначного характера.

1. Значение характера от единичного элемента поля $GF(p)$ всегда равно единице:

$$\psi(1) = 1. \quad (6.19)$$

Справедливость подобного утверждения непосредственно вытекает из следующего соотношения $\alpha^0 = 1 \Rightarrow \log_\alpha 1 = 0$.

2. Характер является мультипликативной функцией, т.е. характер от произведения двух ненулевых элементов есть произведение их характеров. Действительно, из (6.17) и (6.18) следует

$$\psi(xy) = (-1)^{\log_\alpha(xy)} = (-1)^{\log_\alpha x + \log_\alpha y} = (-1)^{\log_\alpha x} \cdot (-1)^{\log_\alpha y} = \psi(x) \cdot \psi(y). \quad (6.20)$$

3. Свойство сбалансированности: сумма характеров всех ненулевых элементов поля $GF(p)$ равна нулю:

$$\sum_{x=1}^{p-1} \psi(x) = \sum_{x=1}^{p-1} (-1)^{\log_\alpha x} = 0. \quad (6.21)$$

Доказательство данного равенства основано на том факте, что при переборе x всех $p-1$ ненулевых значений $\log_\alpha x$ пробегает в некотором порядке по диапазону из $p-1$ целых чисел вида $0, 1, \dots, p-2$. Вследствие нечетности p число целых чисел в указанном диапазоне четно, так что $(p-1)/2$ из них являются четными и $(p-1)/2$ – нечетными. Следовательно, выше приведенная сумма содержит одинаковое число плюс и минус единиц и,

значит, равна нулю.

4. Рассмотрим характер элемента, противоположного единице, т.е. $\psi(-1)$. Поскольку все элементы α^i , $i = 0, 1, \dots, p-2$ различны, то только два из них удовлетворяют уравнению $x^2 = 1$, а именно 1 и -1 , а само уравнение удовлетворяет значению $\alpha^{\frac{p-1}{2}}$, причем последний элемент не может быть иным, чем -1 . В результате получаем, что $\log_{\alpha}(-1) = \frac{p-1}{2}$ и выражение

$$\psi(-1) = (-1)^{\frac{p-1}{2}} = \begin{cases} 1, & p \equiv 1 \pmod{4}, \\ -1, & p \equiv 3 \pmod{4}. \end{cases} \quad (6.22)$$

Пример 6.8.2. Продолжая пример 6.8.1, отметим, что в поле $GF(5)$ $\psi(2) = \psi(3) = -1$, а $\psi(1) = \psi(4) = +1$. Тогда в соответствие со свойством мультипликативности $\psi(2 \cdot 4) = \psi(3) = -1 = \psi(2) \cdot \psi(4) = (-1) \cdot (+1) = -1$ и согласно свойству сбалансированности $\psi(1) + \psi(2) + \psi(3) + \psi(4) = 0$. Кроме того, $\psi(-1) = \psi(4) = 1$, что согласуется с (6.22), поскольку $5 = 4 \cdot 1 + 1$. \square

Другим наименованием двузначного характера, особенно популярного в теории чисел, является символ Лежандра, что объясняет название последовательностей, рассматриваемых в следующем параграфе.

6.9. Последовательности Лежандра.

Сформируем бинарную последовательность нечетной простой длины $N = p$, отождествляя номер позиции i элемента $a_i = \pm 1$ с элементом простого поля $GF(p)$. Тогда для любого $i \in \{1, 2, \dots, N-1\}$ определен характер $\psi(i)$ и последовательность Лежандра есть просто последовательность характеров номеров i за исключением $i = 0$, для которого элемент последовательности устанавливается равным $+1$ ¹. Для периодической версии последовательности Лежандра правило формирования определяется в виде

$$a_i = \begin{cases} +1, & i \equiv 0 \pmod{N}, \\ \psi(i), & i \not\equiv 0 \pmod{N}. \end{cases} \quad (6.23)$$

Периодичность последовательности (6.23) с периодом N следует из трактовки номеров i в $\psi(i)$ как элементов $GF(p)$, причем сложение выполняется по модулю p , в результате чего приходим к выражению $\psi(i + N) = \psi(i + p) = \psi(i)$.

Для исследования периодической АКФ последовательности Лежандра подставим (6.23) в (6.7) и вынесем из-под знака суммы члены, содержащие a_0 :

$$R_p(m) = \sum_{i=0}^{N-1} a_i a_{i-m} = a_0 a_{-m} + a_m a_0 + \sum_{\substack{i=1 \\ i \neq m}}^{N-1} a_i a_{i-m} = \psi(-m) + \psi(m) + \sum_{\substack{i=1 \\ i \neq m}}^{p-1} \psi(i) \psi(i-m). \quad (6.24)$$

Очевидно, что интересно оценить только уровень боковых лепестков, т.е. при тех сдвигах m , когда они не кратны $N = p$. Применение свойства мультипликативности (6.20) дает $\psi(-m) = \psi(-1)\psi(m)$ и $\psi(i-m) = \psi[i(1-mi^{-1})] = \psi(i)\psi(1-mi^{-1})$, где всегда определен обратный элемент, поскольку нулевое значение i вынесено из-под знака суммы в (6.24). В

¹ Другой выбор, приводящий к аналогичному конечному результату, состоит в присваивании этому элементу значения -1 .

результате получаем

$$R_p(m) = \psi(m)[1 + \psi(-1)] + \sum_{\substack{i=1 \\ i \neq m}}^{p-1} \psi^2(i)\psi(1-mi^{-1}). \quad (6.25)$$

Обратимся теперь к (6.22). Для любой длины вида $N = p = 1 \pmod 4$ $\psi(-1) = 1$ и первое слагаемое в вышеприведенном соотношении равно $2\psi(m) = \pm 2$ для любого $m \neq 0 \pmod N$. С другой стороны, для длин типа $N = 3 \pmod 4$ $\psi(-1) = -1$ и то же самое слагаемое в (6.25) аннулируется. Для вычисления второго слагаемого в (6.25) отметим, во-первых, что для любого ненулевого i из $GF(p)$ $\psi^2(i) = 1$. Во-вторых, при пробегании i по всем ненулевым элементам $GF(p)$ i^{-1} , как и $-mi^{-1}$ ($m \neq 0 \pmod p$), также пробегает указанный диапазон, но в некотором другом порядке. Следовательно, $1-mi^{-1}$ пробегает множество $p-1$ элементов поля, включая нуль, но исключая 1, поскольку $-mi^{-1}$ не может принимать нулевого значения. В действительности же нулевой элемент также должен быть исключен из числа возможных значений $1-mi^{-1}$, поскольку i во втором слагаемом (6.24) не принимает значения $i = m$, отвечающего $1-mi^{-1} = 0$, тогда полный диапазон значений $1-mi^{-1}$ простирается от 2 до $p-1$. Резюмируя все вышесказанное, приходим к следующему результату

$$\sum_{\substack{i=1 \\ i \neq m}}^{p-1} \psi^2(i)\psi(1-mi^{-1}) = \sum_{x=2}^{p-1} \psi(x) = \sum_{x=1}^{p-1} \psi(x) - \psi(1) = -1,$$

где завершающий шаг следует из свойств характера (6.19) и (6.21). В итоге периодическая АКФ последовательности Лежандра оказывается возможной одного из двух типов:

1. Если длина имеет вид $N = 4h + 1$ (т.е. $N = 1 \pmod 4$), то

$$R_p(m) = \begin{cases} N, & m = 0 \pmod N, \\ -3 \text{ или } +1, & m \neq 0 \pmod N; \end{cases} \quad (6.26)$$

2. Если же длина представима как $N = 4h + 3$ (т.е. $N = 3 \pmod 4$), то

$$R_p(m) = \begin{cases} N, & m = 0 \pmod N, \\ -1, & m \neq 0 \pmod N; \end{cases} \quad (6.27)$$

где h – натуральное число.

Последний результат, повторяющий (6.12), демонстрирует, что последовательности Лежандра длин $N = 4h + 3$ являются минимаксными, т.е. обладают оптимальными периодическими корреляционными свойствами возможными для бинарных последовательностей нечетных длин.

Пример 6.9.1. Длина $N = 7$ принадлежит множеству вида $N = 4h + 3$. В поле $GF(7)$ элемент 3 является примитивным, поскольку возведение его в степень 0, 1, ..., 5 дает все различные ненулевые элементы: $3^0 = 1, 3^1 = 3, 3^2 = 2, 3^3 = 6, 3^4 = 4, 3^5 = 5$. Как непосредственно следует из этой серии, логарифмы 1, 2, и 4 четны, тогда как 3, 5 и 6 – нечетны. Следовательно, $\psi(1) = \psi(2) = \psi(4) = 1$, а $\psi(3) = \psi(5) = \psi(6) = -1$. Теперь, согласно (6.23), расстановка плюс единицы на позициях $i = 0, 1, 2, 4$ и минус единицы на позициях $i = 3, 5, 6$ дает последовательность Лежандра длины $N = 7$. Вычисления, иллюстрируемые табл. 6.4, подтверждают оптимальность корреляционных свойств полученной бинарной последовательности. \square

Последовательности Лежандра образуют достаточно мощный класс бинарных ко-

дов с минимаксной периодической АКФ. Условие их существования (любая простая длина вида $N = 4h + 3$) значительно мягче, чем у m – последовательностей ($N = 2^n - 1$), благодаря чему последовательности Лежандра более доступны по сравнению с m – по-

Таблица 6.4. Вычисление периодической АКФ последовательности Лежандра.

m	a_0	a_1	a_2	a_3	a_4	a_5	a_6	$R_p(m)$
0	+	+	+	–	+	–	–	+7
1	–	+	+	+	–	+	–	–1
2	–	–	+	+	+	–	+	–1
3	+	–	–	+	+	+	–	–1

следовательностями. Например, на интервале длин от 50 до 1500 m – последовательности существуют только при пяти значениях длин, тогда как доступное число последовательностей Лежандра составляет 114.

6.10. Бинарные коды с хорошей аperiodической АКФ (продолжение).

После накопления необходимых знаний о бинарных последовательностях с хорошими периодическими АКФ можно возвратиться к идее, сформулированной в 6.4 и заключающейся в использовании этих последовательностей в качестве исходного материала для поиска кодов с привлекательной аperiodической АКФ. Рассмотрим некоторую последовательность a_0, a_1, \dots, a_{N-1} длины N . Любой ее циклический сдвиг $a_s, a_{s+1}, \dots, a_{N-1}, a_0, \dots, a_{s-1}$, $1 \leq s \leq N-1$, обладает такой же периодической АКФ, что и исходный код, поскольку периодическая АКФ инвариантна к циклическому сдвигу (см. задачу 5.5). Аperiodическая же АКФ циклически сдвинутой копии может отличаться от АКФ первоначальной. Вместе с границей (6.5) данный факт устанавливает основу для распространенного алгоритма поиска кодов с приемлемой аperiodической АКФ, описанного ниже.

На первом шаге для заданной длины N некоторым образом формируется множество последовательностей кандидатов с хорошей периодической АКФ. Оно может включать все известные последовательности заданной длины N [35-37], уровень боковых лепестков периодической АКФ которых, согласно (6.5), позволяет надеяться на получение низкого значения $\rho_{a, \max}$ или быть ограниченным в соответствии технологическими предпочтениями проектировщика. Например, если необходимы бинарные коды длины $N = 63$, то первоначальное множество может быть ограничено только всеми m – последовательностями данной длины (при указанной длине не существует последовательность Лежандра, т. к. N не простое число) или включать некоторые другие последовательности с удовлетворительной периодической АКФ. Если же необходима длина $N = 127$, то множество может охватывать все m – последовательности наряду с последовательностями Лежандра¹ или содержать другие последовательности с достаточно низкими периодическими боковыми лепестками.

¹ При одной и той же степени могут существовать различные примитивные полиномы, каждый из которых генерирует отличные m – последовательности одной и той же длины. Следовательно, при фиксированной длине могут существовать несколько m – последовательностей, и все они подходят для поиска хороших аperiodических кодов. В отличие от предыдущего существуют только две последовательности Лежандра одинаковой длины, отличающиеся значением первого символа (+1 в одной из них и –1 в другой (см. сноску на стр.172)).

На втором этапе осуществляется исчерпывающий поиск по критерию наименьшего уровня максимума бокового лепестка аperiodической АКФ среди всех однопериодных сегментов последовательностей кандидатов. В частности, берется однопериодный сегмент первой последовательности кандидата, вычисляется его аperiodическая АКФ и запоминается в памяти уровень максимального бокового лепестка наряду с номерами последовательности кандидата и его сдвига. Затем осуществляется циклический сдвиг сегмента на одну позицию, и производятся все необходимые вычисления. Если новое значение максимума аperiodического бокового лепестка окажется ниже предыдущего, то его значение и номер нового сдвига заменяют ранее записанные в памяти данные, в противном случае зарегистрированные значения сохраняются без изменения. Данная процедура повторяется N раз, т.е. для всех циклических сдвигов первой последовательности кандидата, после чего подобному исследованию подвергается следующая последовательность кандидат и т.д. Результатом поиска служит последовательность с минимальным значением $\rho_{a,\max}$ среди всех последовательностей, отобранных на первом этапе. Очевидно, отсутствуют гарантии того, что полученный результат будет наилучшим среди всех возможных бинарных последовательностей данной длины.

Данная процедура, впервые предложенная в начале 60-х годов, в последствие широко использовалась многими авторами, постепенно охватывая все более и более обширное множество кандидатов среди бинарных последовательностей. Один из наиболее подробных списков бинарных кодов, синтезированных подобным образом, может быть найден в [34].

Пример 6.10.1. Длина $N = 2^3 - 1 = 7$ удовлетворяет условию существования m – последовательности. Имеются два примитивных бинарных полинома степени 3: $f(x) = x^3 + x + 1$ и $f(x) = x^3 + x^2 + 1$. Непосредственная проверка показывает, что m – последовательности, порождаемые ими, являются зеркальным отображением друг друга, т.е. одна из них получается из другой считыванием справа налево. Подобное преобразование не изменяет ни периодическую, ни аperiodическую АКФ (см задачу 5.5). Следовательно, достаточно включить во множество кандидатов только одну m – последовательность, например, построенную в примере 6.7.1: $-1, +1, +1, -1, +1, -1, -1$. Кроме того, $N = 7$ является простым числом, удовлетворяющим соотношению $N = 4h + 3$, при котором также существуют и минимаксные последовательности Лежандра, т.е. последовательность, полученная в примере 6.9.1: $+1, +1, +1, -1, +1, -1, -1$, а также ее копия с первым символом, замененным -1 . Последняя полностью повторяет выбранную m – последовательность, тогда как первая – после замены знаков всех элементов – совпадает с циклически сдвинутым зеркальным отображением m – последовательности. Поскольку изменение полярности также не затрагивает ни периодическую, ни аperiodическую АКФ (см. задачу 5.5), то из четырех возможных кандидатов в множество для анализа достаточно включить только одну минимаксную последовательность. Пусть это будет последовательность Лежандра, начинающаяся символом $+1$. Вычисление ее аperiodической АКФ дает следующие значения $R_a(m)$, $m = 1, 2, \dots, 6$: $0, +1, 0, -1, -2, -1, +1$ и $\rho_{a,\max} = 2/7$. После циклического сдвига влево последовательность превращается в $+1, +1, -1, +1, -1, -1, +1$, а вычисления приводят к следующему результату $R_a(4) = -3$, $\rho_{a,\max} = 3/7$, т.е. максимальный аperiodический боковой лепесток хуже, чем у исходного кода. Следующий циклический сдвиг дает $+1, -1, +1, -1, -1, +1, +1$ и $R_a(1) = -2$, $\rho_{a,\max} = 2/7$, т.е. аналогично первоначальному результату. После следующего сдвига приходим к последовательности $-1, +1, -1, -1, +1, +1, +1$, имеющей аperiodическую АКФ с боковыми лепестками $R_a(m) = 0, -1$; $m \neq 0$, т.е. с $\rho_{a,\max} = 1/7$. Данная последовательность является глобально оптимальной среди всех ФМ кодов, поскольку ни один из подобных кодов не может иметь

меньший уровень максимального аperiodического бокового лепестка (см. (6.4)). Действительно, найден код Баркера длины 7, который является зеркальной копией последовательности, приведенной в табл. 6.1. □

Пример 6.10.2. Предположим, что $N = 257 = 64 \times 4 + 1$. Поскольку число 257 простое, то существуют две последовательности Лежандра данной длины, отличающиеся только первым символом. Вместе с тем, N представимо в виде $4h + 1$ и, следовательно, максимальный боковой лепесток ПАКФ этих последовательностей составляет величину $\rho_{p, \max} = 3/N = 3/257$ (см. (6.26)). Несмотря на это упомянутые последовательности могут рассматриваться как перспективные с точки зрения $\rho_{a, \max}$, поскольку нижняя граница (6.4) устанавливает $\rho_{a, \max} \geq 1.5/257$. Применение ранее описанной процедуры приводит к последовательности с максимальным ненормированным аperiodическим боковым лепестком, равным 12, т.е. $\rho_{a, \max} = 12/257$ или -26.6 дБ (сравним полученный результат с наиболее длинным бинарным кодом Баркера, для которого $\rho_{a, \max} = 1/13$ или -22.3 дБ). Последовательность, получаемая после отбрасывания последнего символа, превращается в код длины $N = 256$ с тем же значением максимального не нормированного бокового лепестка и, значит, $\rho_{a, \max} = 12/256 = 3/64$, т.е. снова приблизительно -26.6 дБ. Аperiodическая АКФ данного кода показана на рис. 6.16, а. Интересно, что в стандарте UMTS 3-го поколения в качестве кода первичной синхронизации используется бинарная последова-

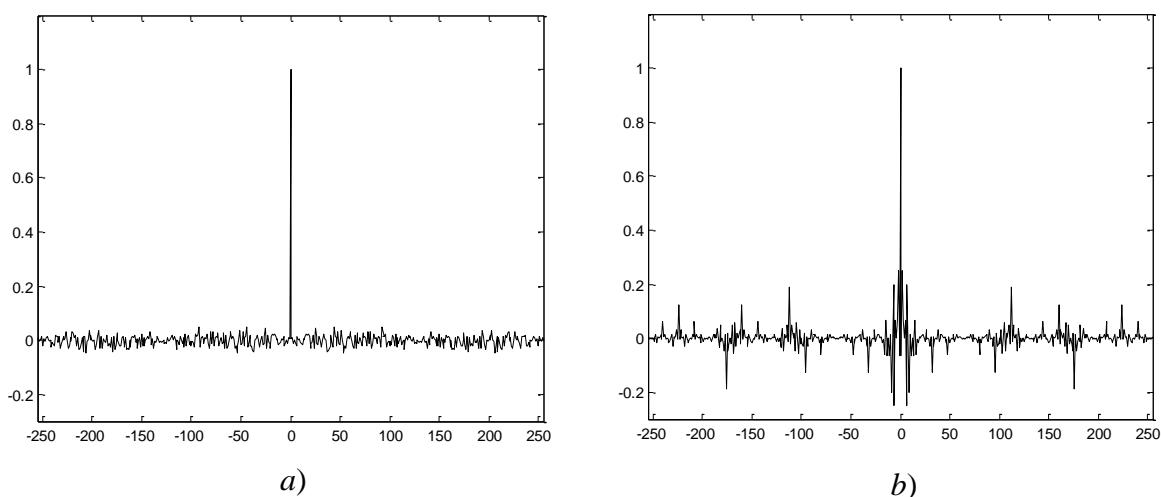


Рис.6.16. Аperiodическая АКФ двух бинарных кодов длины 256: код из примера 6.10.2 (а) и код первичной синхронизации UMTS (b).

тельность точно такой же длины $N = 256$, обладающей аperiodическими боковыми лепестками вплоть до $1/4$ (рис. 6.16, b), т.е. значительно большими в сравнении с только что найденной последовательностью. С другой стороны, выбор кода для поиска соты в UMTS производился с учетом многих других требований, включая реализационные, которые, возможно, перевесили критерий хорошей автокорреляции. □

На рис. 6.17 представлен еще один иллюстративный материал по оптимизации бинарных кодов с точки зрения максимального бокового лепестка аperiodической АКФ, демонстрирующий зависимость $\rho_{a, \max}$ от длины N для предположительно лучших бинарных последовательностей, взятых из [25-27, 34]. Пунктирная линия отвечает кривой $\rho_{a, \max} \approx 0.77/\sqrt{N}$, аппроксимирующей зависимость $\rho_{a, \max} = f(N)$ как a/\sqrt{N} со значением a , найденным с помощью метода наименьших квадратов. Как следует из рисунка,

точность подобной аппроксимации достаточно высока, особенно для $N > 100$.

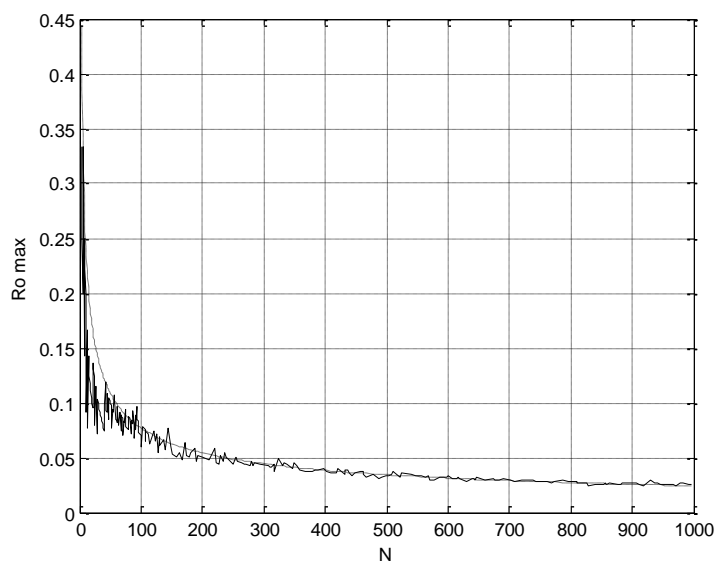


Рис.6.17. Зависимость минимизированного $\rho_{a, \max}$ от длины N .

6.12. Подавление боковых лепестков вдоль оси задержек.

Предположим, что проектировщик системы не склонен отвергать бинарные $\{\pm 1\}$ последовательности и, в то же время, не удовлетворен достижимым уровнем боковых лепестков их периодической АКФ ($\rho_{p, \max} \geq 1/N$). В подобных условиях эффективным способом разрешения этих противоречивых устремлений служит «имитация» идеальной периодической АКФ путем отказа от согласованной фильтрации в пользу специальной расогласованной обработки, позволяющей подавить боковые лепестки на всем периоде сигнала. Очень близкие идеи лежат в основе уменьшения или подавления аperiodических боковых лепестков [39,44,45], также как в стремлении побороть межсимвольную интерференцию с помощью нуль-форсирующих эквалайзеров [2,5,7], однако в наиболее прозрачной форме они проявляются в случае применения к периодическим сигналам [39,46,47].

6.12.1. Фильтр подавления боковых лепестков.

Рассмотрим некоторую последовательность $\dots, a_{i-1}, a_i, a_{i+1}, \dots$ периода N , которая манипулирует чипы длительности Δ , и фильтр с конечным импульсным откликом (FIR), осуществляющий суммирование N сигнальных копий, задержанных на Δ и взвешенных

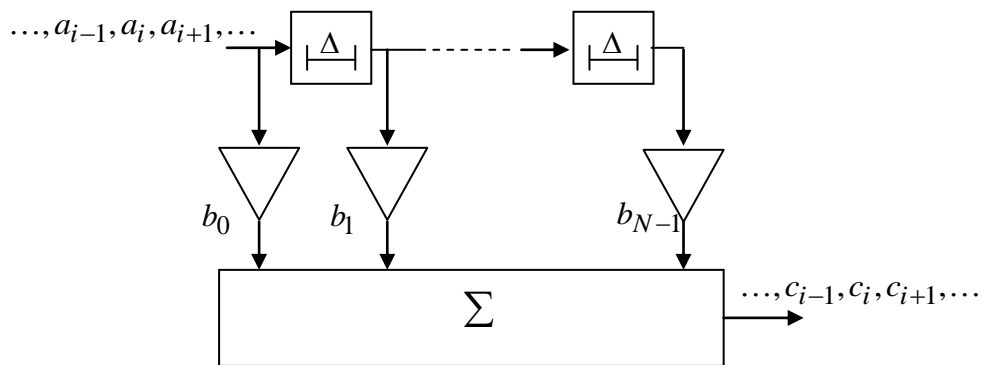


Рис. 6.20. FIR фильтр для последовательности длины N .

коэффициентами $b_i, i = 0, 1, \dots, N-1$, как это показано на рис.6.20. В принципе, то, что представлено ниже, может быть применено к последовательностям произвольного алфавита, однако представляется рациональным ограничиться только бинарным $\{\pm 1\}$ алфавитом, поскольку вне этого ограничения существует множество последовательностей с идеальной периодической АКФ, тем самым лишая задачу подавления боковых лепестков обоснованной мотивации. Соответственно, положим, что коэффициенты фильтра $b_i, i = 0, 1, \dots, N-1$ являются вещественными.

При подаче последовательности $a_i, i = \dots, -1, 0, 1, \dots$ отклик фильтра описывается последовательностью $c_i, i = \dots, -1, 0, 1, \dots$, элементы которой находятся операцией свертки

$$c_i = \sum_{l=0}^{N-1} a_{i-l} b_l, \quad i = \dots, -1, 0, 1, \dots$$

При периодической входной последовательности $a_i = a_{i+N}, i = \dots, -1, 0, 1, \dots$ выходная также будет периодической с тем же периодом N : $c_i = c_{i+N}, i = \dots, -1, 0, 1, \dots$. Тогда N

элементов c_0, c_1, \dots, c_{N-1} полностью определяют выходную последовательность, и c_i становится циклической сверткой

$$c_i = \sum_{l=0}^{N-1} a_{i-l} b_l, \quad i=0,1, \dots, N, \quad (6.36)$$

где вычитание в индексе осуществляется по модулю N .

Предъявим к фильтру следующие требования

$$c_0 \neq 0, \quad c_i = 0, \quad i=1,2, \dots, N-1, \quad (6.37)$$

физический смысл которых состоит в том, что выходной сигнал фильтра имеет ненулевой основной лепесток, повторяющийся с периодом $N\Delta$, тогда как все боковые лепестки между ними равны нулю. Подобный фильтр, называемый в дальнейшем *фильтром подавления боковых лепестков* (ФПБЛ) (*sidelobe suppression filter* (SLSF)), имитирует своим откликом идеальную периодическую АКФ. Поскольку для бинарных кодов идеальная АКФ не достижима (за исключением единственного тривиального сигнала), то ФПБЛ является рассогласованным и, следовательно, уступает согласованному фильтру в отношении сигнал-шум.

Кратчайшим путем получения точного выражения для коэффициентов фильтра служит применение дискретного преобразования Фурье (ДПФ). Последовательность $a_i, i=0,1, \dots, N-1$ и ее компоненты ДПФ спектра $\tilde{a}_k, k=0,1, \dots, N-1$ связаны друг с другом прямым и обратным ДПФ:

$$\tilde{a}_k = \sum_{i=0}^{N-1} a_i \exp\left(-j \frac{2\pi i k}{N}\right), \quad a_i = \frac{1}{N} \sum_{k=0}^{N-1} \tilde{a}_k \exp\left(j \frac{2\pi i k}{N}\right), \quad i, k=0,1, \dots, N-1.$$

Нашей целью является достижение дискретной дельта-функции (6.37) на выходе фильтра, имеющей только один ненулевой элемент на периоде. Спектр подобной функции является равномерным: $\tilde{c}_k = c_0, k=0,1, \dots, N-1$. Тогда вследствие теоремы о свертке [1] спектр последовательности (6.36) на выходе фильтра может быть представлен в виде $\tilde{c}_k = \tilde{a}_k \tilde{b}_k = c_0, k=0,1, \dots, N-1$, где спектр последовательности коэффициентов фильтра \tilde{b}_k есть ничто иное, как передаточная функция ФПБЛ:

$$\tilde{b}_k = \frac{c_0}{\tilde{a}_k}, \quad k=0,1, \dots, N-1. \quad (6.38)$$

Как видно, передаточная функция ФПБЛ обратно пропорциональна спектру сигнала, что послужило причиной присвоения фильтрам этого типа названия инверсных фильтров. Инверсный фильтр как раз и выравнивает входной спектр. Как показывает последнее соотношение, ФПБЛ физически реализуем для любой периодической последовательности, ДПФ спектр которой не имеет нулевых компонент. Применение обратного преобразования Фурье к (6.38) дает точное выражение для коэффициентов ФПБЛ

$$b_i = \frac{c_0}{N} \sum_{k=0}^{N-1} \frac{1}{\tilde{a}_k} \exp\left(j \frac{2\pi i k}{N}\right), \quad i=0,1, \dots, N-1. \quad (6.39)$$

6.12.2. Вычисление потерь в отношении сигнал шум.

В случае согласованного фильтра (см. рис.6.20) его коэффициентами будут служить (игнорируя несущественный общий множитель) значения символов зеркально ото-

браженной входной последовательности¹: $b_i = a_{N-i}$, $i = 0, 1, N-1$ и амплитуда выходной последовательности составит $A_{mf} = \sum_{i=0}^{N-1} a_i^2 = N$, поскольку входная последовательность бинарная. Для входного шума, обладающего временем корреляции в пределах Δ и дисперсией σ^2 , дисперсия на выходе согласованного фильтра составит величину $\sigma_{mf}^2 = \sigma^2 \sum_{i=0}^{N-1} b_i^2 = \sigma^2 \sum_{i=0}^{N-1} a_{N-i}^2 = N\sigma^2$. Таким образом, на выходе согласованного фильтра отношение сигнал-шум по мощности q_{mf}^2 составит

$$q_{mf}^2 = \frac{A_{mf}^2}{\sigma_{mf}^2} = \frac{N}{\sigma^2}. \quad (6.40)$$

Аналогичным образом, амплитуда выходной последовательности ФПБЛ составляет $A_{sl} = c_0$, а дисперсия шума

$$\sigma_{sl}^2 = \sigma^2 \sum_{i=0}^{N-1} b_i^2.$$

Согласно теореме Парсеваля и свойству временного сдвига ДПФ периодическая АКФ произвольной последовательности u_0, u_1, \dots, u_{N-1} периода N связана с энергетическим спектром последовательности $|\tilde{u}_0|^2, |\tilde{u}_1|^2, \dots, |\tilde{u}_{N-1}|^2$ с помощью обратного ДПФ:

$$R_p(m) = \sum_{i=0}^{N-1} u_i u_{i-m}^* = \frac{1}{N} \sum_{k=0}^{N-1} |\tilde{u}_k|^2 \exp\left(j \frac{2\pi mk}{N}\right), \quad m = 0, 1, \dots, N-1. \quad (6.41)$$

В частности

$$R_p(0) = \sum_{i=0}^{N-1} |u_i|^2 = \frac{1}{N} \sum_{k=0}^{N-1} |\tilde{u}_k|^2.$$

Использование последнего соотношения вместе (6.38) в выражении для дисперсии шума на выходе фильтра дает

$$\sigma_{sl}^2 = \frac{\sigma^2}{N} \sum_{k=0}^{N-1} |\tilde{b}_k|^2 = \frac{c_0^2 \sigma^2}{N} \sum_{k=0}^{N-1} \frac{1}{|\tilde{a}_k|^2},$$

и отношение сигнал-шум на выходе ФПБЛ принимает вид

$$q_{sl}^2 = \frac{A_{sl}^2}{\sigma_{sl}^2} = \frac{N}{\sigma^2} \left(\sum_{k=0}^{N-1} \frac{1}{|\tilde{a}_k|^2} \right)^{-1}.$$

Теперь можно оценить энергетические потери в ФПБЛ по сравнению с согласованным фильтром как

$$\gamma = \frac{q_{mf}^2}{q_{sl}^2} = \sum_{k=0}^{N-1} \frac{1}{|\tilde{a}_k|^2}. \quad (6.42)$$

¹ Циклический сдвиг коэффициентов по сравнению со случаем аperiodического сигнала ($b_i = a_{N-1-i}$) служит для более компактного представления (6.38) путем исключения линейной фазовой экспоненты. В случае периодического сигнала данный сдвиг означает циклический сдвиг выходного периодического сигнала и, следовательно, не сказывается на ее форме или отношении сигнал-шум.

Для лучшего понимания последнего результата отметим, что

$$N = \sum_{i=0}^{N-1} a_i^2 = \frac{1}{N} \sum_{k=0}^{N-1} |\tilde{a}_k|^2 = \overline{|\tilde{a}|^2} \quad \text{и} \quad \left(\frac{1}{N} \sum_{k=0}^{N-1} \frac{1}{|\tilde{a}_k|^2} \right)^{-1} = \left(\overline{|\tilde{a}|^{-2}} \right)^{-1}$$

отвечают соответственно среднему арифметическому и гармоническому энергетического спектра последовательности $|\tilde{a}_k|^2$, $k=0,1,\dots,N-1$. Среднее гармоническое любых неотрицательных чисел никогда не превосходит их среднего арифметического, и эти характеристики совпадают, если только все усредняемые числа одинаковы. Следовательно, отношение данных характеристик может служить некоторой мерой степени разброса усредняемых чисел. Однако в рассматриваемом случае это отношение

$$\frac{\overline{|\tilde{a}|^2}}{\left(\overline{|\tilde{a}|^{-2}} \right)^{-1}} = \sum_{k=0}^{N-1} \frac{1}{|\tilde{a}_k|^2} = \gamma$$

в точности соответствует энергетическим потерям в ФПБЛ. Следовательно, потери в отношении сигнал-шум γ определяются неравномерностью энергетического спектра последовательности $|\tilde{a}_k|^2$, $k=0,1,\dots,N-1$, оцениваемой в терминах различия между средним гармоническим и арифметическим.

Возможность подавления всех периодических боковых лепестков выдвигает новый критерий синтеза бинарных последовательностей, который является альтернативным минимизации максимального бокового лепестка $\rho_{p,\max}$. Действительно, какой смысл заботиться об уровне бокового лепестка, когда все боковые лепестки могут быть приведены к нулю? Более естественным представляется минимизация затрат, обуславливающих устранение боковых лепестков, и количественной мерой этих затрат, конечно, выступает потеря в отношении сигнал-шум γ . Рассмотрим вначале бинарную последовательность (гипотетическую, если $N \neq 4$), обладающую равномерным энергетическим спектром: $|\tilde{a}_k|^2 = N$, $k=0,1,\dots,N-1$. В свете соотношения (6.41) это означает, что последовательность обладает идеальной периодической АКФ. Абсолютно предсказуемо, что, поскольку отсутствует объект подавления, то в этом случае ФПБЛ совпадет с согласованным фильтром, не имеющим потерь в отношении сигнал-шум. Соотношение (6.42) подтверждает данный факт, определяя $\gamma=1$, или, выраженное в дБ, $\gamma_{дБ} = 10 \lg \gamma = 0$ дБ. Поскольку не существует бинарных последовательностей с идеальной периодической АКФ, то подавление их боковых лепестков производится за счет потерь в отношении сигнал-шум ($\gamma_{дБ} > 0$), оправдывая введение критерия синтеза $\gamma = \min$.

Как и во многих задачах, касающихся бинарных последовательностей (см. параграф 6.4), глобально оптимальная бинарная последовательность фиксированной длины N с минимальными потерями γ может быть найдена только в результате исчерпывающего поиска. Данная работа была выполнена для длин вплоть до $N=40$ [48]. Конечно, вследствие экспоненциального роста вычислительных затрат указанный поиск не может продолжаться далеко за указанный диапазон. Однако известны многие регулярные правила построения бинарных последовательностей сколь угодно большой длины с очень малыми потерями γ (хотя и без гарантии их глобальной оптимальности).

Рассмотрим специальный класс бинарных последовательностей, обладающих двухуровневой периодической АКФ, т.е. с постоянным уровнем R боковых лепестков

$$R_p(m) = \begin{cases} N, & m = 0 \bmod N, \\ R, & m \neq 0 \bmod N. \end{cases} \quad (6.43)$$

Все минимаксные последовательности, как и ряд других, относятся к данному типу. Энергетический спектр такой последовательности, как показывает (6.41), есть просто ДПФ АКФ:

$$|\tilde{a}_k|^2 = \sum_{m=0}^{N-1} R_p(m) \exp\left(-j \frac{2\pi mk}{N}\right) = N - R + R \sum_{m=0}^{N-1} \exp\left(-j \frac{2\pi mk}{N}\right), \quad k = 0, 1, \dots, N-1.$$

Последняя сумма уже рассматривалась в 6.11.2 и, как было доказано, она равняется N , если $k = 0$, и нулю в противном случае. Таким образом

$$|\tilde{a}_k|^2 = \begin{cases} N + (N-1)R, & k = 0, \\ N - R, & k \neq 0. \end{cases} \quad (6.44)$$

Подстановка последнего соотношения в (6.42) дает

$$\gamma = \frac{1}{N + (N-1)R} + \frac{N-1}{N-R} = \frac{1 + (N-2)\rho}{(1-\rho)[1 + (N-1)\rho]}, \quad (6.45)$$

где $\rho = R/N$ – нормированный уровень бокового лепестка АКФ.

Для определения структуры ФПБЛ, перепишем (6.39) как

$$b_i = \frac{c_0}{N} \sum_{k=0}^{N-1} \frac{\tilde{a}_k^*}{|\tilde{a}_k|^2} \exp\left(j \frac{2\pi ik}{N}\right), \quad i = 0, 1, \dots, N-1$$

и подставим (6.44) в последнее соотношение, тогда

$$b_i = \frac{c_0}{N} \left[\frac{\tilde{a}_0}{N + (N-1)R} - \frac{\tilde{a}_0}{N-R} + \frac{1}{N-R} \sum_{k=0}^{N-1} \tilde{a}_k^* \exp\left(j \frac{2\pi ik}{N}\right) \right], \quad i = 0, 1, \dots, N-1.$$

Сумма по k в данном соотношении определяет коэффициенты согласованного фильтра, поскольку она может быть записана как

$$\frac{1}{N} \sum_{k=0}^{N-1} \tilde{a}_k^* \exp\left(-j \frac{2\pi(N-i)k}{N}\right) = \left[\frac{1}{N} \sum_{k=0}^{N-1} \tilde{a}_k \exp\left(j \frac{2\pi(N-i)k}{N}\right) \right]^* = a_{N-i}, \quad i = 0, 1, \dots, N-1,$$

где использован тот факт, что элементы бинарной последовательности вещественны. Поскольку c_0 произвольный множитель, положим его равным $N - R$. Тогда

$$b_i = a_{N-i} - \frac{\rho \tilde{a}_0}{1 + (N-1)\rho}, \quad i = 0, 1, \dots, N-1. \quad (6.46)$$

Первое слагаемое в данном выражении соответствует последовательности $\{a_i\}$, считываемой справа налево, т.е. коэффициентам согласованного фильтра. Таким образом, для последовательности, обладающей двухуровневой АКФ (6.43), ФПБЛ получается путем незначительного изменения согласованного фильтра: вычитанием некоторой постоянной величины из всех коэффициентов. Более того, для бинарных последовательностей данного типа коэффициенты ФПБЛ принимают только два возможных значения $\pm 1 - \frac{\rho \tilde{a}_0}{1 + (N-1)\rho}$,

где \tilde{a}_0 – постоянная составляющая последовательности, т.е. разность между числом плюс и минус единиц на периоде: $\tilde{a}_0 = N_+ - N_-$.

Как известно из предыдущего материала, существует множество бинарных последовательностей с АКФ вида (6.43), для которых $R = -1$ (m – последовательности, последовательности Лежандра и другие минимаксные последовательности с АКФ вида (6.12)). Оценка их потерь в ФПБЛ согласно (6.45) дает $\gamma = 2N/(N+1)$, т.е. $\gamma \approx 2$ (3 дБ) для практически интересных длин. Отсюда видно, что наиболее популярные минимаксные бинарные

ные последовательности не играют значительной роли в свете критерия потерь в ФПБЛ: половина их энергии теряется при обработке с помощью ФПБЛ.

С другой стороны, при положительном R , достаточно малым в сравнении с N , $\gamma < 1/(1-\rho)$, т.е. потери оказываются достаточно малыми. Так называемые коды Зингера служат хорошим примером подобных бинарных последовательностей. Код Зингера существует при любой длине вида $N = \frac{q^n - 1}{q - 1}$ и обладает двухуровневой АКФ типа (6.43) со

значением $R = N - 4q^{n-2}$, где $q = p^w$ – натуральная степень простого p , n – натуральное число. Наиболее интересная с рассматриваемых позиций модификация кодов Зингера от-

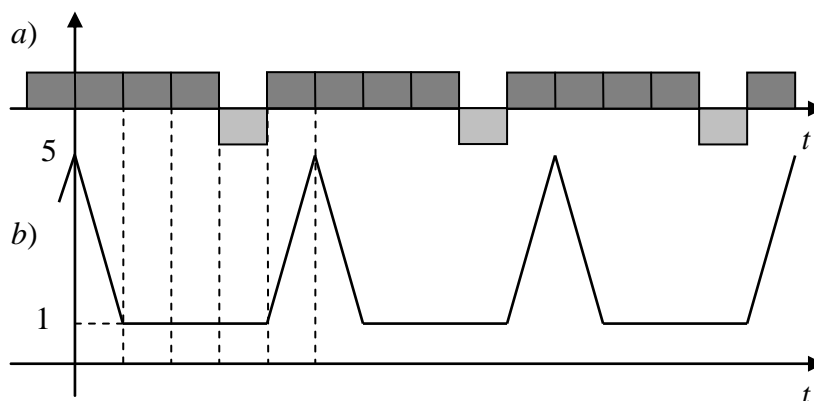


Рис. 6.21. Бинарный сигнал длины $N = 5$ и его периодическая АКФ.

вечает $q = 3$, для которой $\rho = \frac{3^{n-2} - 1}{3^n - 1}$ и $\gamma < \frac{3^n - 1}{8 \cdot 3^{n-2}} < \frac{9}{8} = 1.125$, т.е. $\gamma_{дБ} < 0.51$ дБ. Как вид-

но, данные коды достаточно привлекательны при обработке их ФПБЛ, поскольку потери в отношении сигнал-шум, сопровождающие полное подавление боковых лепестков, малы.

Пример 6.12.1. Рассмотрим периодическую версию бинарного кода Баркера длины $N = 5$ из таблицы 6.1: $+1, +1, +1, -1, +1$, для которого $N_+ = 4$, $N_- = 1$, а постоянная состав-

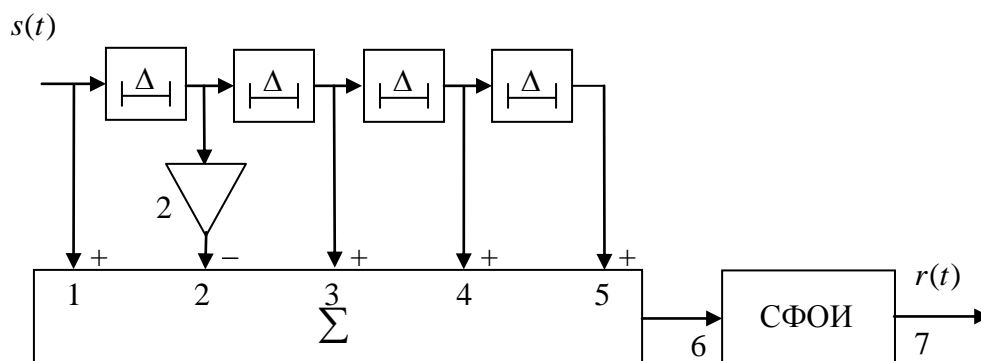


Рис. 6.22 ФПБЛ для последовательности длины $N=5$.

ляющая $\tilde{a}_0 = 3$. Его периодическая АКФ, что может быть проверено непосредственной проверкой, удовлетворяет соотношению (6.43) со значением $R = 1$ ($\rho = 1/5$). В действи-

тельности данная последовательность представляет собой код Зингера с параметрами $q = 4, n = 2$. Рис. 6.21, *b* демонстрирует вид АКФ (т.е. отклик согласованного фильтра) периодического сигнала, модулированного данной последовательностью (рис. 6.21, *a*). Как следует из (6.46), согласованный фильтр для данной последовательности легко трансформируется в ФПБЛ изменением всех коэффициентов $+1$ на $+2/3$ и -1 на $-4/3$, которое после соответствующего масштабирования означает замену -1 на -2 при оставлении без

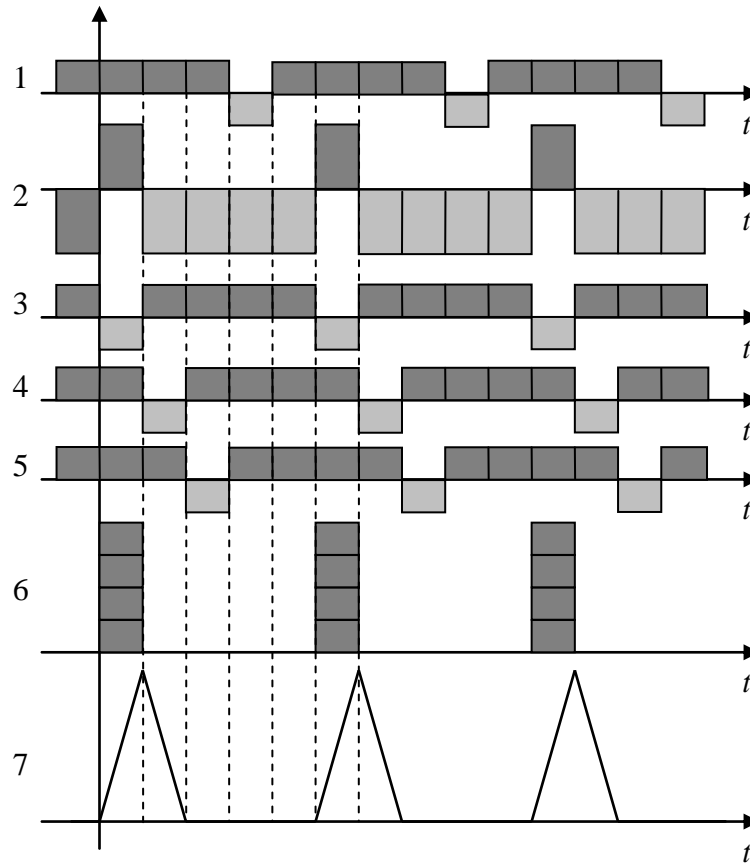


Рис. 6.23 Отклик ФПБЛ для бинарной последовательности $N=5$.

изменения $+1$ (см. рис. 6.22). Отклик ФПБЛ на тот же сигнал построен на рис. 6.23, на котором оцифровка диаграмм соответствует точкам на рис. 6.22. Выходной сигнал фильтра имеет желаемую форму, т.е. нулевой уровень боковых лепестков. Используя соотношение (6.45), легко определить энергетические потери в ФПБЛ как $\gamma = 10/9 = 1.1111\dots$ (0.46 дБ).□

Основываясь, как и для кодов Зингера, на отображении линейных рекуррентных последовательностей над конечными полями на бинарный $\{\pm 1\}$ алфавит, было найдено множество даже более эффективных бинарных кодов. Все они характеризуются одинаково привлекательными чертами: простотой структуры ФПБЛ, коэффициенты которого принимают не более трех различных значений. Не углубляясь далее в детали, которые достаточно сложны, и отсылая заинтересованного читателя к [49-50], отметим только, что среди упомянутых последовательностей имеются семейства, обладающие асимптотически малыми потерями в ФПБЛ: $\gamma_{\text{дБ}} \rightarrow 0$ дБ при $N \rightarrow \infty$.

6.13. ЧМ сигналы с оптимальной аperiodической АКФ.

Краткое обсуждение сути построения ЧМ сигналов с хорошими корреляционными свойствами начнем с соотношения (5.20), напоминающего о том, что требование малого

уровня $R_p(m)$ эквивалентно минимизации числа совпадающих частот в кодовой последовательности частот F_0, F_1, \dots, F_{N-1} и ее копии, сдвинутой на m позиций. Очевидно, что если число чипов (элементарных импульсов) N (т.е. длина) не превосходит числа различных частот M , получение нулевого уровня боковых лепестков $\rho_a(m)$, $m \neq 0$ может быть достигнуто тривиальным путем: использованием частотного кода, все элементы которого F_i различны. На практике, однако, типичной является ситуация, когда $N > M$, что влечет за собой повторения элементов F_i , $i = 0, 1, \dots, N-1$ и, следовательно, по меньшей мере, одно совпадение в сдвинутых копиях частотного кода, т.е. $\rho_{a, \max} \geq 1/N$.

Поскольку частотная последовательность может быть описана с помощью массива (решетки) размера $M \times N$ (см. параграф 5.5), то минимизация $\rho_{a, \max}$ означает построение решетки с минимально возможным числом совпадений помеченных элементов в исходной решетке и ее копии, сдвинутой по горизонтали на m позиций. Одной из ключевых проблем является конструирование т.н. *радарных решеток* (*radar arrays*), определяемых как $M \times N$ массивы, имеющие только один помеченный элемент в каждом столбце и $\rho_{a, \max} = 1/N$, т.е. с числом упомянутых совпадений не более одного. Данное стремление понимается как отыскание максимально длинной радарной решетки с фиксированным значением M , поскольку это означает минимизацию $\rho_{a, \max}$ при ограничениях на частотный ресурс. Следуя [51], попытаемся доказать простейшую верхнюю границу длины радарной решетки.

Рассмотрим последовательность F_0, F_1, \dots, F_{N-1} и отметим, что для получения не более одного совпадения все разности между номерами позиций, отвечающих одинаковым частотам, должны быть различными. Действительно, пусть $F_i = F_k$, $F_s = F_t$ и $i - k = s - t > 0$. Тогда в исходной последовательности и ее копии, сдвинутой на $m = i - k = s - t$ позиций, произойдет не менее двух совпадений. Обозначим через n_i число символов (частот) среди F_0, F_1, \dots, F_{N-1} , повторяющихся i раз. Тогда

$$\sum_i i n_i = N \text{ и } \sum_i n_i = M. \quad (6.47)$$

Подсчитаем теперь число возможных разностей между номерами позиций, содержащих идентичные частоты. Если некоторая частота повторяется i раз, то число подобных разностей для этой частоты будет $i(i-1)$. Поскольку в последовательности содержится n_i частот, повторяющихся i раз, то общее число указанных разностей составит величину $\sum_i i(i-1)n_i$, и так как ни одна из разностей не должна повторяться, то

$$\sum_i i(i-1)n_i \leq N-1, \quad (6.48)$$

где правая часть неравенства дает максимальное число не равных положительных разностей среди номеров $\{0, 1, \dots, N-1\}$. Соотношение $i(i-1) + 3 - 2i = i^2 - 3i + 3$ не имеет вещественных корней и, следовательно, положительно при любом i . Поэтому сумма

$$\sum_i [i(i-1) + 3 - 2i]n_i = \sum_i i(i-1)n_i + 3\sum_i n_i - 2\sum_i i n_i \geq 0,$$

что, с учетом (6.47), (6.48), может быть преобразовано к виду $N - 1 + 3M - 2N \geq 0$, или

$$N \leq 3M - 1. \quad (6.49)$$

В действительности эта граница не является строгой. Известны более точные границы, например в [52], асимптотический вариант которых приводит к результату

$$N \leq \frac{20 + \sqrt{6}}{8} M, \quad M \gg 1, \quad (6.50)$$

что снижает правую часть (6.49) приблизительно на $0.194M$.

Для значений $M \leq 16$ известны абсолютно точные, т.е. действительно достижимые, верхние границы длины N . Для указанного диапазона M максимальная длина N_{\max} радарных решеток ограничена сверху соотношениями [52]

$$N_{\max} = \begin{cases} 3M - 2, & 2 \leq M \leq 4, \\ 3M - 3, & 5 \leq M \leq 9, \\ 3M - 4, & 10 \leq M \leq 13, \\ 3M - 5, & 14 \leq M \leq 16. \end{cases} \quad (6.51)$$

Пример 6.13.1. Частотный код $\{1, 2, 3, 4, 5, 6, 7, 8, 7, 4, 3, 9, 9, 5, 8, 2, 6, 5, 1, 4, 2, 1, 3, 7\}$, который определяет номера частот из некоторого алфавита с $M = 9$ или, что эквивалентно, номера строк в каждой колонке решетки, имеет максимально возможную длину $N = 24$. Свойства радарной решетки, т.е. наличие единственного совпадения частот при всех ненулевых сдвигах, может быть проверено путем непосредственной проверки (задача 6.53). \square

Кроме того, известны регулярные правила построения радарных решеток длины $N = 2.5M$ для любого четного M (см. [51]) и для $M/2$, равного произведению простых чисел, дающего остаток один от деления на 4, т.е. $M = 10, 26, 34, 58, \dots$

Сонарные (sonar) решетки представляют собой обобщение радарных, сохраняющие свойство «не более одного совпадения» для произвольной комбинации горизонтального и вертикального сдвига [53]. Практически это требование отражает стремление иметь слабую корреляцию сигнальных копий, сдвинутых как во времени, так и по частоте. Рассматривая подход к выбору частотного пространства для ЧМ сигналов (параграф 5.5), отмечалось, что частотные сдвиги, вызывающие переход из текущего значения частоты к соседнему, более типичны для сонарных, чем локационных систем. Известно, например, множество регулярных алгоритмов построения массивов Костаса [54], т.е. квадратных ($M = N$) сонарных массивов или ЧМ последовательностей, имеющих одинаковую длину и число различных частот.

6.11. Последовательности с идеальной периодической АКФ.

Как уже неоднократно указывалось, существуют многочисленные практические приложения, в которых периодичность используемых сигналов выдвигает на первый план качество их периодических корреляционных свойств. Другими словами, хорошая периодическая АКФ выполняет не только роль полезного вспомогательного инструмента для синтеза хороших аperiodических последовательностей, но и ценна самостоятельно. Примеры подобного рода дают дальномерные системы с непрерывным излучением, в особенности на больших расстояниях, пилотный канал и канал синхронизации в цифровых системах передачи данных (пилотные каналы «вниз» стандартов cdmaOne и cdma2000, вторичный канал синхронизации стандарта UMTS), радарные и сонарные системы с непрерывным излучением и т.п.

Несмотря на то, что бинарные $\{\pm 1\}$ минимаксные последовательности выглядят достаточно привлекательно, обладая максимальным периодическим боковым лепестком $\rho_{p, \max} = 1/N$, падающим с ростом длины, достаточно вероятны ситуации, когда приемлемое значение $\rho_{p, \max}$ требует фантастически большой длины N . Например, для локационных дальномерных и сонарных систем требование временного разрешения сигналов в динамическом диапазоне, превышающем 80 дБ, является достаточно обычным. Для выполнения данного условия требуются оптимальные бинарные последовательности длины, превышающей 10^4 , что может неоправданно замедлить начальную процедуру поиска (см. главу 8). Очевидно, что для многих подобных сценариев наилучшим выходом могла бы служить идеальная периодическая АКФ (6.6), которая, к сожалению, недостижима на множестве бинарных кодов, наиболее привлекательных с точки зрения практической реализации. В дальнейшей части параграфа будут проанализированы возможные различные пути достижения идеальной периодической АКФ для случаев, когда алфавит последовательности не лимитирован жестким требованием бинарности символов $\{\pm 1\}$.

6.11.1. Бинарные последовательности с не противоположной модуляцией.

Замена противоположного алфавита $\{+1, -1\}$ на некоторый не противоположный бинарный открывает возможность обратить все периодические боковые лепестки любой бинарной минимаксной последовательности, удовлетворяющей (6.12), в нуль. Простейшим путем определения необходимого алфавита является добавление константы c (комплексной в общем случае) к исходному алфавиту $\{+1, -1\}$ последовательности a_0, a_1, \dots, a_{N-1} , заменяя, тем самым, символы $+1$ и -1 на $1+c$ и $-1+c$ соответственно. Периодическая АКФ полученной таким образом последовательности может быть найдена непосредственным подсчетом

$$R_p(m) = \sum_{i=0}^{N-1} (a_i + c)(a_{i-m} + c^*) = \sum_{i=0}^{N-1} a_i a_{i-m} + 2\operatorname{Re}(c\tilde{a}_0) + N|c|^2, \quad (6.28)$$

где $\tilde{a}_0 = \sum_{i=0}^{N-1} a_i$, как и прежде, постоянная составляющая исходной последовательности

a_0, a_1, \dots, a_{N-1} . Соотношение (6.8) показывает, что для любой минимаксной последовательности, удовлетворяющей (6.12), $|\tilde{a}_0|^2 = \sum_{m=0}^{N-1} R_p(m) = N + (N-1)(-1) = 1 \Rightarrow \tilde{a}_0 = \pm 1$. По-

скольку изменение знака всех элементов не затрагивает АКФ, то можно рассматривать только последовательности с $\tilde{a}_0 = -1$. Учтем еще раз, что для любой минимаксной последовательности, удовлетворяющей (6.12), первая сумма в правой части (6.28) равна -1 при

любом $m \neq 0 \pmod N$. Тогда, полагая боковые лепестки последовательности $a_i + c, i = \dots, -1, 0, 1, \dots$ равными нулю, приходим к уравнению для комплексной неизвестной c :

$$|c|^2 + \frac{2}{N} \operatorname{Re}(c\tilde{a}_0) - \frac{1}{N} = |c|^2 - \frac{2}{N} \operatorname{Re}(c) - \frac{1}{N} = 0. \quad (6.29)$$

Данное уравнение при двух вещественных неизвестных (реальной и мнимой частях c) имеет бесконечное множество решений. Найдём потенциально наиболее интересные решения. Если желателен вещественный алфавит, т.е. $\operatorname{Re}(c) = c$ и $|c|^2 = c^2$, то (6.29) превращается в квадратичное уравнение

$$c^2 - \frac{2}{N}c - \frac{1}{N} = 0$$

с корнями $c_{1,2} = \frac{1 \pm \sqrt{N+1}}{N}$. Новые бинарные не противоположные символы $1+c$ и $-1+c$

теперь можно поделить на $1+c$ с целью сохранения $+1$ в качестве символа нового алфавита. После подобной операции приходим к следующему правилу преобразования бинарной минимаксной последовательности с периодической АКФ вида (6.12) в новую с идеальной АКФ: элементы, отвечающие -1 , следует изменить на

$$\frac{-1+c}{1+c} = \frac{-N+1 \pm \sqrt{N+1}}{N+1 \pm \sqrt{N+1}} = -1 \pm \frac{2}{\sqrt{N+1}},$$

а элементы $+1$ остаются без изменения.

Пример 6.11.1. Последовательность Лежандра или m – последовательность длины $N=127$ трансформируется в последовательность с идеальной периодической АКФ путем замены элементов -1 на $-1 \pm \frac{1}{4\sqrt{2}}$. \square

Вышеприведенное решение приводит к алфавиту из двух противоположных символов различной амплитуды, т.е. к амплитудной модуляции (рис. 6.18, а). Другим воз-

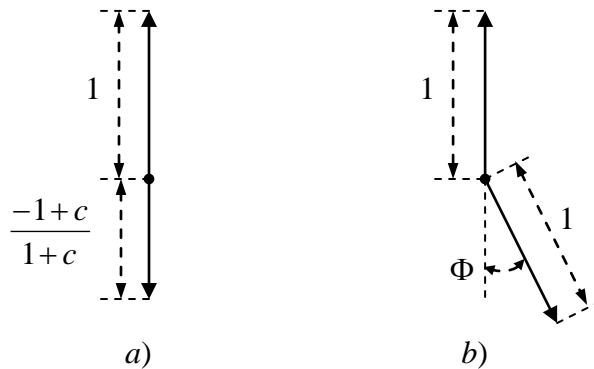


Рис.6.18. Не противоположный бинарный алфавит.

можным вариантом служит ФМ не противоположный алфавит, который может быть определен в явном виде, если представить c как мнимое число $c = jc_1$. Тогда (6.29) имеет

следующие решения $c_1 = \pm \frac{j}{\sqrt{N}}$, а новые символы $1 \pm \frac{j}{\sqrt{N}}$ и $-1 \pm \frac{j}{\sqrt{N}}$ после деления на

$1 \pm \frac{j}{\sqrt{N}}$ станут 1 и $\frac{-\sqrt{N} \pm j}{\sqrt{N} \pm j} = -\frac{N-1}{N+1} \pm \frac{2j\sqrt{N}}{N+1} = -\exp(j\Phi)$, где $\cos \Phi = \frac{N-1}{N+1}$ (рис.6.18, б).

Пример 6.11.2. Для $N = 127$ $\cos \Phi = 63/64$ и $\Phi = \pm \arccos(63/64) \approx \pm 10^\circ 8' 30''$. Заменяя все отрицательные элементы бинарной m – последовательности или последовательности Лежандра длины $N = 127$ на $-\exp(j\Phi)$, получаем последовательность с идеальной периодической АКФ. \square

Только что рассмотренный скромный метод, который неоднократно предлагался и повторялся многими авторами [38, 39], с трудом может быть признан как эффективный с практической точки зрения. Как видно и что подтверждается примерами, он предписывает использование экзотических значений комплексных амплитуд кода, установка и поддержание которых с требуемой точностью может оказаться чрезмерно затрудненной на практике.

6.11.2. Многофазные коды.

Привлечение недвоичной фазовой модуляции с $M > 2$ открывает путь к многочисленным многофазным последовательностям с идеальной периодической АКФ. Известны различные правила их конструирования, однако в большей или меньшей степени все они являются производными двух наиболее распространенных алгоритмов. Первый из них, соответствующий кодам Чу (или квадратичных вычетов), вытекает непосредственным образом и аппроксимирует в дискретной форме закон линейной частотной модуляции (см. 6.2). Коды Чу существуют при произвольном значении длины N и формируются как

$$a_i = \begin{cases} \exp\left(\frac{j\pi i^2}{N}\right), & N - \text{четное}, \\ \exp\left(\frac{j2\pi i^2}{N}\right), & N - \text{нечетное}, \end{cases} \quad (6.30)$$

где $i = \dots, -1, 0, 1, \dots$.

Легко проверить, что $a_i = a_{i+N}$ для всех i и, значит, N – по крайней мере, кратно периоду кода. В процессе вычисления периодической АКФ окончательно прояснится значение периода. Для кода четной длины ненормированная периодическая АКФ определяется в виде

$$R_p(m) = \sum_{i=0}^{N-1} a_i a_{i-m}^* = \exp\left(-\frac{j\pi m^2}{N}\right) \sum_{i=0}^{N-1} \exp\left(\frac{j2\pi im}{N}\right).$$

При $m = 0 \bmod N$ последняя сумма равна N , тогда как коэффициент, стоящий перед ней обращается в 1. Для любого другого m $\exp(j2\pi im/N)$ зависит от i , а упомянутая выше сумма представляет собой сумму корней из единицы некоторой степени, или, что эквивалентно, геометрическую прогрессию с коэффициентом $\exp(j2\pi m/N)$. Вычислив сумму прогрессии, получаем

$$R_p(m) = \exp\left(-\frac{j\pi m^2}{N}\right) \frac{1 - \exp(j2\pi m)}{1 - \exp\left(\frac{j2\pi m}{N}\right)}.$$

Знаменатель последней дроби никогда не обращается в нуль за исключением $m = 0 \bmod N$ и, следовательно, $R_p(m) = 0$ при всех сдвигах, не кратных N . Таким образом, коды Чу, определяемые первой строкой в (6.30), обладают периодом N и имеют идеальную периодическую АКФ. Аналогичным образом осуществляется доказательство и для нечетного значения N (см. задачу 6.29).

Несмотря на то, что коды Чу служат достаточно убедительным академическим примером ФМ последовательностей с идеальной АКФ, их практическая реализация вызы-

вает обоснованные сомнения, поскольку размер фазового алфавита линейно растет с увеличением длины и расстояние между соседними фазами становится чрезвычайно малым. Этим обстоятельством обусловлена возрастающая требовательность к точности формирования символов кода, качеству воспроизведения фаз, условиям эксплуатации и т.п.

Аналогичные недостатки (хотя и в несколько меньшей степени) характерны для второго популярного семейства многофазных кодов: кодов Франка. Они также осуществляют пошаговую аппроксимацию линейной частотной модуляции, однако значительно более грубую, и существуют только при значениях длин, представляющих квадрат целого числа $N = h^2 = 4, 9, 16, 25, 36, 49, \dots$. Правило их формирования описывается соотношением

$$a_i = \exp\left(\frac{j2\pi i}{h} \left[\frac{i}{h}\right]\right), \quad i = \dots, -1, 0, 1, \dots, \quad (6.31)$$

где, как обычно, $[x]$ обозначает округление неотрицательного x в меньшую сторону.

Доказательство идеальности периодических корреляционных свойств кодов Франка отличается от ранее выполненного только незначительными деталями и составляет суть задачи 6.30. Из сравнения (6.31) и (6.30) очевидным образом следует, что фазовая градация кодов Франка уменьшается \sqrt{N} раз, так что увеличение объема алфавита с ростом N происходит значительно медленнее.

Пример 6.11.3. Возьмем $N = 4 \Rightarrow h = 2$. Тогда с приведением значений фаз к интервалу $[0, 2\pi]$ имеем $\frac{2\pi i}{h} \left[\frac{i}{h}\right] = \pi i \left[\frac{i}{2}\right] = 0, 0, 0, \pi$, $i = 0, 1, 2, 3$, и код Франка вида $+1, +1, +1, -1$ является единственным бинарным кодом с идеальной АКФ. \square

Пример 6.11.4. Если $N = 16$ $h = 4$, а фазовый алфавит состоит из 4 символов $\{\pm 1, \pm j\}$, то код Франка данной длины использует QPSK. Поскольку

$$\frac{2\pi i}{h} \left[\frac{i}{h}\right] = \frac{\pi i}{2} \left[\frac{i}{4}\right] = 0, 0, 0, 0, 0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}, 0, \pi, 0, \pi, 0, \frac{3\pi}{2}, \pi, \frac{\pi}{2}, \quad i = 0, 1, \dots, 15,$$

то код имеет вид $+1, +1, +1, +1, +1, +j, -1, -j, +1, -1, +1, -1, +1, -j, -1, +j$. Идеальность его периодической АКФ может быть проверена непосредственным вычислением. \square

Завершая рассмотрение многофазных кодов, отметим еще раз, что технологически они не настолько привлекательны в сравнении с бинарными противоположными кодами. Существуют ли коды, которые не уступают бинарным в простоте практической реализации, но в отличие от них обладают идеальной периодической АКФ? Ответ на этот вопрос будет дан в следующем параграфе.

6.11.3. Троичные последовательности.

Рассмотрим последовательность, элементы a_i которой могут принимать в дополнение к бинарным значениям ± 1 еще и нулевое значение. Другими словами, используется троичный алфавит $\{-1, 0, 1\}$, который с практической точки зрения означает комбинирование бинарной ФМ с паузами, т.е. интервалами времени, в течение которых отсутствует передача символов. Совершенно очевидно, что расширение бинарного алфавита $\{\pm 1\}$ до троичного $\{-1, 0, 1\}$ не приведет к серьезному усложнению схем формирования и обработки, но, как будет показано ниже, откроет путь к получению последовательностей с идеальными периодическими корреляционными свойствами. Напомним, что одной из основных причин проявления интереса к расширению спектра в задачах временного измерения и разрешения служит стремление к достижению высоких показателей при низкой пиковой мощности, т.е. при распределении энергии сигнала на большом временном интервале. В этих условиях совершенно оправданно использовать в качестве показателя эффективности распределения энергии во времени величину пик-фактора ν (см. параграф 2.7.1), т.е. от-

ношение пиковой и средней мощностей. Для любой ФМ, и в частности бинарной, энергия последовательности сигнала равномерно распределена на периоде, так что пиковая и средняя мощности одинакова и, значит, $v = 1$. Введение N_p пауз на периоде последовательности N , как это имеет место при троичном алфавите, нарушит равномерность распределения энергии и увеличит пик-фактор в $v = \frac{N}{N - N_p}$ раз. Следовательно, целевой

функцией синтеза является построение троичных последовательностей, обладающих не только идеальной периодической АКФ, но и малым числом нулей N_p на периоде, т.е. пик-фактором, незначительно превышающим единицу. Без введения подобного ограничения задача становится вырожденной и имеет тривиальное решение: код с одним только ненулевым символом на периоде N , соответствующим одиночному чипу, повторяющемуся с периодом $N\Delta$, обладает, несомненно, идеальной периодической АКФ.

В настоящее время известен целый ряд правил, позволяющих генерировать троичные последовательности с только что упомянутыми свойствами. Наиболее мощное из них формирует последовательности длиной и значением пик-фактора, устанавливаемыми следующими соотношениями

$$N = \frac{q^n - 1}{q - 1}, \quad v = \frac{q^n - 1}{q^n - q^{n-1}} < \frac{q}{q - 1}, \quad (6.32)$$

где $q = p^w$ – натуральная степень простого числа p , n – нечетное. Последовательности этого типа определены при любой комбинации q, n в пределах указанных условий и, следовательно, выбирая достаточно большое q , можно добиться сколь угодно близкого к единице значения пик-фактора.

Построение троичных последовательностей, удовлетворяющих (6.32), основано на некоторых специфических свойствах полей Галуа. Простейшие из них и в то же время, охватывающие большинство длин, определяемых (6.32), отвечают случаю нечетного p ($q = p^w, p > 2$) [40, 41]. Для представления идеи в наиболее явном виде рассмотрим детальное описание алгоритма только для случая $q = p$, т.е. $w = 1$. Наиболее просто это возможно сделать с помощью привлечения p -ичных m -последовательностей.

Пусть $d_i, i = \dots, -1, 0, 1, \dots$ – p -ичная m -последовательность, где p – простое нечетное число. Каждый символ последовательности является элементом простого поля $GF(p)$. Преобразуем последовательность в троичную, отображая нулевой элемент в вещественный нуль, а ненулевые элементы в их двузначные характеры. После подобного преобразования изменим знаки всех элементов, стоящих на нечетных позициях. Формально, предложенный алгоритм может быть представлен следующим соотношением

$$a_i = \begin{cases} (-1)^i \psi(d_i), & d_i \neq 0, \\ 0, & d_i = 0, \end{cases} \quad (6.33)$$

где $i = \dots, -1, 0, 1, \dots$. На рис. 6.19 представлена структура, реализующая данное правило, включая генератор m -последовательности, блок отображения элементов m -последовательности в значения характера или нуль, и умножитель, обеспечивающий изменение полярности.

Для вычисления величины пик-фактора троичной последовательности (6.33) достаточно вспомнить, что период m -последовательности составляет $L = p^n - 1$, а свойство сбалансированности утверждает, что на одном периоде содержится $L_0 = p^{n-1} - 1$ нулей. Все они и никакие другие образуют нули в троичной последовательности, следовательно,

на периодическом сегменте из L элементов троичной последовательности ровно L_0 элементов составляют нули, откуда величина пик-фактора будет

$$v = \frac{L}{L-L_0} = \frac{p^n - 1}{p^n - p^{n-1}} < \frac{p}{p-1},$$

что совпадает с (6.32) при $q = p$. Доказательство того, что последовательность (6.33) обладает периодом, устанавливаемым соотношением (6.32), и идеальной периодической АКФ требует привлечения еще одного свойства псевдослучайности m -последовательностей, доказательство которого заинтересованный читатель может найти в [42]. Для фор-

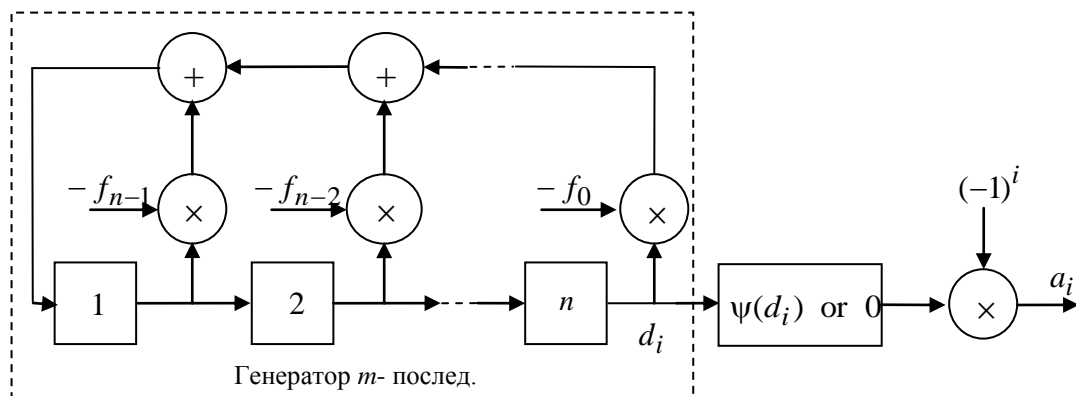


Рис.6.19. Генератор троичной последовательности.

мулирования упомянутого свойства введем обозначение $h = \frac{L}{p-1} = \frac{p^n - 1}{p-1}$ и рассмотрим все пары (d_i, d_{i-m}) элементов p -ичной m -последовательности, разделенные m позициями, при пробегании i интервала одного периода ($i = 0, 1, \dots, L-1$). Тогда (свойство парности), если m не кратно h ($m \neq lh$ для некоторого целого l), то среди пар (d_i, d_{i-m}) пара вида $(0, 0)$ встречается $p^{n-2} - 1$ раз, а любая другая пара (x, y) фиксированных значений $x, y \in GF(p) - p^{n-2}$ раз. В противном случае, если $m = lh$, то в парах (d_i, d_{i-m}) второй элемент строго определяется первым: $d_{i-m} = \alpha^l d_i$, где α , как обычно, примитивный элемент поля $GF(p)$.

С учетом того, что «истинный» (т.е. до сих пор неизвестный) период N троичной последовательности (6.33) есть некоторый делитель периода L исходной m -последовательности, вычислим не нормированную периодическую АКФ троичной последовательности на интервале L , содержащем L/N периодов:

$$R_p(m) = \frac{N}{L} \sum_{i=0}^{L-1} a_i a_{i-m} = (-1)^m \frac{N}{L} \sum_{\substack{i=0 \\ d_i \neq 0 \\ d_{i-m} \neq 0}}^{L-1} \psi(d_i) \psi(d_{i-m}), \quad (6.34)$$

где в последней сумме отброшены слагаемые, для которых $d_i d_{i-m} = 0$, как вносящие нулевой вклад. Рассмотрим первоначально случай, когда сдвиг m не кратен h ($m \neq lh$). Тогда, в соответствие со свойством парности, среди всех пар (d_i, d_{i-m}) в (6.34) любая пара (x, y) ненулевых фиксированных $x, y \in GF(p)$ встречается ровно p^{n-2} раз. Данный факт позволяет вычислить (6.34) следующим образом

$$R_p(m) = (-1)^m p^{n-2} \frac{N}{L} \sum_{x=1}^{p-1} \sum_{y=1}^{p-1} \psi(x)\psi(y) = (-1)^m p^{n-2} \frac{N}{L} \sum_{x=1}^{p-1} \psi(x) \sum_{y=1}^{p-1} \psi(y) = 0, m \neq lh \quad (6.35)$$

вследствие свойства характера (6.21). Обратимся теперь к случаю, когда величина сдвига делится на h ($m = lh$). Тогда согласно свойству парности в сумму (6.34) входят только пары вида $(d_i, d_{i-m}) = (x, \alpha^l x)$, $x \in GF(p)$. Однако, в соответствие со свойством сбалансированности каждый период p -ичной m -последовательности содержит ровно p^{n-1} каждого из ненулевых фиксированных элементов $GF(p)$. Следовательно, используя свойство мультипликативности характеров (6.20), получаем

$$R_p(lh) = (-1)^{lh} p^{n-1} \frac{N}{L} \sum_{x=1}^{p-1} \psi(x)\psi(\alpha^l x) = (-1)^{lh} p^{n-1} \psi(\alpha^l) \frac{N}{L} \sum_{x=1}^{p-1} \psi(x^2)$$

и

$$R_p(lh) = (-1)^{l(h+1)} p^{n-1} (p-1) \frac{N}{L},$$

поскольку $\psi(x^2) = 1$ для любого ненулевого $x \in GF(p)$, а $\psi(\alpha^l) = (-1)^l$ согласно определению (6.18). Так как n – нечетно, то $h = \frac{p^n - 1}{p - 1} = p^{n-1} + p^{n-2} + \dots + 1$ есть сумма нечетного числа нечетных целых и, следовательно, сама нечетна. По этой причине $l(h+1)$ является четным числом вне зависимости от l и, значит, $R_p(lh) = p^{n-1} (p-1) \frac{N}{L}$. Отсюда видно, что значение $R_p(lh)$ одно и то же для любого целого l , тогда как из (6.35) следует, что $R_p(m) = 0$ при $m \neq lh$. Таким образом, $R_p(m)$ как функция от m повторяется с периодом h и, следовательно, истинный период троичной последовательности определяется как $N = h = \frac{L}{p-1} = \frac{p^n - 1}{p - 1}$ в полном соответствии с предсказанным (6.32). В результате приходим к окончательному результату, демонстрирующему идеальность периодической АКФ

$$R_p(m) = \begin{cases} p^{n-1}, & m \neq 0 \pmod{N}, \\ 0, & m = 0 \pmod{N}, \end{cases}$$

где $N = \frac{p^n - 1}{p - 1}$.

Пример 6.11.5. Пусть $p = 3$, $n = 3$, что означает $N = 26/2 = 13$. Для построения троичной последовательности данного периода воспользуемся троичной m -последовательностью из примера 6.6.2: 1,0,0,2,0,2,1,2,2,1,0,2,2,2,0,0,1,0,1,2,1,1,2,0,1,1,... В поле $GF(3)$ имеются только два ненулевых элемента, из которых только 2 является примитивным. Очевидно, что $\psi(1) = 1$, $\psi(2) = -1$ и, следовательно, все ненулевые элементы m -последовательности заменяются, как $1 \rightarrow +1$, $2 \rightarrow -1$, и нули отображаются в вещественный нуль. В результате получаем троичную последовательность периода 26

$$+1,0,0,-1,0,-1,+1,-1,-1,+1,0,-1,-1,-1,0,0,+1,0,+1,-1,+1,+1,-1,0,+1,+1,\dots$$

Замена знака у элементов, стоящих на нечетных позициях (начиная индексацию с нуля), дает результирующую троичную последовательность вида

$$+1,0,0,+1,0,+1,+1,+1,-1,-1,0,+1,-1,+1,0,0,+1,0,+1,+1,+1,-1,-1,0,+1,-1,\dots,$$

имеющую период $N = 13$ и пик-фактор $v = 13/9 \approx 1.445$. Идеальность ее периодической

АКФ может быть проверена непосредственным вычислением. \square

Можно исключить операцию чередования знака у элементов с нечетными значениями позиций в правиле (6.33), а в генераторе на рис. 6.19 вместо m – последовательностей использовать некоторые специальные линейные последовательности меньшего периода. С этой целью коэффициенты f_i в рекурсии (6.13) и обратной связи LFSR генератора должны принадлежать соответствующему не примитивному неприводимому полиному степени n . Теоретическое обоснование этому может быть найдено в [41]. Примеры подобных полиномов третьей степени, позволяющие избавиться от чередования знаков в (6.33), приведены в таблице 6.5 для $p \leq 31$. Последние две колонки таблицы содержат значения не максимального периода L линейной последовательности, генерируемой с помощью регистра сдвига, и периода N результирующей троичной последовательности. Еще одним достоинством этих полиномов является то, что противоположный, по меньшей мере, одну из коэффициентов полинома элемент равен 1, а, значит, умножение на него сводится к простому соединению с сумматором.

Таблица 6.5. Непримитивных полиномов над простым полем.

p	$f(x)$	L	N
3	$x^3 + 2x + 2$	13	13
5	$x^3 + 4x^2 + 4$	31	31
7	$x^3 + 6x + 5$	171	57
11	$x^3 + 10x + 7$	665	133
13	$x^3 + 12x + 9$	1098	183
17	$x^3 + 16x + 15$	2456	307
19	$x^3 + 18x + 15$	3429	381
23	$x^3 + 22x + 19$	6083	553
29	$x^3 + 28x + 28$	1742	871
31	$x^3 + 30x + 22$	14895	993

Пример 6.11.6. Сформируем троичную последовательность, отвечающую $p=3$, $n=3$, используя полином $x^3 + 2x + 2$ из табл. 6.5. Тогда рекуррентное соотношение (6.13) принимает вид $d_i = d_{i-2} + d_{i-3}$, при начальном состоянии $d_0 = 1, d_1 = d_2 = 0$ генерируя линейную последовательность над $GF(3)$ вида 1,0,0,1,0,1,1,1,2,2,0,1,2 периода $L=13$. После отображения ее ненулевых элементов в их характеры, а нулевых в вещественный нуль будет сформирована троичная последовательность периода $N=13$, идентичная полученной в предшествующем примере. \square

Распространение вышеприведенной конструкции на случай $q = p^w$, $p > 2, w > 1$ следует непосредственно, и правило (6.32) сохраняет свою значимость. Единственное отличие заключается в том, что m – последовательность $\{d_i\}$ является теперь q -ичной, т.е. с элементами, принадлежащими расширенному (в отличие от простого) конечному полю $GF(q)$. Арифметические операции в расширенных полях есть нечто иное, чем операции по модулю q , и поэтому не считаем рациональным подробно останавливаться здесь на этих деталях. Заинтересованный читатель может более подробно ознакомиться с ними в [40-41].

В отличие от ранее рассмотренного материала, алгоритм конструирования троич-

ных последовательностей для $q = 2^w$, также обеспечивающих идеальность их периодической АКФ, основан на значительно более сложных математических понятиях, таких как квадратики в конечных полях [42].

Если любую из рассмотренных троичных последовательностей посимвольно умножить на единственную бинарную последовательность 1,1,1,-1, имеющую идеальную периодическую АКФ, результирующая троичная последовательность будет характеризоваться учетверенной длиной без изменения значения пик-фактора и идеальности АКФ. Аналогичным образом, посимвольное произведение двух троичных последовательностей с идеальной АКФ и взаимно простыми длинами N_1, N_2 также будет обладать идеальной АКФ, длиной $N = N_1 N_2$ и пик-фактором $v = v_1 v_2$, где v_i определяет значение пик-фактора i -й последовательности ($i = 1, 2$).

Таблица 6.6 содержит значения длин и пик-фактора последовательностей с параметрами q, p, n из диапазона $N \leq 1057$, формируемых согласно описанному алгоритму. Строки, в которых значения длины представлены в виде произведения, соответствуют последовательностям, получаемым в результате посимвольного произведения исходной троичной последовательности с бинарной последовательностью вида 1,1,1,-1. В этом случае параметры q, p, n отвечают исходной троичной последовательности. Как следует из таблицы, для многих приведенных кодов характерно пренебрежимо малое значение пик-фактора, что дает разработчику достаточно привлекательную альтернативу лучшим бинарным кодам, при желательности идеальной периодической АКФ.

Таблица 6.6. Параметры троичных последовательностей с идеальной ПАКФ.

N	p	n	q	v	N	p	n	q	v
13	3	3	3	1.444	292=4×73	2	3	8	1.141
21	2	3	4	1.312	307	17	3	17	1.062
31	5	3	5	1.240	341	2	5	4	1.332
52=4×13	3	3	3	1.444	364=4×91	3	3	9	1.123
57	7	3	7	1.163	381	19	3	19	1.055
73	2	3	8	1.141	532=4×133	11	3	11	1.099
84=4×21	2	3	4	1.312	553	23	3	23	1.045
91	3	3	9	1.123	651	5	3	25	1.042
121	3	5	5	1.494	732=4×183	13	3	13	1.083
124=4×31	5	3	5	1.240	757	3	3	27	1.0384
133	11	3	11	1.099	781	5	5	5	1.250
183	13	3	13	1.083	871	29	3	29	1.036
228=4×57	7	3	7	1.163	993	31	3	31	1.033
273	2	3	16	1.066	1057	2	3	32	1.032

6.12. Подавление боковых лепестков вдоль оси задержек.

Предположим, что проектировщик системы не склонен отвергать бинарные $\{\pm 1\}$ последовательности и, в то же время, не удовлетворен достижимым уровнем боковых лепестков их периодической АКФ ($\rho_{p, \max} \geq 1/N$). В подобных условиях эффективным способом разрешения этих противоречивых устремлений служит «имитация» идеальной периодической АКФ путем отказа от согласованной фильтрации в пользу специальной несогласованной обработки, позволяющей подавить боковые лепестки на всем периоде сигнала. Очень близкие идеи лежат в основе уменьшения или подавления аperiodических боковых лепестков [39,44,45], также как в стремлении побороть межсимвольную интерференцию с помощью нуль-форсирующих эквалайзеров [2,5,7], однако в наиболее прозрачной форме они проявляются в случае применения к периодическим сигналам [39,46,47].

6.12.1. Фильтр подавления боковых лепестков.

Рассмотрим некоторую последовательность $\dots, a_{i-1}, a_i, a_{i+1}, \dots$ периода N , которая манипулирует чипы длительности Δ , и фильтр с конечным импульсным откликом (FIR), осуществляющий суммирование N сигнальных копий, задержанных на Δ и взвешенных

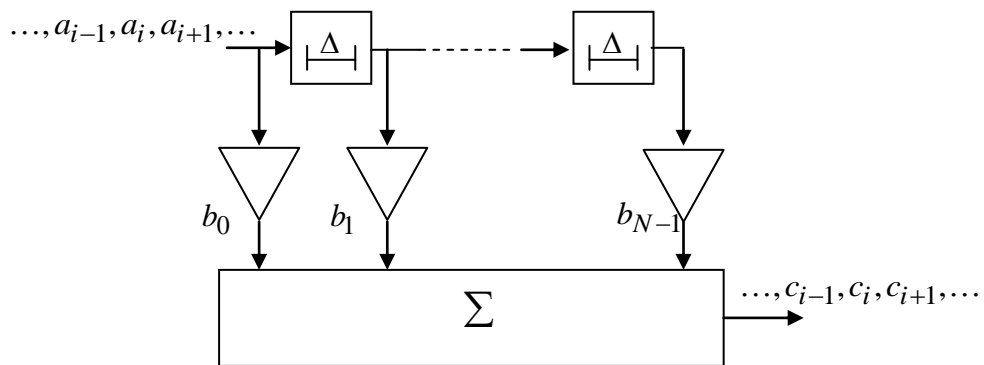


Рис. 6.20. FIR фильтр для последовательности длины N .

коэффициентами $b_i, i = 0, 1, \dots, N-1$, как это показано на рис.6.20. В принципе, то, что представлено ниже, может быть применено к последовательностям произвольного алфавита, однако представляется рациональным ограничиться только бинарным $\{\pm 1\}$ алфавитом, поскольку вне этого ограничения существует множество последовательностей с идеальной периодической АКФ, тем самым лишая задачу подавления боковых лепестков обоснованной мотивации. Соответственно, положим, что коэффициенты фильтра $b_i, i = 0, 1, \dots, N-1$ являются вещественными.

При подаче последовательности $a_i, i = \dots, -1, 0, 1, \dots$ отклик фильтра описывается последовательностью $c_i, i = \dots, -1, 0, 1, \dots$, элементы которой находятся операцией свертки

$$c_i = \sum_{l=0}^{N-1} a_{i-l} b_l, \quad i = \dots, -1, 0, 1, \dots$$

При периодической входной последовательности $a_i = a_{i+N}, i = \dots, -1, 0, 1, \dots$ выходная также будет периодической с тем же периодом N : $c_i = c_{i+N}, i = \dots, -1, 0, 1, \dots$. Тогда N

элементов c_0, c_1, \dots, c_{N-1} полностью определяют выходную последовательность, и c_i становится циклической сверткой

$$c_i = \sum_{l=0}^{N-1} a_{i-l} b_l, \quad i=0,1, \dots, N, \quad (6.36)$$

где вычитание в индексе осуществляется по модулю N .

Предъявим к фильтру следующие требования

$$c_0 \neq 0, \quad c_i = 0, \quad i=1,2, \dots, N-1, \quad (6.37)$$

физический смысл которых состоит в том, что выходной сигнал фильтра имеет ненулевой основной лепесток, повторяющийся с периодом $N\Delta$, тогда как все боковые лепестки между ними равны нулю. Подобный фильтр, называемый в дальнейшем *фильтром подавления боковых лепестков* (ФПБЛ) (*sidelobe suppression filter* (SLSF)), имитирует своим откликом идеальную периодическую АКФ. Поскольку для бинарных кодов идеальная АКФ не достижима (за исключением единственного тривиального сигнала), то ФПБЛ является рассогласованным и, следовательно, уступает согласованному фильтру в отношении сигнал-шум.

Кратчайшим путем получения точного выражения для коэффициентов фильтра служит применение дискретного преобразования Фурье (ДПФ). Последовательность $a_i, i=0,1, \dots, N-1$ и ее компоненты ДПФ спектра $\tilde{a}_k, k=0,1, \dots, N-1$ связаны друг с другом прямым и обратным ДПФ:

$$\tilde{a}_k = \sum_{i=0}^{N-1} a_i \exp\left(-j \frac{2\pi i k}{N}\right), \quad a_i = \frac{1}{N} \sum_{k=0}^{N-1} \tilde{a}_k \exp\left(j \frac{2\pi i k}{N}\right), \quad i, k=0,1, \dots, N-1.$$

Нашей целью является достижение дискретной дельта-функции (6.37) на выходе фильтра, имеющей только один ненулевой элемент на периоде. Спектр подобной функции является равномерным: $\tilde{c}_k = c_0, k=0,1, \dots, N-1$. Тогда вследствие теоремы о свертке [1] спектр последовательности (6.36) на выходе фильтра может быть представлен в виде $\tilde{c}_k = \tilde{a}_k \tilde{b}_k = c_0, k=0,1, \dots, N-1$, где спектр последовательности коэффициентов фильтра \tilde{b}_k есть ничто иное, как передаточная функция ФПБЛ:

$$\tilde{b}_k = \frac{c_0}{\tilde{a}_k}, \quad k=0,1, \dots, N-1. \quad (6.38)$$

Как видно, передаточная функция ФПБЛ обратно пропорциональна спектру сигнала, что послужило причиной присвоения фильтрам этого типа названия инверсных фильтров. Инверсный фильтр как раз и выравнивает входной спектр. Как показывает последнее соотношение, ФПБЛ физически реализуем для любой периодической последовательности, ДПФ спектр которой не имеет нулевых компонент. Применение обратного преобразования Фурье к (6.38) дает точное выражение для коэффициентов ФПБЛ

$$b_i = \frac{c_0}{N} \sum_{k=0}^{N-1} \frac{1}{\tilde{a}_k} \exp\left(j \frac{2\pi i k}{N}\right), \quad i=0,1, \dots, N-1. \quad (6.39)$$

6.12.2. Вычисление потерь в отношении сигнал шум.

В случае согласованного фильтра (см. рис.6.20) его коэффициентами будут служить (игнорируя несущественный общий множитель) значения символов зеркально ото-

браженной входной последовательности¹: $b_i = a_{N-i}$, $i = 0, 1, N-1$ и амплитуда выходной последовательности составит $A_{mf} = \sum_{i=0}^{N-1} a_i^2 = N$, поскольку входная последовательность бинарная. Для входного шума, обладающего временем корреляции в пределах Δ и дисперсией σ^2 , дисперсия на выходе согласованного фильтра составит величину $\sigma_{mf}^2 = \sigma^2 \sum_{i=0}^{N-1} b_i^2 = \sigma^2 \sum_{i=0}^{N-1} a_{N-i}^2 = N\sigma^2$. Таким образом, на выходе согласованного фильтра отношение сигнал-шум по мощности q_{mf}^2 составит

$$q_{mf}^2 = \frac{A_{mf}^2}{\sigma_{mf}^2} = \frac{N}{\sigma^2}. \quad (6.40)$$

Аналогичным образом, амплитуда выходной последовательности ФПБЛ составляет $A_{sl} = c_0$, а дисперсия шума

$$\sigma_{sl}^2 = \sigma^2 \sum_{i=0}^{N-1} b_i^2.$$

Согласно теореме Парсеваля и свойству временного сдвига ДПФ периодическая АКФ произвольной последовательности u_0, u_1, \dots, u_{N-1} периода N связана с энергетическим спектром последовательности $|\tilde{u}_0|^2, |\tilde{u}_1|^2, \dots, |\tilde{u}_{N-1}|^2$ с помощью обратного ДПФ:

$$R_p(m) = \sum_{i=0}^{N-1} u_i u_{i-m}^* = \frac{1}{N} \sum_{k=0}^{N-1} |\tilde{u}_k|^2 \exp\left(j \frac{2\pi mk}{N}\right), \quad m = 0, 1, \dots, N-1. \quad (6.41)$$

В частности

$$R_p(0) = \sum_{i=0}^{N-1} |u_i|^2 = \frac{1}{N} \sum_{k=0}^{N-1} |\tilde{u}_k|^2.$$

Использование последнего соотношения вместе (6.38) в выражении для дисперсии шума на выходе фильтра дает

$$\sigma_{sl}^2 = \frac{\sigma^2}{N} \sum_{k=0}^{N-1} |\tilde{b}_k|^2 = \frac{c_0^2 \sigma^2}{N} \sum_{k=0}^{N-1} \frac{1}{|\tilde{a}_k|^2},$$

и отношение сигнал-шум на выходе ФПБЛ принимает вид

$$q_{sl}^2 = \frac{A_{sl}^2}{\sigma_{sl}^2} = \frac{N}{\sigma^2} \left(\sum_{k=0}^{N-1} \frac{1}{|\tilde{a}_k|^2} \right)^{-1}.$$

Теперь можно оценить энергетические потери в ФПБЛ по сравнению с согласованным фильтром как

$$\gamma = \frac{q_{mf}^2}{q_{sl}^2} = \sum_{k=0}^{N-1} \frac{1}{|\tilde{a}_k|^2}. \quad (6.42)$$

¹ Циклический сдвиг коэффициентов по сравнению со случаем аperiodического сигнала ($b_i = a_{N-1-i}$) служит для более компактного представления (6.38) путем исключения линейной фазовой экспоненты. В случае периодического сигнала данный сдвиг означает циклический сдвиг выходного периодического сигнала и, следовательно, не сказывается на ее форме или отношении сигнал-шум.

Для лучшего понимания последнего результата отметим, что

$$N = \sum_{i=0}^{N-1} a_i^2 = \frac{1}{N} \sum_{k=0}^{N-1} |\tilde{a}_k|^2 = \overline{|\tilde{a}|^2} \quad \text{и} \quad \left(\frac{1}{N} \sum_{k=0}^{N-1} \frac{1}{|\tilde{a}_k|^2} \right)^{-1} = \left(\overline{|\tilde{a}|^{-2}} \right)^{-1}$$

отвечают соответственно среднему арифметическому и гармоническому энергетического спектра последовательности $|\tilde{a}_k|^2$, $k=0,1,\dots,N-1$. Среднее гармоническое любых неотрицательных чисел никогда не превосходит их среднего арифметического, и эти характеристики совпадают, если только все усредняемые числа одинаковы. Следовательно, отношение данных характеристик может служить некоторой мерой степени разброса усредняемых чисел. Однако в рассматриваемом случае это отношение

$$\frac{\overline{|\tilde{a}|^2}}{\left(\overline{|\tilde{a}|^{-2}} \right)^{-1}} = \sum_{k=0}^{N-1} \frac{1}{|\tilde{a}_k|^2} = \gamma$$

в точности соответствует энергетическим потерям в ФПБЛ. Следовательно, потери в отношении сигнал-шум γ определяются неравномерностью энергетического спектра последовательности $|\tilde{a}_k|^2$, $k=0,1,\dots,N-1$, оцениваемой в терминах различия между средним гармоническим и арифметическим.

Возможность подавления всех периодических боковых лепестков выдвигает новый критерий синтеза бинарных последовательностей, который является альтернативным минимизации максимального бокового лепестка $\rho_{p,\max}$. Действительно, какой смысл заботиться об уровне бокового лепестка, когда все боковые лепестки могут быть приведены к нулю? Более естественным представляется минимизация затрат, обуславливающих устранение боковых лепестков, и количественной мерой этих затрат, конечно, выступает потеря в отношении сигнал-шум γ . Рассмотрим вначале бинарную последовательность (гипотетическую, если $N \neq 4$), обладающую равномерным энергетическим спектром: $|\tilde{a}_k|^2 = N$, $k=0,1,\dots,N-1$. В свете соотношения (6.41) это означает, что последовательность обладает идеальной периодической АКФ. Абсолютно предсказуемо, что, поскольку отсутствует объект подавления, то в этом случае ФПБЛ совпадет с согласованным фильтром, не имеющим потерь в отношении сигнал-шум. Соотношение (6.42) подтверждает данный факт, определяя $\gamma=1$, или, выраженное в дБ, $\gamma_{дБ} = 10 \lg \gamma = 0$ дБ. Поскольку не существует бинарных последовательностей с идеальной периодической АКФ, то подавление их боковых лепестков производится за счет потерь в отношении сигнал-шум ($\gamma_{дБ} > 0$), оправдывая введение критерия синтеза $\gamma = \min$.

Как и во многих задачах, касающихся бинарных последовательностей (см. параграф 6.4), глобально оптимальная бинарная последовательность фиксированной длины N с минимальными потерями γ может быть найдена только в результате исчерпывающего поиска. Данная работа была выполнена для длин вплоть до $N=40$ [48]. Конечно, вследствие экспоненциального роста вычислительных затрат указанный поиск не может продолжаться далеко за указанный диапазон. Однако известны многие регулярные правила построения бинарных последовательностей сколь угодно большой длины с очень малыми потерями γ (хотя и без гарантии их глобальной оптимальности).

Рассмотрим специальный класс бинарных последовательностей, обладающих двухуровневой периодической АКФ, т.е. с постоянным уровнем R боковых лепестков

$$R_p(m) = \begin{cases} N, & m = 0 \bmod N, \\ R, & m \neq 0 \bmod N. \end{cases} \quad (6.43)$$

Все минимаксные последовательности, как и ряд других, относятся к данному типу. Энергетический спектр такой последовательности, как показывает (6.41), есть просто ДПФ АКФ:

$$|\tilde{a}_k|^2 = \sum_{m=0}^{N-1} R_p(m) \exp\left(-j \frac{2\pi mk}{N}\right) = N - R + R \sum_{m=0}^{N-1} \exp\left(-j \frac{2\pi mk}{N}\right), \quad k = 0, 1, \dots, N-1.$$

Последняя сумма уже рассматривалась в 6.11.2 и, как было доказано, она равняется N , если $k = 0$, и нулю в противном случае. Таким образом

$$|\tilde{a}_k|^2 = \begin{cases} N + (N-1)R, & k = 0, \\ N - R, & k \neq 0. \end{cases} \quad (6.44)$$

Подстановка последнего соотношения в (6.42) дает

$$\gamma = \frac{1}{N + (N-1)R} + \frac{N-1}{N-R} = \frac{1 + (N-2)\rho}{(1-\rho)[1 + (N-1)\rho]}, \quad (6.45)$$

где $\rho = R/N$ – нормированный уровень бокового лепестка АКФ.

Для определения структуры ФПБЛ, перепишем (6.39) как

$$b_i = \frac{c_0}{N} \sum_{k=0}^{N-1} \frac{\tilde{a}_k^*}{|\tilde{a}_k|^2} \exp\left(j \frac{2\pi ik}{N}\right), \quad i = 0, 1, \dots, N-1$$

и подставим (6.44) в последнее соотношение, тогда

$$b_i = \frac{c_0}{N} \left[\frac{\tilde{a}_0}{N + (N-1)R} - \frac{\tilde{a}_0}{N-R} + \frac{1}{N-R} \sum_{k=0}^{N-1} \tilde{a}_k^* \exp\left(j \frac{2\pi ik}{N}\right) \right], \quad i = 0, 1, \dots, N-1.$$

Сумма по k в данном соотношении определяет коэффициенты согласованного фильтра, поскольку она может быть записана как

$$\frac{1}{N} \sum_{k=0}^{N-1} \tilde{a}_k^* \exp\left(-j \frac{2\pi(N-i)k}{N}\right) = \left[\frac{1}{N} \sum_{k=0}^{N-1} \tilde{a}_k \exp\left(j \frac{2\pi(N-i)k}{N}\right) \right]^* = a_{N-i}, \quad i = 0, 1, \dots, N-1,$$

где использован тот факт, что элементы бинарной последовательности вещественны. Поскольку c_0 произвольный множитель, положим его равным $N - R$. Тогда

$$b_i = a_{N-i} - \frac{\rho \tilde{a}_0}{1 + (N-1)\rho}, \quad i = 0, 1, \dots, N-1. \quad (6.46)$$

Первое слагаемое в данном выражении соответствует последовательности $\{a_i\}$, считываемой справа налево, т.е. коэффициентам согласованного фильтра. Таким образом, для последовательности, обладающей двухуровневой АКФ (6.43), ФПБЛ получается путем незначительного изменения согласованного фильтра: вычитанием некоторой постоянной величины из всех коэффициентов. Более того, для бинарных последовательностей данного

типа коэффициенты ФПБЛ принимают только два возможных значения $\pm 1 - \frac{\rho \tilde{a}_0}{1 + (N-1)\rho}$,

где \tilde{a}_0 – постоянная составляющая последовательности, т.е. разность между числом плюс и минус единиц на периоде: $\tilde{a}_0 = N_+ - N_-$.

Как известно из предыдущего материала, существует множество бинарных последовательностей с АКФ вида (6.43), для которых $R = -1$ (m – последовательности, последовательности Лежандра и другие минимаксные последовательности с АКФ вида (6.12)). Оценка их потерь в ФПБЛ согласно (6.45) дает $\gamma = 2N/(N+1)$, т.е. $\gamma \approx 2$ (3 дБ) для практически интересных длин. Отсюда видно, что наиболее популярные минимаксные бинарные

ные последовательности не играют значительной роли в свете критерия потерь в ФПБЛ: половина их энергии теряется при обработке с помощью ФПБЛ.

С другой стороны, при положительном R , достаточно малым в сравнении с N , $\gamma < 1/(1-\rho)$, т.е. потери оказываются достаточно малыми. Так называемые коды Зингера служат хорошим примером подобных бинарных последовательностей. Код Зингера существует при любой длине вида $N = \frac{q^n - 1}{q - 1}$ и обладает двухуровневой АКФ типа (6.43) со

значением $R = N - 4q^{n-2}$, где $q = p^w$ – натуральная степень простого p , n – натуральное число. Наиболее интересная с рассматриваемых позиций модификация кодов Зингера от-

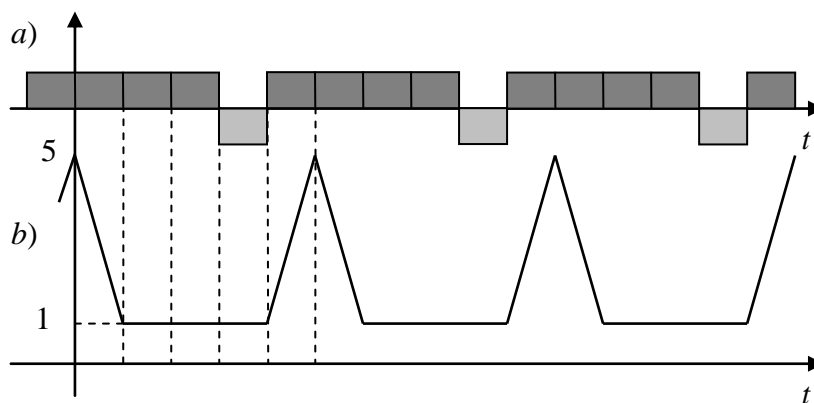


Рис. 6.21. Бинарный сигнал длины $N = 5$ и его периодическая АКФ.

вечает $q = 3$, для которой $\rho = \frac{3^{n-2} - 1}{3^n - 1}$ и $\gamma < \frac{3^n - 1}{8 \cdot 3^{n-2}} < \frac{9}{8} = 1.125$, т.е. $\gamma_{дБ} < 0.51$ дБ. Как видно, данные коды достаточно привлекательны при обработке их ФПБЛ, поскольку потери в отношении сигнал-шум, сопровождающие полное подавление боковых лепестков, малы.

Пример 6.12.1. Рассмотрим периодическую версию бинарного кода Баркера длины $N = 5$ из таблицы 6.1: $+1, +1, +1, -1, +1$, для которого $N_+ = 4$, $N_- = 1$, а постоянная состав-

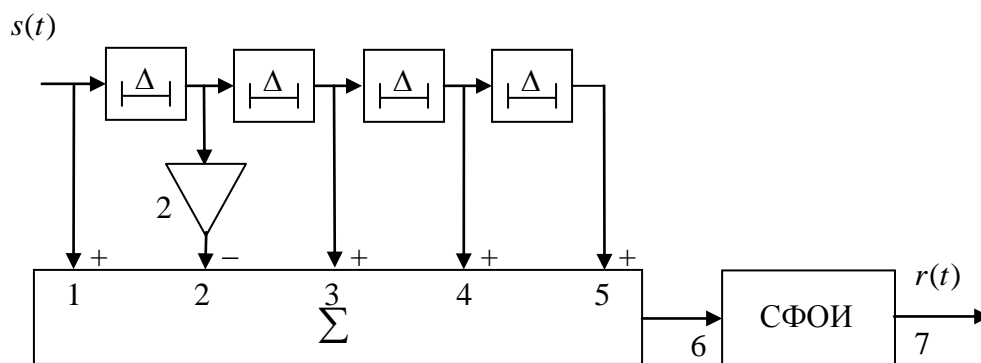


Рис. 6.22 ФПБЛ для последовательности длины $N=5$.

ляющая $\tilde{a}_0 = 3$. Его периодическая АКФ, что может быть проверено непосредственной проверкой, удовлетворяет соотношению (6.43) со значением $R = 1$ ($\rho = 1/5$). В действи-

тельности данная последовательность представляет собой код Зингера с параметрами $q = 4, n = 2$. Рис. 6.21, *b* демонстрирует вид АКФ (т.е. отклик согласованного фильтра) периодического сигнала, модулированного данной последовательностью (рис. 6.21, *a*). Как следует из (6.46), согласованный фильтр для данной последовательности легко трансформируется в ФПБЛ изменением всех коэффициентов $+1$ на $+2/3$ и -1 на $-4/3$, которое после соответствующего масштабирования означает замену -1 на -2 при оставлении без

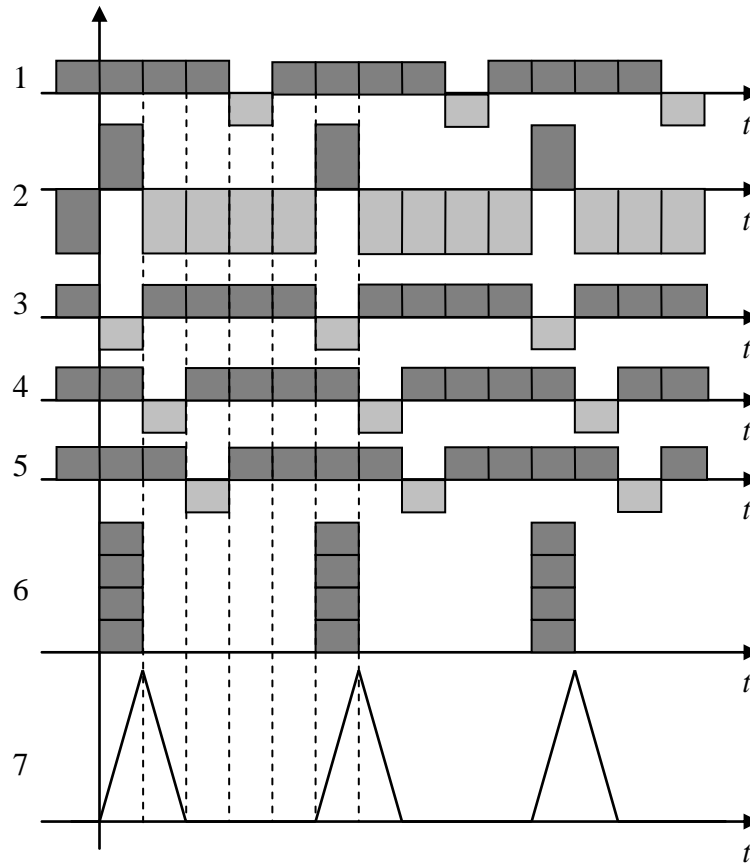


Рис. 6.23 Отклик ФПБЛ для бинарной последовательности $N=5$.

изменения $+1$ (см. рис. 6.22). Отклик ФПБЛ на тот же сигнал построен на рис. 6.23, на котором оцифровка диаграмм соответствует точкам на рис. 6.22. Выходной сигнал фильтра имеет желаемую форму, т.е. нулевой уровень боковых лепестков. Используя соотношение (6.45), легко определить энергетические потери в ФПБЛ как $\gamma = 10/9 = 1.1111\dots$ (0.46 дБ).□

Основываясь, как и для кодов Зингера, на отображении линейных рекуррентных последовательностей над конечными полями на бинарный $\{\pm 1\}$ алфавит, было найдено множество даже более эффективных бинарных кодов. Все они характеризуются одинаково привлекательными чертами: простотой структуры ФПБЛ, коэффициенты которого принимают не более трех различных значений. Не углубляясь далее в детали, которые достаточно сложны, и отсылая заинтересованного читателя к [49-50], отметим только, что среди упомянутых последовательностей имеются семейства, обладающие асимптотически малыми потерями в ФПБЛ: $\gamma_{\text{дБ}} \rightarrow 0$ дБ при $N \rightarrow \infty$.

6.13. ЧМ сигналы с оптимальной аperiodической АКФ.

Краткое обсуждение сути построения ЧМ сигналов с хорошими корреляционными свойствами начнем с соотношения (5.20), напоминающего о том, что требование малого

уровня $R_p(m)$ эквивалентно минимизации числа совпадающих частот в кодовой последовательности частот F_0, F_1, \dots, F_{N-1} и ее копии, сдвинутой на m позиций. Очевидно, что если число чипов (элементарных импульсов) N (т.е. длина) не превосходит числа различных частот M , получение нулевого уровня боковых лепестков $\rho_a(m)$, $m \neq 0$ может быть достигнуто тривиальным путем: использованием частотного кода, все элементы которого F_i различны. На практике, однако, типичной является ситуация, когда $N > M$, что влечет за собой повторения элементов F_i , $i = 0, 1, \dots, N-1$ и, следовательно, по меньшей мере, одно совпадение в сдвинутых копиях частотного кода, т.е. $\rho_{a, \max} \geq 1/N$.

Поскольку частотная последовательность может быть описана с помощью массива (решетки) размера $M \times N$ (см. параграф 5.5), то минимизация $\rho_{a, \max}$ означает построение решетки с минимально возможным числом совпадений помеченных элементов в исходной решетке и ее копии, сдвинутой по горизонтали на m позиций. Одной из ключевых проблем является конструирование т.н. *радарных решеток* (*radar arrays*), определяемых как $M \times N$ массивы, имеющие только один помеченный элемент в каждом столбце и $\rho_{a, \max} = 1/N$, т.е. с числом упомянутых совпадений не более одного. Данное стремление понимается как отыскание максимально длинной радарной решетки с фиксированным значением M , поскольку это означает минимизацию $\rho_{a, \max}$ при ограничениях на частотный ресурс. Следуя [51], попытаемся доказать простейшую верхнюю границу длины радарной решетки.

Рассмотрим последовательность F_0, F_1, \dots, F_{N-1} и отметим, что для получения не более одного совпадения все разности между номерами позиций, отвечающих одинаковым частотам, должны быть различными. Действительно, пусть $F_i = F_k$, $F_s = F_t$ и $i - k = s - t > 0$. Тогда в исходной последовательности и ее копии, сдвинутой на $m = i - k = s - t$ позиций, произойдет не менее двух совпадений. Обозначим через n_i число символов (частот) среди F_0, F_1, \dots, F_{N-1} , повторяющихся i раз. Тогда

$$\sum_i i n_i = N \text{ и } \sum_i n_i = M. \quad (6.47)$$

Подсчитаем теперь число возможных разностей между номерами позиций, содержащих идентичные частоты. Если некоторая частота повторяется i раз, то число подобных разностей для этой частоты будет $i(i-1)$. Поскольку в последовательности содержится n_i частот, повторяющихся i раз, то общее число указанных разностей составит величину $\sum_i i(i-1)n_i$, и так как ни одна из разностей не должна повторяться, то

$$\sum_i i(i-1)n_i \leq N-1, \quad (6.48)$$

где правая часть неравенства дает максимальное число не равных положительных разностей среди номеров $\{0, 1, \dots, N-1\}$. Соотношение $i(i-1) + 3 - 2i = i^2 - 3i + 3$ не имеет вещественных корней и, следовательно, положительно при любом i . Поэтому сумма

$$\sum_i [i(i-1) + 3 - 2i]n_i = \sum_i i(i-1)n_i + 3\sum_i n_i - 2\sum_i i n_i \geq 0,$$

что, с учетом (6.47), (6.48), может быть преобразовано к виду $N - 1 + 3M - 2N \geq 0$, или

$$N \leq 3M - 1. \quad (6.49)$$

В действительности эта граница не является строгой. Известны более точные границы, например в [52], асимптотический вариант которых приводит к результату

$$N \leq \frac{20 + \sqrt{6}}{8} M, \quad M \gg 1, \quad (6.50)$$

что снижает правую часть (6.49) приблизительно на $0.194M$.

Для значений $M \leq 16$ известны абсолютно точные, т.е. действительно достижимые, верхние границы длины N . Для указанного диапазона M максимальная длина N_{\max} радарных решеток ограничена сверху соотношениями [52]

$$N_{\max} = \begin{cases} 3M - 2, & 2 \leq M \leq 4, \\ 3M - 3, & 5 \leq M \leq 9, \\ 3M - 4, & 10 \leq M \leq 13, \\ 3M - 5, & 14 \leq M \leq 16. \end{cases} \quad (6.51)$$

Пример 6.13.1. Частотный код $\{1, 2, 3, 4, 5, 6, 7, 8, 7, 4, 3, 9, 9, 5, 8, 2, 6, 5, 1, 4, 2, 1, 3, 7\}$, который определяет номера частот из некоторого алфавита с $M = 9$ или, что эквивалентно, номера строк в каждой колонке решетки, имеет максимально возможную длину $N = 24$. Свойства радарной решетки, т.е. наличие единственного совпадения частот при всех ненулевых сдвигах, может быть проверено путем непосредственной проверки (задача 6.53). \square

Кроме того, известны регулярные правила построения радарных решеток длины $N = 2.5M$ для любого четного M (см. [51]) и для $M/2$, равного произведению простых чисел, дающего остаток один от деления на 4, т.е. $M = 10, 26, 34, 58, \dots$

Сонарные (sonar) решетки представляют собой обобщение радарных, сохраняющие свойство «не более одного совпадения» для произвольной комбинации горизонтального и вертикального сдвига [53]. Практически это требование отражает стремление иметь слабую корреляцию сигнальных копий, сдвинутых как во времени, так и по частоте. Рассматривая подход к выбору частотного пространства для ЧМ сигналов (параграф 5.5), отмечалось, что частотные сдвиги, вызывающие переход из текущего значения частоты к соседнему, более типичны для сонарных, чем локационных систем. Известно, например, множество регулярных алгоритмов построения массивов Костаса [54], т.е. квадратных ($M = N$) сонарных массивов или ЧМ последовательностей, имеющих одинаковую длину и число различных частот.

Задачи.

6.1. Частота заполнения прямоугольного импульса длительности $T = 10$ мксек линейно спадает со 110 до 90 МГц. Вычислить выигрыш от обработки сигнала. Какова приблизительно длительность сигнала на выходе согласованного фильтра? Изобразить функцию и диаграмму неопределенности.

6.2. Частота заполнения прямоугольного импульса длительности $T = 10$ мксек линейно спадает со 110 до 90 МГц на первой его половине, а затем линейно возрастает от 90 до 110 МГц на второй. Вычислить выигрыш от обработки сигнала. Изобразить функцию неопределенности и ее горизонтальные сечения в верхнем и нижнем уровнях.

6.3. Вычислить аperiodическую и периодическую АКФ для бинарного кода Баркера длины $N = 11$ наиболее экономным образом.

6.4. Возьмем периодическую последовательность $\{a_i\}$ периода N и сформируем новую последовательность $\{b_i\}$, выбирая каждый d -й элемент $\{a_i\}$: $b_i = a_{di}$, где произведение в индексе берется по модулю N . Данная трансформация называется децимированием $\{a_i\}$ с индексом d . Доказать, что если $\{a_i\}$ обладает идеальной периодической АКФ, а d взаимно просто с N , то $\{b_i\}$ также будет обладать идеальной периодической АКФ.

6.5. Бинарный $\{\pm 1\}$ код длины $N = 5$ обладает периодической АКФ вида $R_p(m) = 1, m \neq 0 \pmod{5}$. Его аperiodическая АКФ принимает значения $R_a(1) = 0$, $R_a(2) = -1$. Определить $R_a(3)$ и $R_a(4)$.

6.6. Бинарный $\{\pm 1\}$ код длины $N = 5$ обладает постоянной составляющей $\tilde{a}_0 = +3$ и $R_a(4) = -1$. Определить значения $R_p(m)$ и остальные значения $R_a(m)$.

6.7. Может ли бинарный $\{\pm 1\}$ код нечетной длины $N > 5$ иметь $R_a(5) = -1$? Может ли бинарный $\{\pm 1\}$ код четной длины $N > 6$ иметь $R_a(6) = -1$? Сформулировать и доказать для произвольного бинарного кода связь между соотношением трех величин: длины N , сдвига m и уровня $R_a(m)$?

6.8. Возможно ли, чтобы бинарная $\{\pm 1\}$ последовательность имела $R_a(m) = 1$, $R_a(m+1) = 3$ для некоторого сдвига m ? Что можно сказать о соотношении между $R_a(m)$ и $R_a(m+1)$?

6.9. Возможно ли, чтобы бинарная $\{\pm 1\}$ последовательность имела $R_a(2) = 2$, $R_p(2) = -1$?

6.10. Предположим, кто-то определил, что каждая из ФМ последовательностей длины $N = 100$ в его распоряжении имеет не нормированную периодическую АКФ, принимающую значения ± 12 при некоторых сдвигах $m \in \{1, 2, \dots, N-1\}$. Может ли среди них существовать код с $\rho_{a, \max} < 0.05$.

6.11. Построить согласованный фильтр для бинарного ФМ сигнала, манипулированного последовательностью вида $\{+-++\}$, и привести диаграммы в характерных точках для апериодической и периодической версии сигнала на входе фильтра.

6.12. Построить согласованный фильтр для бинарного ФМ сигнала, манипулированного последовательностью вида $\{+-+--\}$, и привести диаграммы в характерных точках для апериодической и периодической версии сигнала на входе фильтра.

6.13. Студент вычислил периодическую АКФ бинарной $\{\pm 1\}$ последовательности длины $N=21$ и получил следующие значения $R_p(m_1)=9$, $R_p(m_2)=-3$, $R_p(m_4)=-5$, $R_p(m_5)=-7$, $R_p(m_6)=7$. Могут ли все эти результаты оказаться правильными? Какие из них несомненно не правильные?

6.14. Доказать не существование минимаксных бинарных последовательностей $R_p(m)=\pm 1$, $m=1,2,\dots,N-1$ длин $N=17,21,29,33,37,45$. (При доказательстве использовать тот же метод, что и при определении необходимых условий идеальности АКФ бинарных последовательностей).

6.15. Доказать, что децимирование минимаксной бинарной последовательности вновь дает минимаксную бинарную последовательность, если индекс децимации взаимно прост с периодом N .

6.16. Доказать, что операция децимации любой периодической последовательности не изменяет величины максимального бокового лепестка, если индекс децимации взаимно прост с периодом N .

6.17. Вычислить $(3+7)-5(6\cdot 7+4)+1$ в поле $GF(11)$.

6.18. Решить уравнение $6x+7(5+4\cdot 2)^{-1}=1$ в поле $GF(11)$.

6.19. Доказать, что поле $GF(4)$ не может быть построено на основании операций по модулю 4.

6.20. Является ли последовательность длины $L=7$ вида $\{0100110\}$ двоичной m -последовательностью? Каков будет ответ в случае замены всех нулевых элементов на единичные и наоборот?

6.21. Построить двоичную m -последовательность длины $L=15$ с начальным состоянием $d_0=1, d_1=d_2=d_3=0$, нарисовать структуру ее генератора и составить таблицу, описывающую изменение состояний генератора.

6.22. Доказать, что на одном периоде бинарной m -последовательности памяти n число серий последовательных символов (01) , (10) , (11) встречается 2^{n-2} раза, а число серий (00) – на единицу меньше.

6.23. Наблюдается m -последовательность, о которой известны алфавит и значение памяти, но неизвестны коэффициенты рекурсии (6.13). Каково минимально необходимое и достаточное число наблюдаемых символов для определения значений этих коэффициентов?

6.24. Доказать независимость двузначного характера от конкретного выбора примитивного элемента.

6.25. Определить примитивный элемент поля $GF(13)$. Построить таблицу логарифмов и двузначных характеров всех ненулевых элементов $GF(13)$.

6.26. Построить последовательность Лежандра длины $N = 11$, вычислить ее периодическую АКФ и сравнить ее с теоретически предсказываемой ПАКФ.

6.27. Построить последовательность Лежандра длины $N = 13$, вычислить ее периодическую АКФ и сравнить ее с теоретически предсказываемой ПАКФ.

6.28. Определить циклический сдвиг последовательности из задачи 6.26, обеспечивающий минимальный боковой лепесток аperiodической АКФ.

6.29. Доказать идеальность периодической АКФ кодов Чу нечетной длины.

6.30. Доказать идеальность периодической АКФ кодов Франка. (При доказательстве использовать представление вида $i = i_1h + i_2$, $m = m_1h + m_2$; $0 \leq i_1, i_2, m_1, m_2 \leq h - 1$ и суммирование по i_1, i_2 в (5.9)).

6.31. Существует ли последовательность с ФМ-8 длины $N = 64$ и идеальной периодической АКФ? Если существует, то построить ее.

6.32. Доказать не существование последовательностей нечетной длины с квадратурной ФМ и идеальной периодической АКФ.

6.33. Доказать не существование последовательности с квадратурной ФМ длины $N = 30$ и идеальной периодической АКФ.

6.34. Доказать, что для троичных $\{0, \pm 1\}$ последовательностей с идеальной ПАКФ число ненулевых элементов на периоде всегда является квадратом целого числа.

6.35. Доказать не существование троичных $\{0, \pm 1\}$ последовательностей, обладающих идеальной ПАКФ и единственным нулевым элементом на периоде, для любой нечетной длины.

6.36. Построить ФПБЛ, продемонстрировать эффект подавления боковых лепестков и определить потери в отношении сигнал-шум для бинарной последовательности вида $\{+++++-\}$.

6.37. Определить энергетические потери в ФПБЛ для последовательности вида $\{++----+-\}$.

6.38. Определить максимальную длину ЧМ последовательности с объемом частотного алфавита $M = 4$ и не более одним совпадением частот.

7. Ансамбли широкополосных сигнатур в CDMA приложениях.

7.1. Широкополосная передача данных.

Из рассмотренного в 4.3–4.6 материала очевидным образом следует, что в CDMA сети каждый из K абонентов передает или получает свои индивидуальные данные посредством использования некоторой специфической для пользователя сигнатуры, причем тщательно выбранный ансамбль из K сигнатур обеспечивает максимально возможную совместимость абонентов. С целью обеспечения транспортировки k -й сигнатурой потока данных необходимо применение некоторого вида модуляции, которая, учитывая широкополосную природу CDMA сигнатур, часто называется широкополосной модуляцией. Различают две классические версии широкополосной модуляции: *прямой последовательностью* (*direct sequence* (DS)) и *прыгающей частотой* (*frequency hopping* (FH)). Первая из них наиболее типична для современных коммерческих многопользовательских беспроводных приложений, тогда как вторая будет лишь кратко рассмотрена.

7.1.1. Прямое расширение спектра: бинарная манипуляция данных и бинарные сигнатуры.

Общая идея прямого расширения спектра состоит в амплитудно-фазовой модуляции APSK сигнатур потоком данных. Для облегчения восприятия данной концепции начнем рассмотрение с простейшего случая передачи данных с помощью бинарной модуляции без расширения спектра. Пусть $B_k(t)$ информационный сигнал k -го пользователя (рис. 7.1), в котором импульсы положительной и отрицательной полярности длительности T_b отвечают передаче бита информации, равного 0 или 1 соответственно. Если, как и в

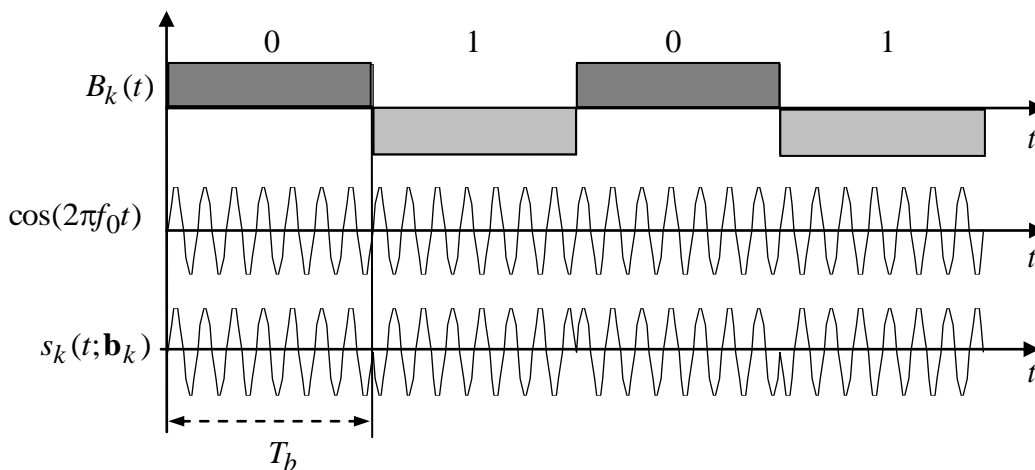


Рис.7.1. Иллюстрация передачи данных с помощью бинарной ФМ

главе 4, $\mathbf{b}_k = (\dots, b_{k,-1}, b_{k,0}, b_{k,1}, \dots)$ описывает битовый поток k -го пользователя, то $B_k(t) = b_{k,i} = \pm 1$, $(i-1)T_b < t \leq iT_b$. Передача $B_k(t)$ с использованием бинарной ФМ просто означает умножение ее на непрерывную несущую частоту f_0 , позволяя прийти к отсылаемому модулированному сигналу (см. рис. 7.1):

$$s_k(t; \mathbf{b}_k) = B_k(t) \cos(2\pi f_0 t). \quad (7.1)$$

Таким образом, практическая реализация бинарного ФМ модулятора представляет собой

умножитель, представленный на рис. 7.2, а. После распространения по каналу сигнал приобретает временную задержку τ_k и начальную фазу φ_k , а также испытывает ослабление, которое не учитывается в дальнейшем рассмотрении. Тогда принятый полезный сигнал имеет вид

$$s_{kr}(t; \mathbf{b}_k) = B_k(t - \tau_k) \cos(2\pi f_0 t + \varphi_k). \quad (7.2)$$

Типичный приемник бинарной ФМ информации содержит схемы слежения за задержкой и фазой несущей частоты, которые оценивают текущие значения задержки τ_k и начальной фазы φ_k . В настоящий момент вопрос о точности оценивания может быть оставлен в стороне, допуская, что приемнику известны «истинные» значения τ_k и φ_k . Если рассматриваемый сигнал искажен АБГШ, то оптимальная (максимально правдоподобная) процедура (см. 2.2) восстановления i -го переданного бита состоит в вычислении корреляции наблюдения $y(t) = s_{kr}(t; \mathbf{b}_k) + n(t)$ с разностью сигналов, отвечающих значениям информационного бита 0 и 1 соответственно, которая в рассматриваемом случае есть просто $2\cos(2\pi f_0 t + \varphi_k)$. Поскольку для решения о принятом бите используется только полярность корреляции и поскольку i -й бит на выходе канала занимает временной интервал $[(i-1)T_b + \tau_k, iT_b + \tau_k]$, то обсуждаемая корреляция представляет собой выражение

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

а решение $\hat{b}_{k,i} = 0$ или $\hat{b}_{k,i} = 1$ принимается в зависимости от положительного или отрицательного знака z_k . Возможная и широко распространенная структура демодулятора, реализующего данное правило, представлена на рис. 7.2, б. Она содержит коррелятор, выполненный в виде перемножителя наблюдения и генерируемого местного образца $\cos(2\pi f_0 t + \varphi_k)$ 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,

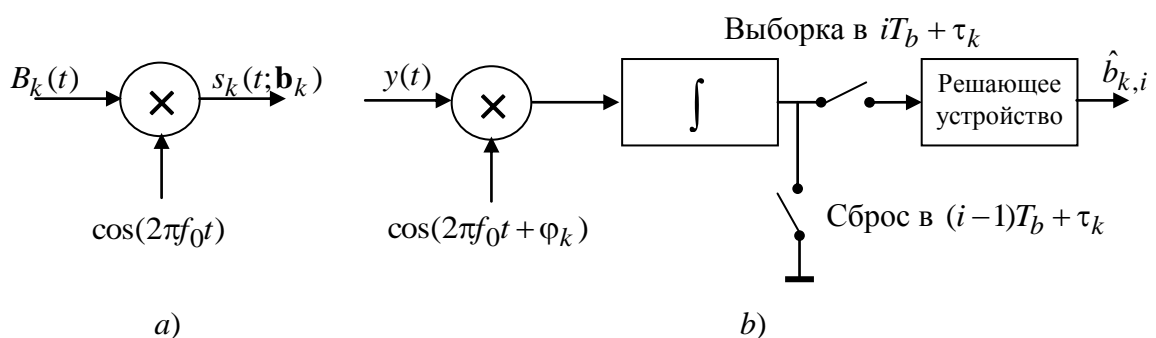


Рис.7.2. Бинарный ФМ модулятор (а) и демодулятор (б).

подготавливая его к работе на следующем битовом интервале.

Рассмотрим теперь изменения, которые необходимо осуществить в случае передачи бинарных ФМ данных с расширением спектра бинарной ФМ прямой последовательностью. Пусть $s_k(t)$ – k -я пользовательская сигнатура, т.е. дискретный сигнал, состоящий из чипов длительности Δ , манипулированная некоторой специфической для каждого пользователя бинарной последовательностью. Пусть на интервале одного бита данных содержится N чипов сигнатуры. Тогда расширение прямой последовательностью бинарно-

го ФМ сигнала заключается только во введении еще одного умножения в (7.1) – на сигнатуру $s_k(t)$:

$$s_k(t; \mathbf{b}_k) = s_k(t)B_k(t)\cos(2\pi f_0 t). \quad (7.3)$$

Поскольку ширина полосы сигналов (7.1) и (7.3) есть величина обратная длительности бита $T_b = 1/R$ и длительности чипа $\Delta = T_b/N = 1/RN$ соответственно, то распределение прямой последовательностью расширяет спектр в N раз. Последнее объясняет еще одно наименование, используемое для обозначения частотно-временного произведения и выигрыша от обработки $WT = N$, *показателя расширения (spreading factor)*. На практике операции умножения в (7.3) могут быть выполнены в произвольном порядке, например, как это показано на рис. 7.3 (расширение бинарной m -последовательностью длины $N = 7$, $T_b = N\Delta$) и 7.5, а, где битовый поток $B_k(t)$ первоначально умножается на сигнатуру $s_k(t)$, а результат произведения $s_k(t)B_k(t)$ модулирует непрерывную несущую. Можно сказать, что в данном случае битовый поток первоначально модулирует бинарную сигнатуру, а результат используется для бинарной фазовой манипуляции несущей.

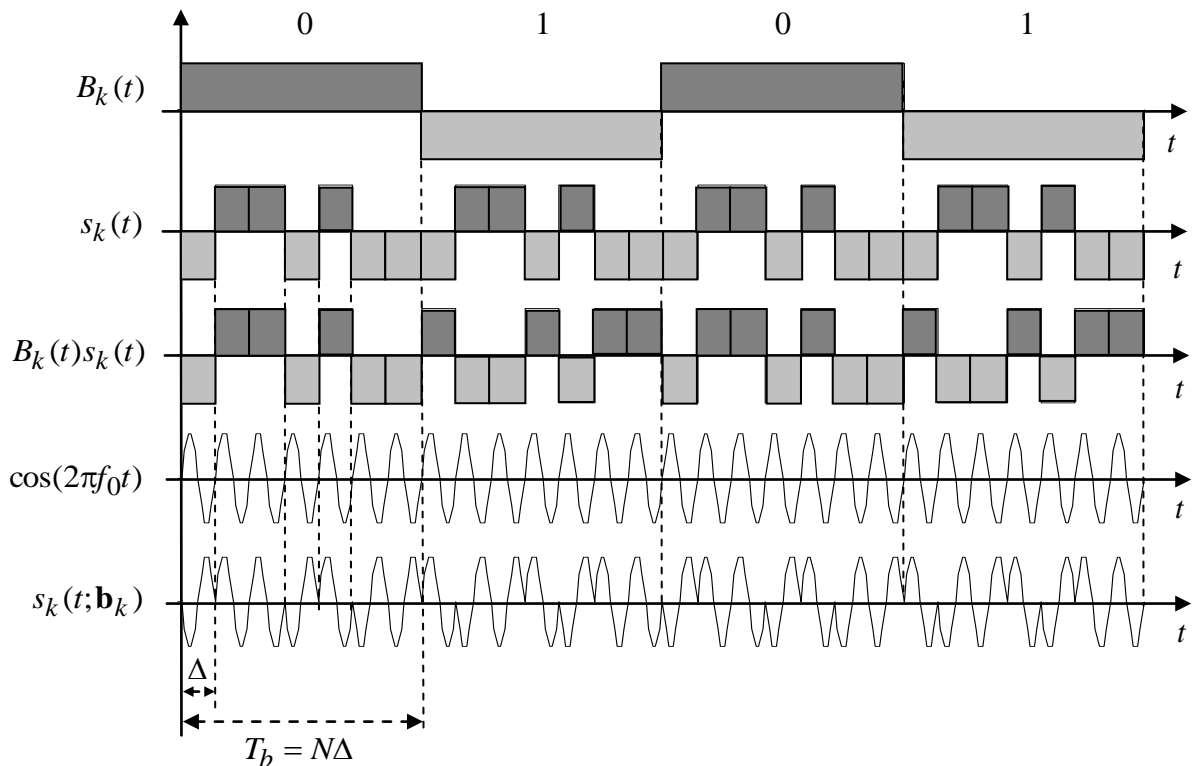


Рис.7.3. Прямое расширение бинарных ФМ данных бинарной сигнатурой.

После распространения по каналу, в ходе которого приобретает задержка τ_k и начальная фаза φ_k , сигнал принимает вид

$$s_{kr}(t; \mathbf{b}_k) = s_k(t - \tau_k)B_k(t - \tau_k)\cos(2\pi f_0 t + \varphi_k). \quad (7.4)$$

Предполагая, как и ранее, полное знание о параметрах τ_k, φ_k , приемник для восстановления текущего (i -го) бита должен лишь сделать выбор между сигналом $s_k(t - \tau_k)\cos(2\pi f_0 t + \varphi_k)$ и ее противоположной копией. Для выполнения этой операции оптимальным образом вначале может быть вычислена корреляция

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

между наблюдаемым колебанием $y(t)$ и формируемой на месте полосной копией сигнатуры $s_k(t - \tau_k) \cos(2\pi f_0 t + \varphi_k)$, а ее полярность использована при вынесении решения. Интересно отметить, что та же самая оптимальная процедура может быть реализована в два этапа: на первом снимается расширение, 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 $s_k(t - \tau_k)$ сигнатуры, в точности синхронизированную с принимаемым сигналом. Полезная составляющая (7.4) наблюдения после этой операции примет вид

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

где учтена бинарная природа сигнатур ($s_k(t) = \pm 1$), вследствие чего $s_k^2(t) = 1$. Как видно, после данного шага принятый сигнал не обладает никакими чертами широкополосного, полностью совпадая с простым сигналом (7.2) с бинарной манипуляцией, определяемой потоком данных. Вследствие этого процедура умножения наблюдения на копию сигнатуры называется *снятием расширения (despreading)*. Рис. 7.4 иллюстрирует процедуру трансформации широкополосного сигнала с прямым расширением в обычный бинарный ФМ сигнал, модулированный потоком данных.

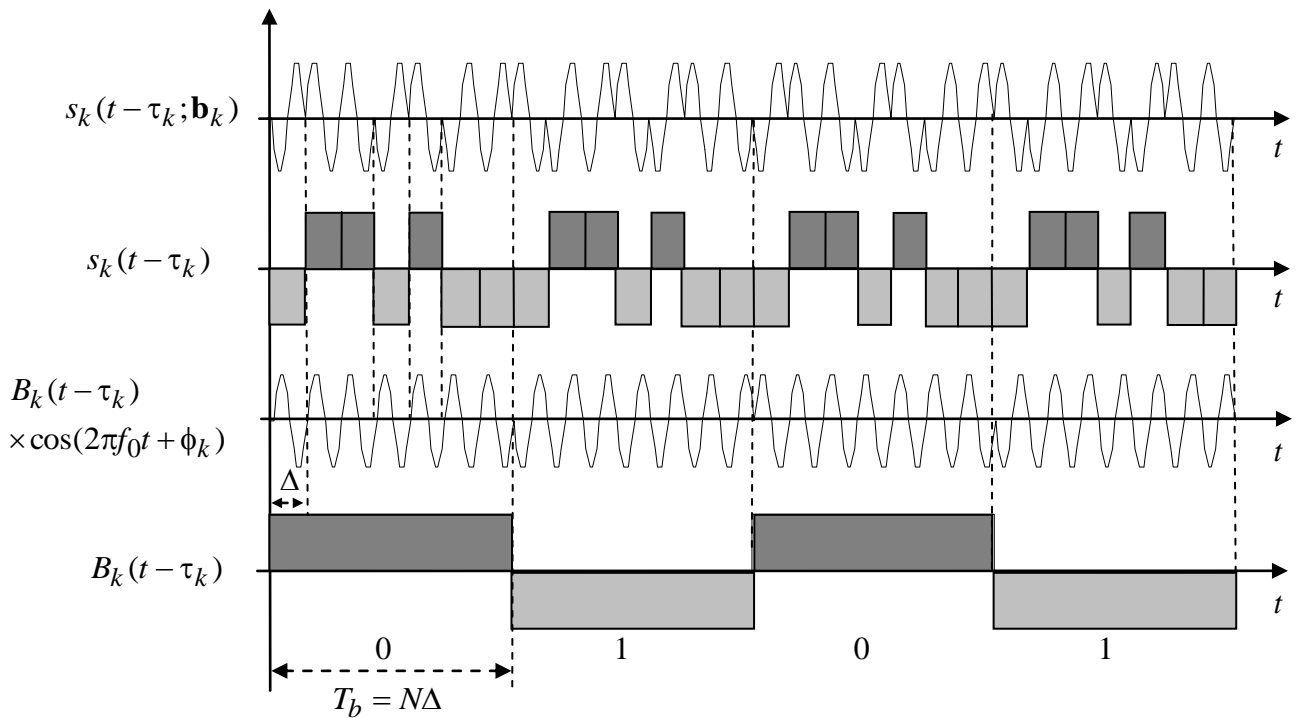


Рис.7.4. Снятие расширения бинарного ФМ сигнала.

Поскольку сигнал со снятым расширением представляет собой непрерывную несущую с бинарной фазовой манипуляцией данными, то дальнейшее извлечение информации осуществляется привычным демодулятором бинарного ФМ сигнала, например таким, который представлен на рис. 7.2, *b*. Полный цикл расширения и сжатия демонстрирует рис. 7.5.

Для проведения обсуждения в терминах частотной области обратимся к рис. 7.6, на

котором приведены спектральные плотности мощности $\tilde{S}_b(f)$, $\tilde{S}_{bs}(f)$ исходного потока данных $B_k(t)$ и его широкополосной версии $s_k(t)B_k(t)$ соответственно. Для последовательности $B_k(t)$ битовых импульсов длительности T_b , полярности которых случайны и

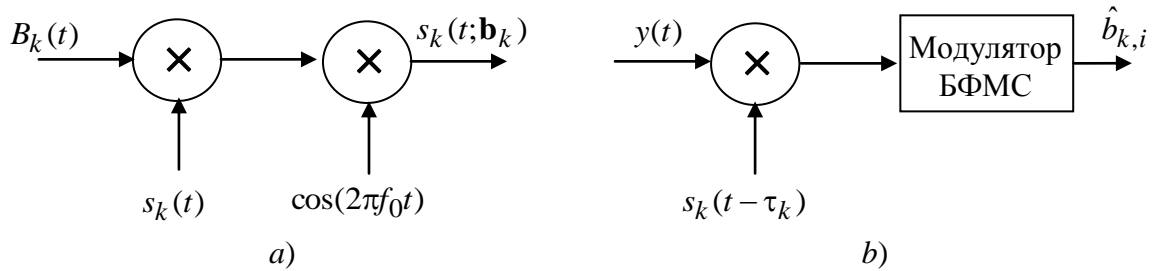


Рис.7.5. Расширение (a) и сужение (b) сигнала с бинарной ФМ.

независимы, спектр мощности определяется как $\tilde{S}_b(f) = T_b \text{sinc}^2(\pi f T_b)$. 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 раз большую полосу: $\tilde{S}_{bs}(f) = \Delta \text{sinc}^2(\pi f \Delta) = (T_b / N) \text{sinc}^2(\pi f T_b / N)$. Передача в эфире широкополосного сигнала обладает всеми преимуществами распределенного спектра (см. главы 3, 4), но на приемной стороне снятие расширения возвращает спектр в исходную полосу, превращая сигнал в узкополосный, что позволяет применять простейшие технологии демодуляции данных.

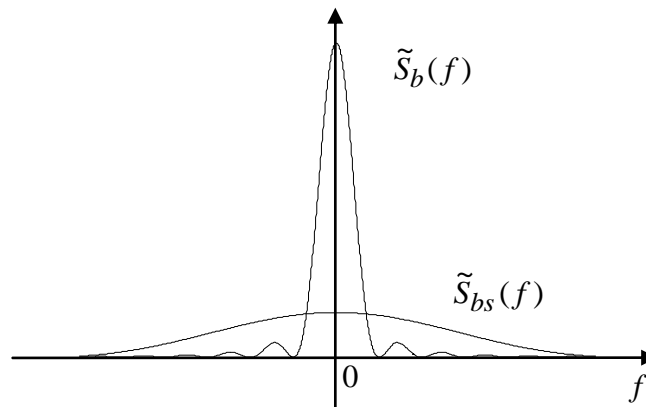


Рис.7.6. Спектр мощности исходного и расширенного потока

7.1.2. Прямое расширение: общий случай.

Идея прямого расширения спектра, рассмотренная выше для случая передачи данных с помощью бинарной ФМ и применения бинарных сигнатур, легко может быть обобщена на более широкий диапазон методов модуляции данных и сигнатур. Пусть $\dot{B}_k(t)$ обозначает комплексное колебание, отвечающее потоку данных \mathbf{b}_k k -го пользователя, передаваемых в некотором цифровом формате модуляции данных (АМ, M -ичная ФМ, КАМ и др.) При M -ичной цифровой передаче данных $\dot{B}_k(t)$ состоит из соприкасающихся прямоугольников длительности $T = (\log M)T_b$, манипулированных комплексными симво-

лами, принадлежащими конкретному алфавиту M -ичной модуляции. Например, в случае ФМ-8 прямоугольники длительности $T = 3T_b$ манипулированы комплексными амплитудами, принадлежащими алфавиту $\{\exp(jl\pi/4): l=0,1,\dots,7\}$, которые изображены на рис. 2.6, с. Если же выбрана манипуляция КАМ-16, тогда прямоугольники имеют длительность $T = 4T_b$, а их комплексные амплитуды принимают значения, определяемые рис. 2.6, b и т.д. В случае обычной (не широкополосной) M -ичной модуляции передаваемый сигнал, переносящий поток данных \mathbf{b}_k (теперь удобно полагать, что компоненты $b_{k,i}$ потока \mathbf{b}_k представляют собой M -ичные комплексные символы $\dot{B}_k(t) = b_{k,i}$, $(i-1)T < t \leq iT$), имеет вид

$$s_k(t; \mathbf{b}_k) = \text{Re}[\dot{B}_k(t) \exp(j2\pi f_0 t)].$$

Тогда принятый сигнал

$$s_{kr}(t; \mathbf{b}_k) = \text{Re}\{\dot{B}_k(t - \tau_k) \exp[j(2\pi f_0 t + \phi_k)]\}$$

обладает комплексной огибающей вида

$$\dot{S}_{kr}(t; \mathbf{b}_k) = \dot{B}_k(t - \tau_k) \exp(j\phi_k). \quad (7.5)$$

Передаваемый i -й символ данных $b_{k,i}$ есть ничто иное, как комплексная амплитуда непрерывной несущей, постоянной на интервале $((i-1)T, iT]$. Следовательно, для восстановления этого символа приемник должен принять решение относительно M копий прямоугольного импульса, существующего на интервале $((i-1)T + \tau_k, iT + \tau_k]$ и обладающего различными конкурирующими значениями комплексной амплитуды. Для выполнения указанной задачи необходимо вычислить корреляцию между наблюдаемой комплексной огибающей $\dot{Y}(t)$, выделенной из зашумленного наблюдения $y(t)$, и $\exp(j\phi_k)$, представимую в виде

$$\dot{z}_k = \int_{(i-1)T + \tau_k}^{iT + \tau_k} \dot{Y}(t) \exp(-j\phi_k) dt,$$

которая (после нормировки к энергии символа опорного образца) используется для вынесения необходимой оценки $\hat{b}_{k,i}$ о символе $b_{k,i}$.

Демодулятор на рис. 7.7, где двойной круг символизирует операцию комплексного

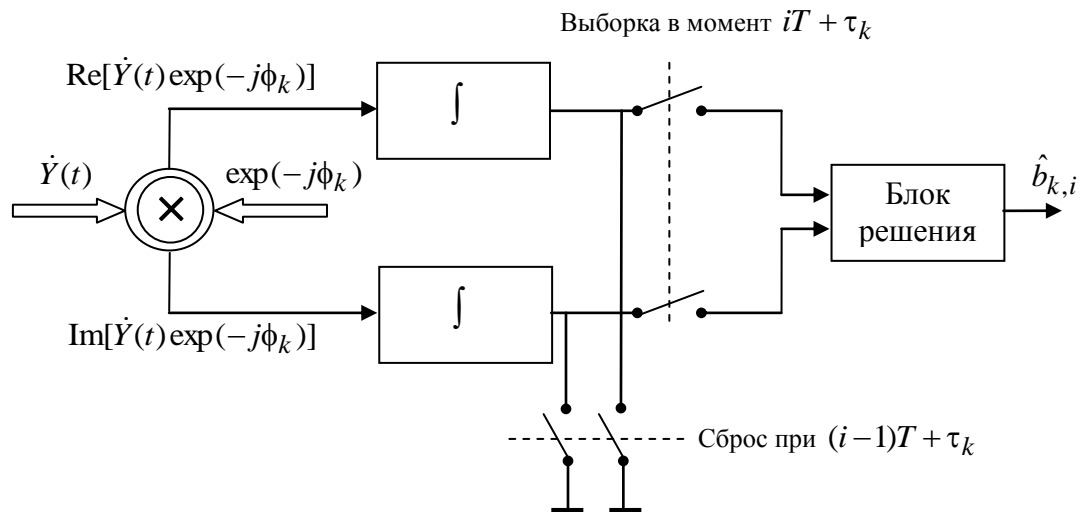


Рис.7.7. M -ичный демодулятор.

умножения, реализует указанную процедуру. Данная схема является обобщением корреляционной

ляционной структуры, представленной на рис. 7.1, *b*, на случай произвольной цифровой модуляции данных. Первоначально наблюдаемая комплексная огибающая $\dot{Y}(t)$ умножается на $\exp(-j\varphi_k)$ с целью компенсации канального фазового сдвига φ_k . После этой операции полезная компонента (7.5) наблюдения $\dot{Y}(t)$ становится $\dot{B}_k(t - \tau_k)$. Любое комплексное колебание эквивалентно двум вещественным (реальной и мнимой частям), так что выход умножителя на рис. 7.7 трактуется в терминах реальной и мнимой частей произведения $\dot{Y}(t)\exp(-j\varphi_k)$. Полезными компонентами произведения служат реальная и мнимая части $\dot{B}_k(t - \tau_k)$, а последующее их интегрирование, как и ранее, служит выделению из шума. Отсчеты с выхода интегратора в моменты $iT + \tau_k$ являются оценками реальной и мнимой частей принятого M -ичного символа данных, которые используются в блоке решения для выдачи демодулированного символа.

Рассмотрим теперь, как прямое расширение спектра может быть включено в данную схему модуляции – демодуляции. Пусть $\dot{S}_k(t)$ будет комплексной огибающей k -й пользовательской CDMA сигнатурой. Ее алфавит может быть выбран независимо от алфавита модуляции данных, например, может быть бинарным, квадратурным, АФМ и т.д. Тогда прямое расширение означает умножение колебаний модулированных данных $\dot{B}_k(t)$ на сигнатуру $\dot{S}_k(t)$ с целью использования их произведения $\dot{S}_k(t)\dot{B}_k(t)$ как комплексной огибающей передаваемого сигнала:

$$s_k(t; \mathbf{b}_k) = \text{Re}[\dot{S}_k(t)\dot{B}_k(t)\exp(j2\pi f_0 t)]. \quad (7.6)$$

Принятый полезный сигнал представляет собой задержанную и сдвинутую по фазе копию сигнала (7.6)

$$s_{kr}(t; \mathbf{b}_k) = \text{Re}[\dot{S}_k(t - \tau_k)\dot{B}_k(t - \tau_k)\exp[j(2\pi f_0 t + \varphi_k)]]$$

с комплексной огибающей

$$\dot{S}_{kr}(t; \mathbf{b}_k) = \dot{S}_k(t - \tau_k)\dot{B}_k(t - \tau_k)\exp(j\varphi_k). \quad (7.7)$$

Постоянство $\dot{B}_k(t - \tau_k)$ на интервале $((i-1)T + \tau_k, iT + \tau_k]$ снова означает, что для выделения i -го символа приемник должен принять решение об истинности одной из M конкурирующих копий одной и той же сигнатуры $\dot{S}_k(t - \tau_k)\exp(-j\varphi_k)$, умноженной на различные символы данных $b_{k,i}$. Тогда корреляция вида

$$\dot{z}_k = \int_{(i-1)T + \tau_k}^{iT + \tau_k} \dot{Y}(t)\dot{S}_k^*(t - \tau_k)\exp(-j\varphi_k)dt \quad (7.8)$$

послужит (после соответствующей нормировки) для получения необходимой оценки $b_{k,i}$ и снова может трактоваться, как пара отсчетов с выходов интеграторов демодулятора, изображенного на рис. 7.7, при изменении опорного сигнала в комплексном умножителе с $\exp(-j\varphi_k)$ на $\dot{S}_k^*(t - \tau_k)\exp(-j\varphi_k)$. После умножения на подобный опорный сигнал полезная компонента наблюдаемой комплексной огибающей

$$\dot{S}_{kr}(t; \mathbf{b}_k)\dot{S}_k^*(t - \tau_k) = |\dot{S}_k(t - \tau_k)|^2 \dot{B}_k(t - \tau_k)\exp(j\varphi_k)$$

на интервале i -го символа данных становится просто одной из M возможных копий видеосигнала $|\dot{S}_k(t - \tau_k)|^2$, умноженного на различные комплексные коэффициенты $b_{k,i}$. Если чипы сигнатуры не имеют амплитудной модуляции, т.е. $\dot{S}_k(t)$ подвергается только

фазовой манипуляции, то $|\dot{S}_k(t - \tau_k)|^2 = 1$ и, как показывает предыдущее соотношение, умножение на $\dot{S}_k^*(t - \tau_k)$ превращает комплексную огибающую принятого сигнала в одну из характеристик обычной (не широкополосной) M -ичной модуляции данных, т.е. осуществляет снятие расширения. Благодаря этому приемник снова может рассматриваться как двухэтапный: первоначально осуществляется снятие расширения, а затем обычная M -ичная демодуляция, использующая, например, схему на рис. 7.7.

Остановимся более подробно на реализации комплексного умножения и выделении комплексной огибающей $\dot{Y}(t)$ из действительно наблюдаемого вещественного колебания $y(t)$. Вспомнив основные правила комплексной арифметики

$$\operatorname{Re}(xy) = \operatorname{Re}(x)\operatorname{Re}(y) - \operatorname{Im}(x)\operatorname{Im}(y), \quad \operatorname{Im}(xy) = \operatorname{Re}(x)\operatorname{Im}(y) + \operatorname{Im}(x)\operatorname{Re}(y),$$

видно, что множитель двух комплексных величин x и y содержит четыре обычных ум-

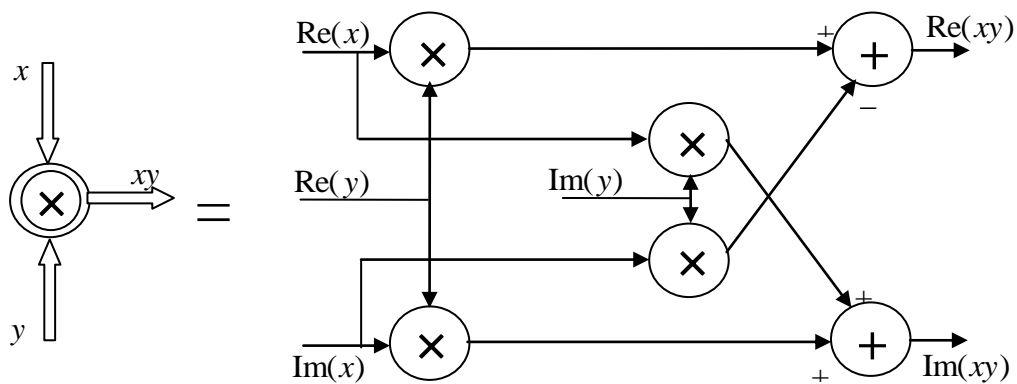


Рис.7.8. Комплексный умножитель.

ножителя и два сумматора (рис. 7.8). Входные комплексные величины x, y определены своими реальными и мнимыми частями, а результат представляет собой двумерный вектор, состоящий из вещественной и мнимой частей произведения xy .

Получение комплексной огибающей наблюдения основывается на определении $\dot{Y}(t)$: $y(t) = \operatorname{Re}[\dot{Y}(t)\exp(j2\pi f_0 t)]$. Применив вышеприведенное правило комплексного умножения и формулу Эйлера, получаем $y(t) = \operatorname{Re}[\dot{Y}(t)]\cos(2\pi f_0 t) - \operatorname{Im}[\dot{Y}(t)]\sin(2\pi f_0 t)$. Умножение обеих частей этого выражения на $2\cos(2\pi f_0 t)$ и $-2\sin(2\pi f_0 t)$, а также применение тригонометрического преобразования приводит к следующему результату

$$\begin{aligned} 2y(t)\cos(2\pi f_0 t) &= \operatorname{Re}[\dot{Y}(t)] + \operatorname{Re}[\dot{Y}(t)]\cos(4\pi f_0 t) - \operatorname{Im}[\dot{Y}(t)]\sin(4\pi f_0 t) \\ -2y(t)\sin(2\pi f_0 t) &= \operatorname{Im}[\dot{Y}(t)] - \operatorname{Im}[\dot{Y}(t)]\cos(4\pi f_0 t) - \operatorname{Re}[\dot{Y}(t)]\sin(4\pi f_0 t). \end{aligned} \quad (7.9)$$

Первые слагаемые в правых частях соотношений (7.9) представляют собой низкочастотные колебания (поскольку комплексная огибающая определяет закон модуляции, т.е. является низкочастотной), тогда как остальные являются полосными сигналами с центральной частотой $2f_0$. Полоса закона модуляции значительно меньше f_0 (см. рис. 7.9, а). Следовательно, фильтр низких частот легко может отфильтровать высокочастотные компоненты в (7.9), пропустив только вещественную и мнимую части желаемой комплексной огибающей $\dot{Y}(t)$. Данный принцип восстановления комплексной огибающей из реального наблюдения $y(t)$ реализуется схемой, представленной на рис. 7.9, б.

В качестве заключения приведенному обсуждению рис. 7.10 иллюстрирует опера-

ции, выполняемые на передающей и приемной сторонах обычной широкополосной системы с прямым расширением спектра. Модулятор (рис.7.10, а) реализует алгоритм, опре-

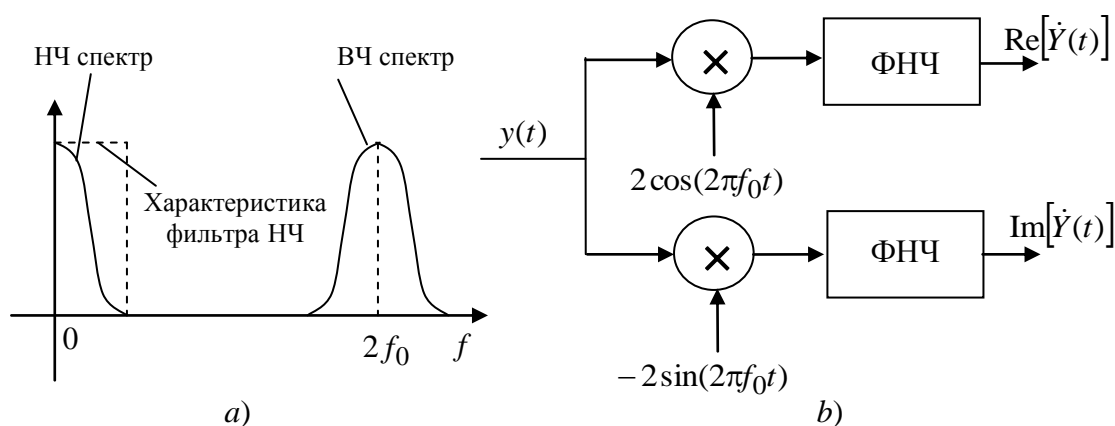


Рис.7.9. Восстановление комплексной огибающей.

деляемый соотношением (7.6), используя только вещественную часть комплексного произведения. В демодуляторе (рис. 7.10, б) комплексная огибающая наблюдения, восстанавливаемая согласно схеме на рис. 7.9, б, в дальнейшем подается на вход обычного M -ичного демодулятора (см. рис. 7.7) для принятия решения о принятых символах.

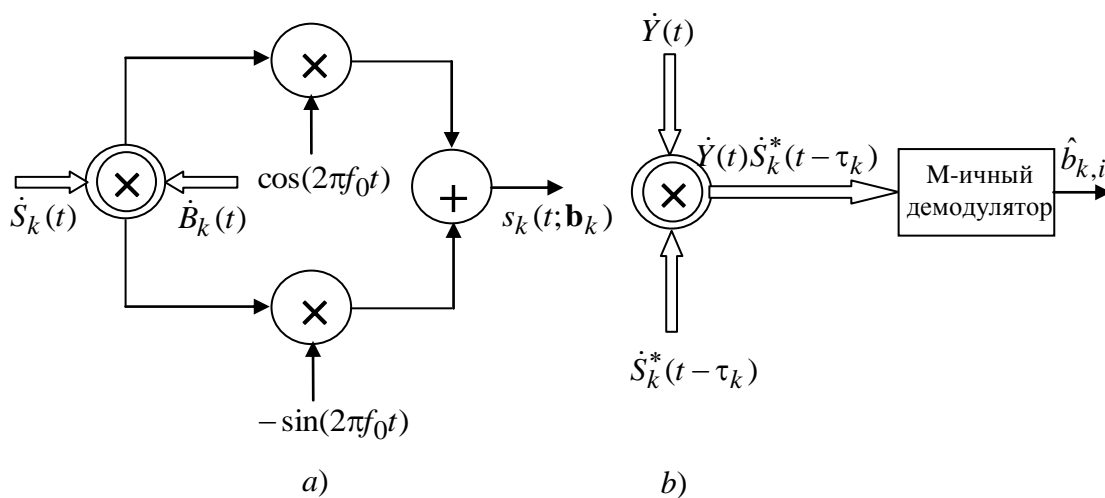


Рис.7.10 Модуляция (а) и демодуляция (б) при прямом расширении спектра.

Отметим также, что только что описанный способ реализации расширения–сжатия не является единственным и что существует многообразие конкретных схемных решений этих операций. Например, перемножение комплексных огибающих может быть выполнено неявным образом в ходе процедуры гетеродинирования. В частности, если $u_i(t) = U_i(t) \cos[2\pi f_i t + \gamma_i(t)]$, $i = 1, 2$ – два полосных сигнала с несущими частотами f_i и комплексными огибающими $\dot{U}_i(t) = U_i(t) \exp[j\gamma_i(t)]$, то их произведение будет

$$u_1(t)u_2(t) = \frac{1}{2}U_1(t)U_2(t) \cos[2\pi(f_1 - f_2)t + \gamma_1(t) - \gamma_2(t)] + \frac{1}{2}U_1(t)U_2(t) \cos[2\pi(f_1 + f_2)t + \gamma_1(t) + \gamma_2(t)].$$

Два слагаемых в последнем выражении представляют собой полосные сигналы с несущи-

ми частотами $f_1 - f_2$ и $f_1 + f_2$. Если нижняя несущая частота $f_1 - f_2$ превосходит полосу произведения $\dot{U}_1(t)\dot{U}_2^*(t)$, то после отфильтровывания слагаемого с более высокой частотой оставшийся полосный сигнал с более низкой частотой будет обладать комплексной огибающей вида $\dot{U}_1(t)\dot{U}_2^*(t)$, т.е. в точности произведением, получаемым после снятия расширения. Аналогичным образом, слагаемое с более высокой частотой представляет собой полосный сигнал, комплексная огибающая которого является аналогичным произведением, но без комплексного сопряжения.

Прямое расширение используется во всех CDMA стандартах 2-го и 3-го поколений: IS-95 (cdmaOne), UMTS и cdma2000. В них используются различные комбинации алфавитов манипулированных данных и сигнатур, которые более подробно будут рассмотрены в параграфе 11.2.

7.1.3. Расширение спектра прыгающей частотой.

При расширении спектра с помощью прыгающей частоты (*frequency hopping* (FH)) используются ЧМ сигнатуры, а модуляция данных, как правило, также осуществляется с помощью ЧМ. Традиционно различают два типа расширения прыгающей частотой: *быстрое* (*fast*) и *медленное* (*slow*), критерием классификации которой служит соотношение между длительностями чипа Δ и символа данных T . Для быстрой FH $\Delta = T/l$, где $l > 1$ – натуральное число, тогда как для медленной FH $\Delta = lT$, где $l \geq 1$ – натуральное. Другими словами, при быстрой FH на один символ данных приходится несколько скачков частоты, тогда как при медленной FH в течение одного частотного скачка сигнатуры могут быть переданы несколько символов данных. Для лучшего понимания сущности этого типа расширения обратимся к примеру.

Пример 7.1.1. Возьмем ЧМ сигнатуру примера 5.5.1 и используем ее для расширения спектра с помощью быстрой FH в комбинации с бинарной модуляцией данных. В этом случае число различных частот в сигнатуре $M = 5$, длина сигнатуры $N = 8$, а один символ данных передает один бит информации, так что $T = T_b$. Предположим, что в схеме быстрой FH $l = N = 8$, т.е. на один бит данных приходится 8 скачков частоты. Тогда полная последовательность ЧМ чипов, представленная на рис. 5.3, передается в течение одного бита. Если бит данных равен нулю, то частотный образец излучается на несущей частоте, равной f_0 , тогда как для бита, равного единице, несущая частота изменяется на f_1 . Очевидно, что разность между частотами f_1 и f_0 должна быть не меньше полосы, занимаемой сигнатурой, т.е. MF . На рис. 7.11, *a* изображен передаваемый частотный образец, соответствующий потоку битовых данных вида 01011. Как можно видеть, спектр одиночного бита данных, полоса которого перед расширением была примерно $1/T_b$, расширяется до полосы, равной $MF \approx M/\Delta = MN/T_b$, т.е. становится в MN раз шире (см. параграф 5.6).

На приемной стороне снятие расширения заключается в обратном преобразовании наблюдаемого колебания на промежуточную частоту f_i . С этой целью используется опорное колебание несущей частоты $f_0 - f_i$, модулированное согласно ЧМ образца сигнатуры с необходимой задержкой во времени (рис. 7.11, *b*). В результате сигнал со снятым расширением представляет собой обычное узкополосное колебание, частотно-манипулированное в соответствии с передаваемыми данными, где нулевой бит данных передается более низкой частотой, чем бит, отвечающий единице. Теперь спектр отдельного символа данных возвращается к полосе $1/T_b$ и может быть использован обычный бинарный ФМ демодулятор для восстановления принятых данных. \square

Следующий пример иллюстрирует алгоритм расширения спектра медленной FH.

Пример 7.1.2. Используем ту же самую сигнатуру снова в комбинации бинарной ФМ данных и приравняем длительности чипа и символа данных: $T = T_b = \Delta$. Последнее означает, что текущая частота остается постоянной в течение всей длительности бита данных, а скачок частоты происходит только при переходе от одного бита к следующему.

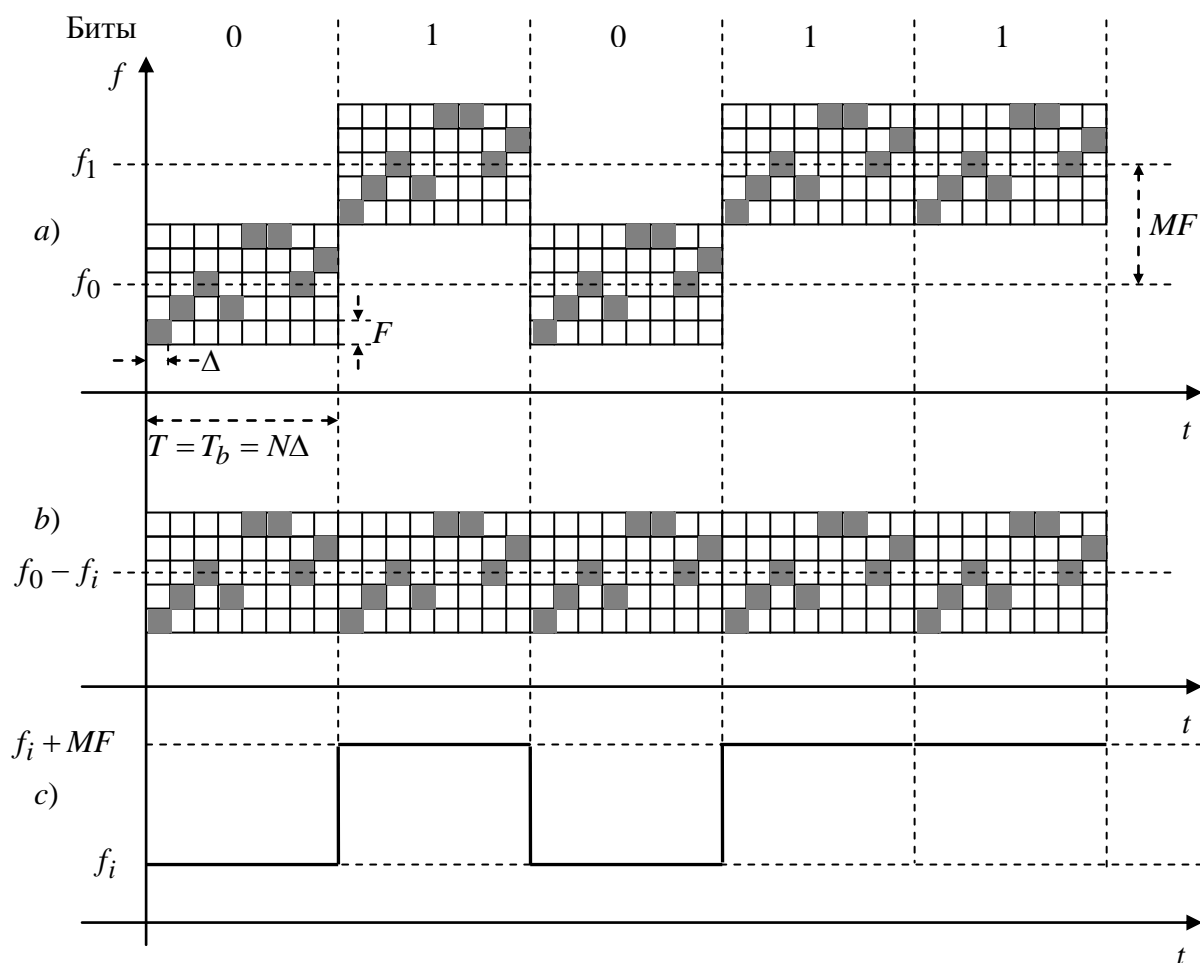


Рис.7.11. Расширение-сужение с помощью быстрой FH (fast FH).

Частотный образец сигнатуры растягивается во времени и его длина охватывает N бит данных (рис. 7.12, *a*). Предположим, что в течение бита данных с номером i частота сигнатуры равна F_i . Тогда частота передачи становится равной $f_0 + F_i$ в случае нулевого бита данных и $f_1 + F_i$ для бита данных, равного единице. Рис. 7.12, *b* служит иллюстрацией этого для битового потока 00101101. Принципиальное отличие между быстрой и медленной FH теперь очевидно: последняя не расширяет спектр отдельного символа данных, увеличивая только полосу, занимаемую системой. Это аналогично тому, как если бы система просто время от времени переключалась бы с одной рабочей частоты на другую, но не происходило бы переключения на интервале фиксированной группы символов данных. На приемной стороне обратное преобразование на промежуточную полосу f_i выполняется с помощью опорного сигнала на частоте $f_0 - f_i$, повторяющего частотный образец сигнатуры (задержанным во времени соответствующим образом) (рис. 7.12, *c*). Данная операция возвращает колебание в полосу, соответствующую простой (без скачков частоты) ЧМ модуляции данных (рис. 7.12, *d*), так что для восстановления передаваемых данных может быть использован обычный ЧМ демодулятор (рис. 7.12, *e*). □

Методы, иллюстрируемые приведенными выше примерами для передачи данных с

помощью бинарной ФМ, легко обобщаются на случай модуляции данных произвольной ФМ (см. задачи 7.5-7.7).

Расширение спектра с помощью прыгающей частоты обладает некоторыми осо-

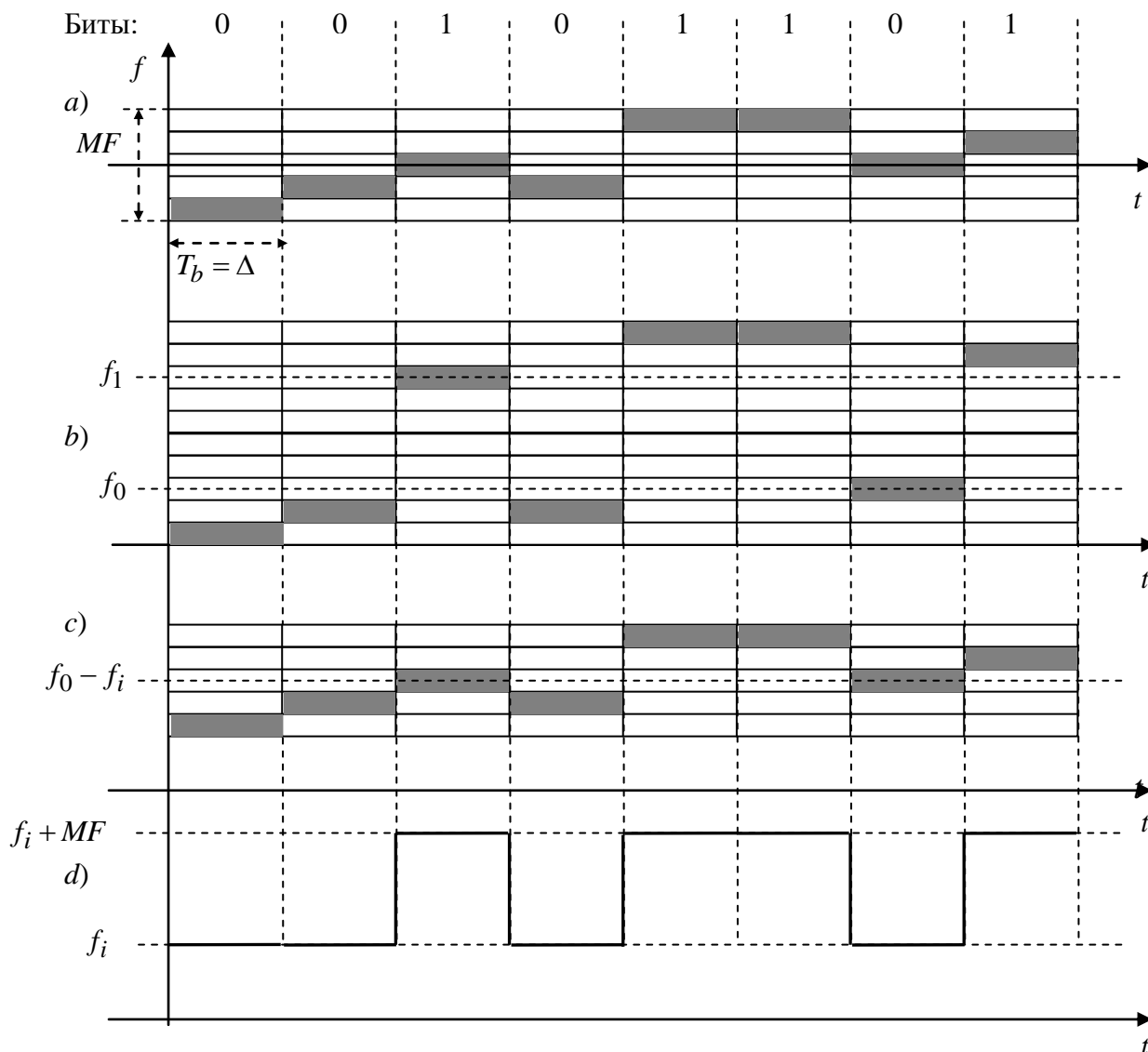


Рис.7.12. Расширение-сжатие с помощью медленной FH.

бенностями, делающими этот метод особенно привлекательным для военных приложений, в частности в различных антагонистических сценариях борьбы против систем подавления [3, 6]. В то же время, его коммерческое использование до недавнего времени совершенно незначительно, по крайней мере в отношении FFH. Однако внедрение технологии Bluetooth [55] указывает, что этот тип расширения спектра также обладает хорошими коммерческими перспективами.

7.2. Синтез ансамблей сигнатур для синхронного CDMA с прямым расширением спектра.

7.2.1. Постановка задачи.

Рассмотрим K –пользовательскую CDMA систему с прямым расширением спектра, в которой все пользовательские потоки данных и все сигнатуры жестко синхронизованы, т.е. характеризуются нулевым взаимным временным сдвигом на входе приемника. Как указывалось в параграфе 4.4, классическим примером подобной системы может служить канал «вниз» система мобильной радиосвязи с CDMA, в которой базовая станция полностью контролирует временное положение сигналов, адресованных всем пользователям в пределах соты. Очевидно, что групповой сигнал достигает приемника мобильного абонента с сохранением начальной синхронизации между сигналами, адресованными различным индивидуальным пользователям. В приводимом здесь анализе будем считать, что имеет место идеализированная модель канала, в котором многолучевая задержка на распространение τ_{\max} меньше периода следования чипов Δ пользовательских сигнатур или используется эффективный эквалайзер, исключаяющий все многолучевые компоненты, задержка которых превышает Δ . Данное предположение позволяет игнорировать любое потенциальное нарушение идеальности синхронизации компонент принятого сигнала.

Согласно концепции прямого расширения спектра, комплексная огибающая принятого группового сигнала $\dot{S}(t; \mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_K)$ представляет собой сумму по k комплексных огибающих сигнатур, манипулированных пользовательскими потоками данных, каждая из которых определяется соотношением (7.6). В общем случае каждый из пользовательских сигналов может обладать собственной амплитудой, однако ограничимся простейшим случаем равной интенсивности сигналов. Поскольку предположение об идеальном синхронизме позволяет положить все задержки τ_k и начальные фазы φ_k в (7.6) равными нулю, получаем

$$\dot{S}(t; \mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_K) = \sum_{k=1}^K \dot{B}_k(t) \dot{S}_k(t). \quad (7.10)$$

Сосредоточим наше внимание на интервале длительности T одиночного символа данных. Снова, вследствие полного синхронизма, начало и конец текущих символов данных всех пользователей строго совпадают. Если через b_k обозначить k –й текущий пользовательский символ данных, то на одиночном символьном интервале (7.10) может быть записано в следующем виде

$$\dot{S}(t; \mathbf{b}) = \sum_{k=1}^K b_k(t) \dot{S}_k(t), \quad (7.11)$$

где $\mathbf{b} = (b_1, b_2, \dots, b_K)$ – K –мерный вектор текущих символов данных всех пользователей.

Вспомним теперь, что каждая сигнатура в CDMA системе с прямым расширением является АФМ сигналом, описываемым моделью (5.2):

$$\dot{S}(t) = \sum_{i=0}^{N-1} a_{k,i} \dot{S}_0(t - i\Delta), \quad (7.12)$$

где $\{a_{k,0}, a_{k,1}, \dots, a_{k,N-1}\}$ – кодовая последовательность, манипулирующая чипы k –й сигнатуры, а N – коэффициент расширения, т.е. число чипов на один символ данных. Рис. 7.13 демонстрирует строгое взаимное расположение между чипами сигнатуры и границами передаваемых символов данных при синхронном варианте CDMA.

Основываясь на (7.11)–(7.12), можно сформулировать несколько подходов к выбо-

ру ансамблей сигнатур для сетей с синхронным CDMA и прямым расширением спектра. Среди основных факторов, оказывающих влияние на процедуру и результаты оптимизации множества сигнатур, следует отметить связь между числом пользователей K и пока-

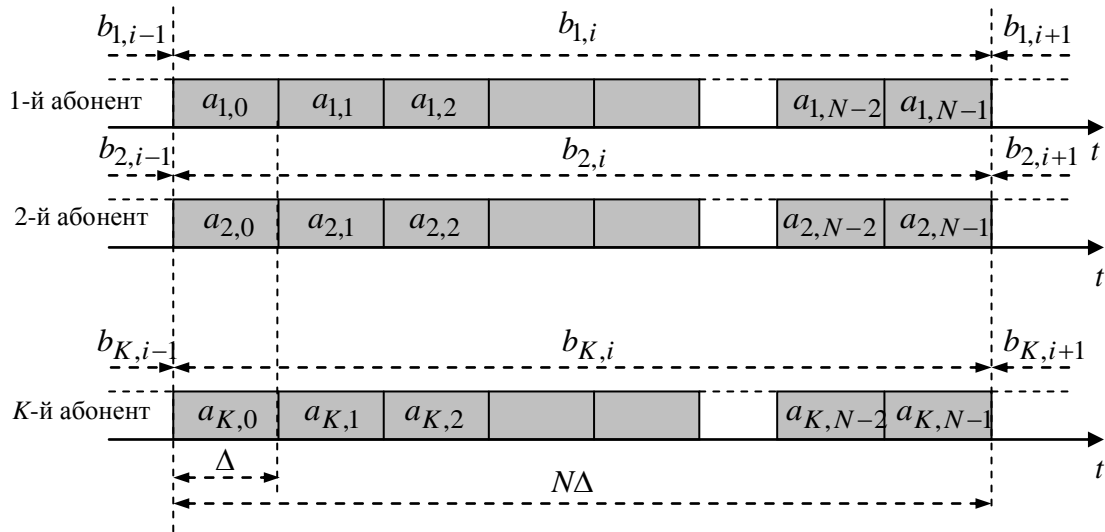


Рис.7.13. Положение символа данных и чипа в синхронной CDMA.

зателем расширения N , а также алгоритмом приема (многопользовательский или обычный).

7.2.2. Оптимизация множества сигнатур по критерию минимума расстояния.

Предположим, что допустим приемник любой сложности и, следовательно, разрешается использовать оптимальный (многопользовательский) алгоритм оценивания вектора данных \mathbf{b} , основанный на поиске значения $\hat{\mathbf{b}}$, минимизирующего расстояние между наблюдением $y(t)$ и претендентом группового сигнала $s(t; \mathbf{b})$ (см. параграф 4.1). В терминах комплексной огибающей последнее означает минимизацию по \mathbf{b} квадрата расстояния $\|\dot{Y}(t) - \dot{S}(t; \mathbf{b})\|^2$, где $\dot{S}(t; \mathbf{b})$ определяется (7.11). Тогда строгое теоретическое обоснование очевидным образом указывает на нахождение ансамбля K сигнатур $\{\dot{S}_1(t), \dot{S}_2(t), \dots, \dot{S}_K(t)\}$, минимизирующего вероятность ошибки в оценке $\hat{\mathbf{b}}$ K -пользовательского вектора данных $\mathbf{b} = (b_1, b_2, \dots, b_K)$. Возвращаясь к материалу параграфа 2.3, вспомним, что асимптотически (т.е. при достаточно большом отношении сигнал-шум), минимизация вероятности ошибки эквивалентна максимизации минимума расстояния в созвездии из M передаваемых сигналов. В рассматриваемом случае подлежащие обсуждению альтернативные сигналы представляют собой копии (7.11), отвечающие различным векторам данных \mathbf{b} . Следовательно, задачу оптимизации множества сигнатур можно сформулировать как задачу максимизации минимума квадрата расстояния:

$$d_{\min}^2 = \min_{\substack{\mathbf{b}, \mathbf{b}' \\ \mathbf{b} \neq \mathbf{b}'}} d^2(\mathbf{b}, \mathbf{b}') = \max, \quad (7.13)$$

где минимальное расстояние d_{\min} ищется по всем различным парам векторов данных $\mathbf{b} = (b_1, b_2, \dots, b_K)$, $\mathbf{b}' = (b'_1, b'_2, \dots, b'_K)$, $\mathbf{b} \neq \mathbf{b}'$ и, согласно (2.43), определяется как

$$d^2(\mathbf{b}, \mathbf{b}') = \|s(t; \mathbf{b}) - s(t; \mathbf{b}')\|^2 = \frac{1}{2} \|\dot{S}(t; \mathbf{b}) - \dot{S}(t; \mathbf{b}')\|^2. \quad (7.14)$$

Рассмотрим подробно передачу бинарных данных, когда символами являются би-

ты, непосредственно передаваемые с помощью бинарной ФМ, так что $b_k, b'_k = \pm 1, k = 1, 2, \dots, K$. Подобное сужение диапазона упрощает анализ битов, допуская непосредственный переход к последующему расширению до произвольной ФМ. Тогда использование (7.11), (2.41) и (2.42) в (7.14) приводит к следующему результату

$$d^2(\mathbf{b}, \mathbf{b}') = \frac{1}{2} \left\| \sum_{k=1}^K \varepsilon_k \dot{S}_k(t) \right\|^2 = \frac{1}{2} \int_0^T \left| \sum_{k=1}^K \varepsilon_k \dot{S}_k(t) \right|^2 dt = E_b \sum_{k=1}^K \sum_{l=1}^K \varepsilon_k \varepsilon_l \dot{\rho}_{kl}, \quad (7.15)$$

где $\varepsilon_k = b_k - b'_k$ принимает одно из трех возможных значений: 0 или ± 2 ;

$E_b = \frac{1}{2} \int_0^T |\dot{S}_k(t)|^2 dt$ – энергия k -й сигнатуры, затрачиваемая для передачи одного бита (будем полагать, что эти энергии одинаковы для всех K пользователей); и

$$\dot{\rho}_{kl} = \frac{1}{2E_b} \int_0^T \dot{S}_k(t) \dot{S}_l^*(t) dt$$

коэффициент корреляции между комплексными огибающими k -й и l -сигнатур. При использовании свойств коэффициента корреляции, вытекающих из его определения: $\dot{\rho}_{kk} = 1$ и $\dot{\rho}_{kl} = \dot{\rho}_{lk}^*$, соотношение (7.15) принимает вид, ясно свидетельствующий, что расстояние всегда вещественно:

$$d^2(\mathbf{b}, \mathbf{b}') = E_b \sum_{k=1}^K \varepsilon_k^2 + 2E_b \sum_{k=1}^{K-1} \sum_{l=k+1}^K \varepsilon_k \varepsilon_l \operatorname{Re}(\dot{\rho}_{kl}). \quad (7.16)$$

Возьмем два вектора данных (битовых образца) \mathbf{b}, \mathbf{b}' , отличающихся только одним элементом, например, первым компонентом. Тогда $\varepsilon_k = 0, k = 2, 3, \dots, K, \varepsilon_1 = \pm 2$ и из (7.16) получаем, что $d^2(\mathbf{b}, \mathbf{b}') = 4E_b$. Поскольку d_{\min}^2 никогда не превосходит значения квадрата расстояния для некоторой конкретной пары \mathbf{b}, \mathbf{b}' , то

$$d_{\min}^2 \leq 4E_b. \quad (7.17)$$

Данная верхняя граница свидетельствует, что ансамбль сигнатур, для которого $d_{\min}^2 = 4E_b$, следует считать оптимальным по критерию максимума минимального расстояния (7.13). Одним из достаточных условий достижения границы (7.17) является *слабая* (weak) ортогональность комплексных огибающих сигнатур:

$$\operatorname{Re}(\dot{\rho}_{kl}) = \begin{cases} 1, & k = l, \\ 0, & k \neq l, \end{cases} = \delta_{kl}. \quad (7.18)$$

Причина, по которой комплексные огибающие, удовлетворяющие (7.18), называются слабо ортогональными, становится ясной после сравнения выражений (7.18) и (2.46) ($\dot{\rho}_{kl} = \delta_{kl}$). Последнее является более жестким и требует, чтобы сигналы $s_k(t), s_l(t)$ с комплексными огибающими $\dot{S}_k(t), \dot{S}_l(t)$ сохраняли ортогональность при любом взаимном фазовом сдвиге. В то же время два сигнала, промодулированные $\dot{S}(t)$ и $\dot{S}(t) \exp(j\frac{\pi}{2}) = j\dot{S}(t)$, т.е. просто квадратурными (сдвинутыми по фазе на $\pi/2$) копиями одного и того же сигнала, являются ортогональными, однако теряют свойство ортогональности, если их взаимный фазовый сдвиг отличается от $\pm\pi/2$. Следовательно, $\dot{S}(t)$ и $j\dot{S}(t)$ являются только слабо ортогональными. Конечно, любые ортогональные (в смысле (2.46)) сигнатуры являются слабо ортогональными, но не наоборот.

Для сигнатур, удовлетворяющих (7.18), соотношение (7.16) приводится к виду

$d^2(\mathbf{b}, \mathbf{b}') = E_b \sum_{k=1}^K \varepsilon_k^2$. По крайней мере, одно из слагаемых в данной сумме не является нулевым, так что $d_{\min}^2 \geq 4E_b$, что совместно с (7.17) дает $d_{\min}^2 = 4E_b$. Следствием этого является тот факт, что ансамбль из K слабо ортогональных сигнатур оптимален по критерию минимума расстояния, а, значит, (асимптотически) по вероятности перепутывания различных пользовательских битовых образцов.

Существует множество способов построения ортогональных (удовлетворяющих (2.46)) широкополосных сигналов различной длины (коэффициента расширения) N . Одним из примеров являются функции Уолша или, в общем случае, матрицы Адамара, рассмотренные в 2.7.3, которые позволяют построить ортогональные бинарные коды. Другим возможным выбором служат циклически сдвинутые копии любой последовательности с идеальной периодической АКФ, например, троичной, многофазной, др. (см. параграф 6.11). Любой ансамбль из K' ортогональных сигнатур тривиальным образом трансформируется во множество слабо ортогональных сигнатур, состоящее из $2K'$ сигналов, путем добавления квадратурных копий каждого из сигналов: факт, неоднократно упоминавшийся ранее (см. параграфы 2.5, 4.1).

При любом конкретном выборе ортогональных сигнатур размерность сигнального пространства жестко ограничивает их число (а, значит, и число пользователей K) (см. параграф 2.5). Согласно соотношению (7.12), при заданной форме чипа N -мерный вектор $\mathbf{a}_k = (a_{k,0}, a_{k,1}, \dots, a_{k,N-1})$ k -й кодовой последовательности полностью определяет k -ю сигнатуру, и ортогональность k -й и l -й сигнатур эквивалентна ортогональности векторов $\mathbf{a}_k, \mathbf{a}_l$. Действительно, повторение вывода (2.52) для комплексных огибающих (альтернативно использованию (5.7)) позволяет записать скалярное произведение $\dot{S}_k(t), \dot{S}_l(t)$ как

$$(\dot{S}_k, \dot{S}_l) = 2E_0 \sum_{i=0}^{N-1} a_{k,i} a_{l,i}^* = 2E_0 (\mathbf{a}_k, \mathbf{a}_l), \quad (7.19)$$

подтверждая тем самым, что ортогональность векторов $\mathbf{a}_k, \mathbf{a}_l$ является необходимым и достаточным условием ортогональности $\dot{S}_k(t), \dot{S}_l(t)$. Размерность N пространства векторов \mathbf{a}_k кодовых последовательностей явным образом определяет максимальное число K' ортогональных сигнатур $\dot{S}_k(t)$. Еще раз подчеркнем, что если допустимо квадратурное расщепление каждой сигнатуры, то максимальное число пользователей, размещаемое в рамках ансамбля сигнатур, определяется как $K = 2K' = 2N$. Однако, если по каким-либо причинам точный фазовый сдвиг $\pm \pi/2$ между квадратурными копиями сигнатур не может быть поддержан, то слабой ортогональности не достаточно, и максимальное число пользователей будет в два раза меньше: $K = K' = N$.

Отметим, что слабая ортогональность является только достаточным, но не необходимым условием равенства в (7.17), и, в частности, особенно интересен вопрос о возможности достижения верхней границы (7.17) при числе сигнатур, превышающем размерность сигнального пространства n_s . Как следует из предшествующего обсуждения, значение n_s равняется либо $2N$, либо N в зависимости от допустимости или нет квадратурного расщепления сигнатур. Синхронные CDMA системы, в которых $K > n_s$, называются *перенасыщенными (oversaturated)*, подчеркивая, что избыточное число вовлеченных кодовых векторов исключает вероятность их ортогональности (возможно слабой).

Возможность и алгоритм достижения минимального расстояния, равного верхнему пределу в (7.17) в перенасыщенных системах были доказаны в [56]. Для более прозрачного обсуждения данной идеи и упрощения используемого обозначения первоначально отвергнем возможность удвоения размерности сигнального пространства за счет квадратур-

ного расщепления, полагая $n_s = N$. Возьмем N ортонормированных N -мерных векторов $\mathbf{a}_k, k = 1, 2, \dots, N$, таких, что $(\mathbf{a}_k, \mathbf{a}_l) = \delta_{kl}$, и прибавим к ним еще один вектор, образованный как

$$\mathbf{a}_{N+1} = \frac{1}{\sqrt{N}} \sum_{k=1}^N \mathbf{a}_k. \quad (7.20)$$

Использование полученных таким образом $N+1$ векторов $\mathbf{a}_k, k = 1, 2, \dots, N+1$ для формирования согласно (7.12) $K = N+1$ сигнатур означает образование $(N+1)$ -й сигнатуры в виде

$$\dot{S}_{N+1}(t) = \frac{1}{\sqrt{N}} \sum_{k=1}^N \dot{S}_k(t),$$

и, модулируя все сигнатуры бинарными символами данных $b_k = \pm 1$, получаем групповой сигнал как

$$\dot{S}(t; \mathbf{b}) = \sum_{k=1}^{N+1} b_k \dot{S}_k(t) = \sum_{k=1}^N b_k \dot{S}_k(t) + \frac{1}{\sqrt{N}} b_{N+1} \sum_{k=1}^N \dot{S}_k(t) = \sum_{k=1}^N \left(b_k + \frac{1}{\sqrt{N}} b_{N+1} \right) \dot{S}_k(t). \quad (7.21)$$

Разность между двумя версиями группового сигнала, отвечающим двум битовым образцам $\mathbf{b} = (b_1, b_2, \dots, b_{N+1})$, $\mathbf{b}' = (b'_1, b'_2, \dots, b'_{N+1})$, составит

$$\dot{S}(t; \mathbf{b}) - \dot{S}(t; \mathbf{b}') = \sum_{k=1}^N \left(\varepsilon_k + \frac{1}{\sqrt{N}} \varepsilon_{N+1} \right) \dot{S}_k(t), \quad \varepsilon_k = b_k - b'_k = 0, \pm 2.$$

Используя тот же подход, что и при выводе (7.15), и ортогональность первых N сигнатур, приходим к выражению

$$d^2(\mathbf{b}, \mathbf{b}') = \|\dot{S}(t; \mathbf{b}) - \dot{S}(t; \mathbf{b}')\|^2 = E_b \sum_{k=1}^N \left(\varepsilon_k + \frac{1}{\sqrt{N}} \varepsilon_{N+1} \right)^2. \quad (7.22)$$

Поскольку битовые образцы различны, то, по крайней мере, одна из величин $\varepsilon_k, k = 1, 2, \dots, N+1$ отлична от нуля, т.е. равняется ± 2 . Если $\varepsilon_{N+1} = 0$, то таковой является ε_k одна из множества $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N$, и тогда $d^2(\mathbf{b}, \mathbf{b}') \geq 4E_b$. Если же $\varepsilon_{N+1} = \pm 2$, то слагаемые в (7.22) с $\varepsilon_k = 0$ равны $4/N$, тогда как все остальные будут $4(\sqrt{N} \pm 1)^2/N$, приводя к $d^2(\mathbf{b}, \mathbf{b}') \geq 4E_b \min \left\{ 1, (\sqrt{N} - 1)^2 \right\}$. Учитывая полученные результаты, приходим к следующей нижней оценке минимума квадрата расстояния

$$d_{\min}^2 \geq \min \left\{ 4E_b, 4(\sqrt{N} - 1)^2 E_b \right\} = \begin{cases} 4(\sqrt{N} - 1)^2 E_b, & N < 4, \\ 4E_b, & N \geq 4. \end{cases}$$

Сравнение полученного результата с (7.17) показывает возможность добавления еще одной сигнатуры к N ортогональным без изменения минимального расстояния при $N > 4$. Обобщение этой идеи лежит в основе следующей процедуры построения оптимального перенасыщенного ансамбля сигнатур [56-57]. Пусть вектора $\mathbf{a}_0^0, \mathbf{a}_1^0, \dots, \mathbf{a}_{N-1}^0$ образуют ортонормированный базис N -мерного пространства, где $N = 4^l, l$ - натуральное число. Используем их в качестве кодов N первичных сигнатур. Организуем перенасыщение дополнительными сигнатурами как результат применения l -этапной (слоистой) процедуры. Кодовыми последовательностями дополнительных сигнатур s -го слоя будут

$$\mathbf{a}_k^s = \frac{1}{2} \sum_{m=0}^3 \mathbf{a}_{4k+m}^{s-1} = \frac{1}{2^s} \sum_{m=0}^{4^s-1} \mathbf{a}_{4^s k+m}^0, \quad k = 0, 1, \dots, \frac{N}{4^s} - 1, \quad s = 1, 2, \dots, l. \quad (7.23)$$

Иными словами, первый слой дополнительных сигнатур формируется путем расщепления

базового множества $\{a_0^0, a_1^0, \dots, a_{N-1}^0\}$ на 4^{l-1} групп, каждая из которых содержит по четыре исходные сигнатуры. Линейная комбинация (7.20) (при $N=4$) этих четырех базисных сигнатур добавляется к своей группе, образуя в общем случае $N+N/4$ сигнатур. На

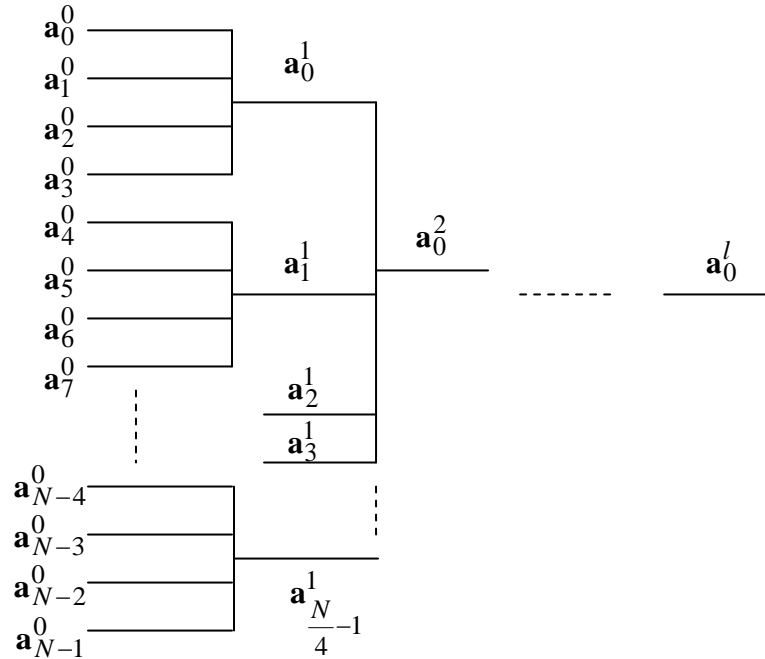


Рис.7.14. Конструирование множества перенасыщенных сигнатур.

втором этапе (уровне) аналогичным образом делим на группы по четыре все дополнительные сигнатуры первого слоя и снова вводим в каждую группу линейную комбинацию (7.20) и т.д. Дерево на рис. 7.14 служит иллюстрацией к данной процедуре. Тогда на первом этапе образуется 4^{l-1} дополнительных сигнатур, на втором этапе – 4^{l-2} , и на произвольном s -м шаге – 4^{l-s} , что в общем случае составит

$$4^{l-1} + 4^{l-2} + \dots + 4 + 1 = \frac{4^l - 1}{3} = \frac{N-1}{3}$$

дополнительных сигнатур, а с учетом первоначальных, отождествленных с нулевым словом, общее число сигнатур составит

$$K = \frac{4N-1}{3} = \left\lceil \frac{4N}{3} \right\rceil.$$

Поскольку нормы всех векторов (7.23) остаются равными единице, то дополнительные сигнатуры сохраняют прежнее значение энергии на бит E_b , что и исходные.

Пусть $\dot{S}_k^s(t)$ и b_k^s являются комплексной огибающей k -й сигнатуры из s -го слоя и пользовательским потоком бит, передаваемых этой сигнатурой. Тогда образование группового сигнала, аналогичного (7.21), приводит к выражению

$$\dot{S}(t; \mathbf{b}) = \sum_{k=0}^{N-1} b_k^0 \dot{S}_k^0(t) + \sum_{k=0}^{\frac{N}{4}-1} b_k^1 \dot{S}_k^1(t) + \dots + \sum_{k=0}^{\frac{N}{4^{l-1}}-1} b_k^{l-1} \dot{S}_k^{l-1}(t) + b_0^l \dot{S}_0^l(t) = \sum_{s=0}^l \sum_{k=0}^{\frac{N}{4^s}-1} b_k^s \dot{S}_k^s(t),$$

которое после подстановки (7.23) обращается в

$$\dot{S}(t; \mathbf{b}) = \sum_{s=0}^l \frac{1}{2^s} \sum_{k=0}^{4^s-1} \sum_{m=0}^{4^s-1-k} b_k^s \dot{S}_{4^s k+m}^0(t).$$

Двойная сумма по k, m в последнем выражении содержит N слагаемых независимо от значения s . Следовательно, возможно преобразование их в одиночную сумму после изменения индекса суммирования как $4^s k + m = n \Rightarrow k = \left\lfloor \frac{n}{4^s} \right\rfloor$, приводящего (при переобозначении $n \rightarrow k$) к виду

$$\dot{S}(t; \mathbf{b}) = \sum_{s=0}^l \frac{1}{2^s} \sum_{k=0}^{N-1} b_{\left\lfloor \frac{k}{4^s} \right\rfloor}^s \dot{S}_k^0(t) = \sum_{k=0}^{N-1} \left(b_k^0 + \frac{1}{2} b_{\left\lfloor \frac{k}{4} \right\rfloor}^1 + \dots + \frac{1}{2^{l-1}} b_{\left\lfloor \frac{k}{4^{l-1}} \right\rfloor}^{l-1} + \frac{1}{2^l} b_0^l \right) \dot{S}_k^0(t).$$

Тогда квадрат расстояния между групповыми сигналами, отвечающими различным битовым образцам, обобщает (7.22) как

$$d^2(\mathbf{b}, \mathbf{b}') = E_b \sum_{k=0}^{N-1} \left(\varepsilon_k^0 + \frac{1}{2} \varepsilon_{\left\lfloor \frac{k}{4} \right\rfloor}^1 + \dots + \frac{1}{2^{l-1}} \varepsilon_{\left\lfloor \frac{k}{4^{l-1}} \right\rfloor}^{l-1} + \frac{1}{2^l} \varepsilon_0^l \right)^2, \quad (7.24)$$

где, как и раньше, $\varepsilon_n^s = 0; \pm 2$ – разность значений бит, передаваемых сигнатурой $\dot{S}_k^s(t)$ в пользовательских битовых образцах \mathbf{b}, \mathbf{b}' . Если все $\varepsilon_m^s, s > 0$ нулевые (биты образцов \mathbf{b}, \mathbf{b}' во всех дополнительных сигнатур одинаковы), тогда, по меньшей мере, одно из ε_k^0 равно ± 2 и $d^2(\mathbf{b}, \mathbf{b}') \geq 4E_b$. Если u – максимальное число слоев, для которых $\varepsilon_m^u = \pm 2, u > 0$, то тогда слагаемые в (7.24), содержащие ε_m^u , могут быть представлены в виде

$$\frac{1}{4^{u-1}} \left| \left(2^u x_0 \pm 2^{u-1} x_1 \pm \dots \pm 2x_{u-1} \right) \pm 1 \right|^2,$$

где $x_s = \varepsilon_n^s / 2 = 0, \pm 1, s = 0, 1, \dots, u-1$. Число, содержащееся в круглых скобках под знаком модуля, всегда четно, так что квадрат модуля никогда не меньше единицы. Поскольку в (7.24) входит ровно 4^u членов, содержащих ε_m^u при любом фиксированном m , приходим к оценке $d(\mathbf{b}, \mathbf{b}') \geq 4^u E_b / 4^{u-1} = 4E_b$, доказывающее, что перенасыщенный ансамбль подобного сорта не уменьшает минимальное расстояние первичного ортогонального множества.

В своей общей форме описанная процедура не гарантирует, что дополнительные кодовые последовательности (7.23), получаемые из первичных бинарных последовательностей, также будут бинарными. Для выполнения указанного требования может быть использована версия алгоритма, предложенная в [57], в которой первичные последовательности формируются как строки l -й кронекеровской степени матрицы Адамара 4-го порядка, содержащей нечетное число $+1$ в любом столбце.

Пример 7.2.1. Построим перенасыщенный ансамбль бинарных сигнатур длины $N = 16 = 4^2$. Согласно только что рассмотренной схеме, к $N = 16$ первичным ортогональным сигнатурам можно добавить пять дополнительных (четыре из первого слоя $s = 1$ и одну из второго слоя $s = 2$), обеспечив общее число пользователей $K = 21$. Для построения только бинарных сигнатур воспользуемся матрицей Адамара вида

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

содержащей одну или три +1 в своих столбцах и образуем ее кронекеровский квадрат в виде

$$\mathbf{H}_{16} = \mathbf{H}_4 \otimes \mathbf{H}_4 = \begin{bmatrix} \mathbf{H}_4 & \mathbf{H}_4 & \mathbf{H}_4 & \mathbf{H}_4 \\ -\mathbf{H}_4 & -\mathbf{H}_4 & \mathbf{H}_4 & \mathbf{H}_4 \\ \mathbf{H}_4 & -\mathbf{H}_4 & \mathbf{H}_4 & -\mathbf{H}_4 \\ \mathbf{H}_4 & -\mathbf{H}_4 & -\mathbf{H}_4 & \mathbf{H}_4 \end{bmatrix}.$$

Первичные сигнатуры представляют собой строки этой матрицы, т.е. в нормированной форме:

$$\begin{pmatrix} \mathbf{a}_0^0, \mathbf{a}_1^0, \mathbf{a}_2^0, \mathbf{a}_3^0, \mathbf{a}_4^0, \mathbf{a}_5^0, \mathbf{a}_6^0, \mathbf{a}_7^0, \mathbf{a}_8^0, \mathbf{a}_9^0, \mathbf{a}_{10}^0, \mathbf{a}_{11}^0, \mathbf{a}_{12}^0, \mathbf{a}_{13}^0, \mathbf{a}_{14}^0, \mathbf{a}_{15}^0 \end{pmatrix}^T =$$

$$= \frac{1}{4} \begin{pmatrix} + & + & + & + & + & + & + & + & + & + & + & + & + & + & + & + \\ - & - & + & + & - & - & + & + & - & - & + & + & - & - & + & + \\ + & - & + & - & + & - & + & - & + & - & + & - & + & - & + & - \\ + & - & - & + & + & - & - & + & + & - & - & + & + & - & - & + \\ - & - & - & - & - & - & - & - & + & + & + & + & + & + & + & + \\ + & + & - & - & + & + & - & - & - & - & + & + & - & - & + & + \\ - & + & - & + & - & + & - & + & - & + & - & + & - & + & - & + \\ - & + & + & - & - & + & + & - & - & + & + & - & - & + & + & - \\ + & + & + & + & - & - & - & - & + & + & + & + & - & - & - & - \\ - & - & + & + & + & + & - & - & - & - & + & + & + & + & - & - \\ + & - & + & - & - & + & - & + & - & - & + & - & - & + & - & + \\ + & + & + & + & - & - & - & - & - & - & + & + & + & + & + & + \\ - & - & + & + & + & + & - & + & - & - & - & - & + & + & - & + \\ + & - & + & - & - & + & - & + & - & + & - & + & - & + & - & + \\ + & - & - & + & - & + & + & - & - & + & + & - & + & - & - & + \end{pmatrix}.$$

Применение правила (7.23) к строкам данной матрицы дает пять дополнительных бинарных сигнатур:

$$\begin{pmatrix} \mathbf{a}_0^1 \\ \mathbf{a}_1^1 \\ \mathbf{a}_2^1 \\ \mathbf{a}_3^1 \\ \mathbf{a}_0^2 \end{pmatrix} = \frac{1}{4} \begin{pmatrix} + & - & + & + & + & - & + & + & + & - & + & + & - & + & + \\ - & + & - & - & - & + & - & - & + & - & + & + & - & + & + \\ + & - & + & + & - & + & - & - & + & - & + & + & - & + & - \\ + & - & + & + & - & + & - & - & - & + & - & - & + & - & + \\ + & - & + & + & - & + & - & - & + & - & + & + & - & + & + \end{pmatrix}.$$

Читателю самому предоставляется возможность проверить свойства этого перенасыщенного ансамбля в отношении минимума расстояния. \square

Вспомним, что критерий минимума расстояния полностью адекватен только случаю многопользовательского приема. С этой точки зрения нас не волновал вопрос сложности реализации многопользовательского приемника. Для случая системы без насыщения ($K \leq N$) данная проблема не критична, поскольку, учитывая оптимальность ортогонального ансамбля, многопользовательский прием вырождается в вариант однопользовательского (см. параграф 4.1). С другой стороны, при анализе перенасыщенной системы очень важно учитывать возможности упрощения многопользовательского алгоритма за счет соответствующего построения сигнатур. Один из путей, реализующий данный подход, вновь заключается в идее расщепления общего N -мерного сигнального пространства на ортогональные подпространства меньшей размерности n . Однако, в противоположность тому, что обсуждалось ранее, в дальнейшем каждое из подпространств перенасыщается автономно, обеспечивая $n + n_{ov}$ сигнатурами, так что все сигнатуры из различных

подпространств остаются ортогональными. Обоснование подобным действиям состоит в расщеплении всего многопользовательского алгоритма на N/n параллельных, каждый из которых осуществляется в n -мерном подпространстве независимо от других. При рациональном выборе n эти частичные алгоритмы достаточно просты и делают общую структуру приемника технологически возможной. Общее число пользователей, достижимое в подобной системе, составляет величину $\frac{N}{n}(n+n_{ov}) = N(1+\frac{n_{ov}}{n})$. Задача оптимизации подобного типа ансамбля в известном смысле не тривиальна и не заключается в простом добавлении n_{ov} дополнительных сигнатур к n исходным ортогональным сигналам. За деталями заинтересованный читатель может обратиться к [58-59]. Альтернативным путем является построение ансамблей сигнатур, допускающих применение многопользовательских алгоритмов в виде различных вычислительно эффективных итерационных формах [60-61].

7.2.3. Последовательности, удовлетворяющие границе Велча.

Обратимся теперь к другому сценарию, в котором а priori жесткое ограничение на сложность приемника допускает только его простейший вариант, т.е. применение однопользовательского или обычного алгоритма приема. В этом случае решение \hat{b}_k о текущем символе данных b_k k -го пользователя определяется только корреляцией (7.8), как если бы на входе приемника отсутствовали другие помехи, а присутствовал только АБГШ. Без потери общности можно считать, что текущий символ принимается на интервале $[0, T]$ и полагать задержку τ_k и фазу φ_k в (7.8), равными нулю:

$$\dot{z}_k = \int_0^T \dot{Y}(t) \dot{S}_k^*(t) dt. \quad (7.25)$$

Когда все сигнатуры идеально синхронизированы, а их число K не превосходит N , ортогональные сигнатуры снова представляют собой наилучший вариант, поскольку в этом случае однопользовательский алгоритм идентичен многопользовательскому (максимально правдоподобному) алгоритму. Очевидно, что в этом случае отсутствуют помехи множественного доступа (МАИ), так что игнорирование всех сигналов других пользователей не нарушает оптимальности приемника. В противоположность рассмотренному варианту, перенасыщенная ($K > N$) система представляет особый интерес, поскольку в данном случае не все сигнатуры ортогональны и, значит, неизбежно возникновение МАИ. Возвращаясь к (7.11), представим наблюдаемую комплексную огибающую в виде

$$\dot{Y}(t) = \dot{S}(t; \mathbf{b}') + \dot{N}(t) = \sum_{l=1}^K b'_l \dot{S}_l(t) + \dot{N}(t),$$

где $\dot{N}(t)$ – комплексная огибающая шума, а обозначение $\mathbf{b}' = (b'_1, b'_2, \dots, b'_K)$ снова (как и в (4.8)) символизирует истинный (т.е. не известный на приемной стороне) образец данных, передаваемых K пользователями, с целью отличия его от образца $\mathbf{b} = (b_1, b_2, \dots, b_K)$, гипотетически рассматриваемого в ходе принятия решения. После подстановки упомянутых величин в (7.25) получаем

$$\dot{z}(b_k) = 2b'_k E + 2E \sum_{\substack{l=1 \\ l \neq k}}^K b'_l \dot{\rho}_{lk} + \int_0^T \dot{N}(t) \dot{S}_k^*(t) dt, \quad (7.26)$$

где $E = \frac{1}{2} \int_0^T |\dot{S}_k(t)|^2 dt$ полагается одинаковой для всех пользовательских сигнатур энергией,

приходящейся на один передаваемый символ, а $\dot{\rho}_{lk} = \dot{\rho}_{kl}^*$, как и всегда, коэффициент кор-

реляции комплексных огибающих l -й и k -й сигнатур. Второе слагаемое в (7.26) представляет собой помеху множественного доступа, т.е. взаимную помеху, создаваемую чужими сигналами на выходе приемника, «настроенного» на k -й пользовательский сигнал. Каждое слагаемое $b_l' \dot{\rho}_{lk}$ в сумме по l (т.е. вклад l -го пользовательского сигнала в общую помеху МАИ) случайно вследствие случайности пользовательских символов данных b_l' . Для произвольной ФМ данных $|b_l'|=1$, и средняя мощность (дисперсия) каждого вклада в помеху МАИ составляет $4E^2 |\dot{\rho}_{lk}|^2$. Естественно, все пользователи передают свои данные независимо, так что полная средняя мощность (дисперсия) помех МАИ P_{lk} на выходе k -го приемника представляет собой сумму по l мощностей индивидуальных вкладов

$$P_{lk} = 4E^2 \sum_{\substack{l=1 \\ l \neq k}}^K |\dot{\rho}_{kl}|^2.$$

Поскольку эта величина оценивает мощность помехи МАИ только для приемника k -го пользователя, то для охвата всей системы можно просуммировать по k , придя к результату

$$P_l = \sum_{k=1}^K P_{lk} = 4E^2 \sum_{k=1}^K \sum_{\substack{l=1 \\ l \neq k}}^K |\dot{\rho}_{kl}|^2. \quad (7.27)$$

Теперь нетрудно увидеть, что при постулировании однопользовательского приема совершенно адекватным критерием оптимизации синхронизированных сигнатур является минимизация мощности помех множественного доступа или, что эквивалентно, суммы квадратов корреляций в вышеприведенном соотношении. Конечно, для случая $K \leq N$ ортогональное множество сигнатур снова не создает помех МАИ, т.е. обращает эту сумму в нуль, так что только перенасыщенные ансамбли служат предметом интереса.

Только что введенный критерий, как правило, используется в литературе под названием минимума *полного квадрата корреляции* (*total squared correlation (TSC)*)

$$TSC = \sum_{k=1}^K \sum_{l=1}^K |\dot{\rho}_{kl}|^2 = \min, \quad (7.28)$$

который не отличается от исходного выражения, поскольку рассматриваемая сумма больше, содержащейся в (7.27), на постоянную величину K , т.к. ($\dot{\rho}_{kk} = 1$).

Существует фундаментальная нижняя граница величины TSC, известная как граница Велча [62]. Установим ее, выразив первоначально коэффициенты корреляции в терминах элементов $a_{k,i}$ кодовых последовательностей сигнатур. Полагая все вектора $\mathbf{a}_k = (a_{k,0}, a_{k,1}, \dots, a_{k,N-1})$ кодовых последовательностей нормированными, так что $\|\mathbf{a}_k\| = N$, соотношение (7.19) дает

$$\dot{\rho}_{kl} = \frac{(\dot{\mathbf{S}}_k, \dot{\mathbf{S}}_l)}{2E} = \frac{(\dot{\mathbf{S}}_k, \dot{\mathbf{S}}_l)}{2\|\mathbf{a}_k\|\|\mathbf{a}_l\|E_0} = \frac{(\mathbf{a}_k, \mathbf{a}_l)}{N} = \frac{1}{N} \sum_{i=0}^{N-1} a_{k,i} a_{l,i}^*.$$

Подстановка последнего соотношения в определение (7.28) TSC приводит к следующему результату

$$\begin{aligned} TSC &= \frac{1}{N^2} \sum_{k=1}^K \sum_{l=1}^K \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} a_{k,i} a_{l,i}^* a_{k,j} a_{l,j}^* = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \sum_{k=1}^K a_{k,i} a_{k,j}^* \sum_{l=1}^K a_{l,i} a_{l,j}^* = \\ &= \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \left| \sum_{k=1}^K a_{k,i} a_{k,j}^* \right|^2. \end{aligned}$$

Поскольку слагаемые по i, j все неотрицательны, то пренебрежение теми из них, которые

отвечают различным i, j , не может увеличить значения суммы, так что

$$TSC \geq \frac{1}{N^2} \sum_{i=0}^{N-1} \left(\sum_{k=1}^K |a_{k,i}|^2 \right)^2. \quad (7.29)$$

Для получения окончательного результата можно воспользоваться неравенством Шварца, однако этот шаг становится необязательным для наиболее интересного случая ФМ сигнатур. Для любого алфавита с ФМ $|a_{k,i}|=1$, что завершает вывод границы Велча:

$$TSC \geq \frac{1}{N^2} \sum_{i=0}^{N-1} \left(\sum_{k=1}^K 1 \right)^2 = \frac{K^2}{N}.$$

В отсутствие перенасыщения ($K \leq N$) непосредственным следствием определения (7.28) является строгая граница, основанная на том факте, что при ортогональных сигнатурах все слагаемые в (7.28) с разными k, l исчезают, и значение TSC достигает своего минимума, равного K . Сочетание полученных результатов приводит к следующей общей форме границы Велча

$$TSC \geq \begin{cases} K, & K \leq N, \\ \frac{K^2}{N}, & K > N. \end{cases} \quad (7.30)$$

Очевидно, что множество последовательностей, достигающих (7.30) (*последовательностей, удовлетворяющих границе Велча (Welch-bound sequences)*), являются наилучшими по критерию общей помехи MAI в случае однопользовательского приемника. Однако на самом деле значимость этих множеств превышает указанное качество, поскольку последовательности, удовлетворяющие границе Велча, максимизируют шенноновскую пропускную способность CDMA каналов с АБГШ и гауссовским входом, причем последнее ограничение теряет свое значение при достаточно малом отношении сигнал-шум на входе приемника. Детали доказательства этого замечательного свойства читатель может найти в [63].

Поскольку величина TSC включает в себя K квадратов корреляций векторов самих с собой, каждая из которых равна единице, то разность $TSC - K$ охватывает только нежелательные корреляции между несовпадающими векторами, значения которых желательно иметь минимально возможными. В состав TSC входит $K(K-1)$ подобных пар векторов, так что среднеквадратическое значение корреляции $\overline{\rho^2}$ на одну пару составит величину

$$\overline{\rho^2} = \frac{TSC - K}{K(K-1)},$$

которая совместно с (7.30) дает следующую нижнюю границу этого параметра

$$\overline{\rho^2} \geq \begin{cases} 0, & K \leq N, \\ \frac{K-N}{N(K-1)}, & K \geq N. \end{cases} \quad (7.31)$$

Из способа получения границы (7.30) можно понять, как построить ансамбль, удовлетворяющий границе Велча. Очевидно, что следует обсудить только нетривиальный случай перенасыщения, поскольку способы формирования ортогональных последовательностей были обсуждены ранее. Во-первых, равенство в (7.29) является достаточным (и, конечно, необходимым) условием достижения равенства в (7.30), или, рассматривая соотношение, предшествующее (7.29), последовательности, для которых

$$\sum_{k=1}^K a_{k,i} a_{k,j}^* = 0, \quad i \neq j, \quad (7.32)$$

являются последовательностями, удовлетворяющими границе Велча. Предположим, что

все вектора $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_K$ кодовых последовательностей сигнатур записаны в виде столбцов $N \times K$ матрицы сигнатур \mathbf{A} :

$$\mathbf{A} = [\mathbf{a}_1 \ \mathbf{a}_2 \ \dots \ \mathbf{a}_K] = \begin{bmatrix} a_{1,0} & a_{2,0} & \dots & a_{K,0} \\ a_{1,1} & a_{2,1} & \dots & a_{K,1} \\ \dots & \dots & \dots & \dots \\ a_{1,N-1} & a_{2,N-1} & \dots & a_{K,N-1} \end{bmatrix},$$

тогда (7.32) означает ничто иное, как ортогональность строк матрицы \mathbf{A} . Следовательно, для построения перенасыщенного ($K > N$) ансамбля последовательностей, удовлетворяющих границе Велча, требуется только построить $N \times K$ матрицу \mathbf{A} с ортогональными строками. Поскольку размерность строк подобной матрицы превышает их число, то не существует принципиальных препятствий к их существованию. Тогда желаемые последовательности есть просто столбцы матрица \mathbf{A} .

Теперь можно оценить пороговое отношение сигнал-помеха SIR (т.е. без учета шума) для перенасыщенного ансамбля последовательностей, удовлетворяющих границе Велча. Общая мощность помех МАИ P_I может быть найдена из (7.27)-(7.28) как $P_I = 4E^2(TSC - K)$. Поскольку эта величина была получена путем суммирования помех МАИ по всем K однопользовательским приемникам, то средняя выходная мощность МАИ, приходящаяся на один приемник, будет $\overline{P_{Ik}} = P_I / K$. Полезный эффект на выходе k -го приемника (т.е. обусловленный k -й сигнатурой), определяемый первым членом в соотношении (7.26), составляет $4E^2$ (при условии применения фазовой модуляции), так что согласно (7.30) пороговое отношение по мощности SIR относительно средней мощности МАИ составит

$$q_I^2 = \frac{4E^2}{\overline{P_{Ik}}} = \frac{K}{TSC - K} = \frac{N}{K - N}. \quad (7.33)$$

Пример 7.2.2. Построим бинарный ансамбль из $K = 16$ последовательностей длины $N = 14$, удовлетворяющий границе Велча. Для решения данной задачи воспользуемся матрицей \mathbf{H}_{16} из примера 7.2.2 и отбросим в ней две произвольные (например, последние) строки. Полученная таким образом матрица \mathbf{A} в точности соответствует необходимой, а ее 16 столбцов представляют собой сигнатуры длины 14, удовлетворяющие границе Велча. Для найденного таким путем ансамбля значение TSC в точности равняется величине, определяемой соотношением (7.30): $TSC = \frac{K^2}{N} = \frac{256}{14}$. Пороговое отношение SIR, оцениваемое относительно средней мощности помех МАИ на один приемник, составит согласно (7.33) величину $q_I^2 = N / (K - N) = 7$. \square

Если ограничиться алфавитом кодовых последовательностей сигнатур, определяемым только фазовой модуляцией, то предложенный выше алгоритм конструирования множеств, удовлетворяющих границе Велча, является универсальным. Например, в качестве строк матрицы \mathbf{A} всегда могут быть взяты K циклически сдвинутых копий последовательности Чу длины K . Как было показано в параграфе 6.11.2, коды Чу существуют при любой длине, и все их различные циклические сдвиги являются ортогональными. С другой стороны, если сигнатуры должны быть бинарными (т.е. $a_{k,i} = \pm 1$), то ортогональность всех N строк матрицы \mathbf{A} при $N > 2$ возможна только для K кратного четырем (см. задачу 7.14). Данное утверждение влечет, что для бинарных сигнатур с $K \neq 0 \pmod{4}$ граница Велча (7.30) не является точной, и должны существовать более точные нижние границы. Вывод их заинтересованный читатель может найти в [64, 65] (а также в задаче 7.17).

7.3. Подходы к синтезу ансамблей сигнатур для асинхронного CDMA с прямым расширением спектра.

Распространим задачу синтеза сигнатур на случай асинхронного CDMA с прямым расширением спектра, в котором временная задержка и сдвиг начальной фазы между отдельными пользовательскими сигналами случайны. При предположении об использовании однопользовательского приемника решение о значении текущего символа k -го пользователя снова принимается на основе величины корреляции (7.25). Однако теперь точное выравнивание границ символов данных и чипов различных пользователей не поддерживается вследствие произвольных взаимных временных сдвигов пользовательских сигналов. Предположим, что исследуется приемник данных k -го пользователя и τ_l – задержка l -го сигнала относительно k -го сигнала. С целью концентрации внимания только на задаче синтеза кодов сигнатур допустим, что границы чипов всех K сигнатур синхронизированы, т.е. взаимные задержки кратны Δ : $\tau_l = n_l \Delta$, где n_l – целое число, такое что $0 \leq n_l < N$. Тогда ситуация хорошо поясняется рис. 7.15 (для случая $k=1$), подчеркивающим, что в асинхронном варианте CDMA в отличие от синхронного (см. рис. 7.13) символы данных других пользователей могут изменяться в течение приема текущего символа k -го пользователя. Еще одним фактором, затрудняющим синтез множества асинхронных сигнатур, является необходимость различения каждой сигнатуры относительно всех возможных сдвинутых копий других сигнатур, что излишне при синхронном CDMA.

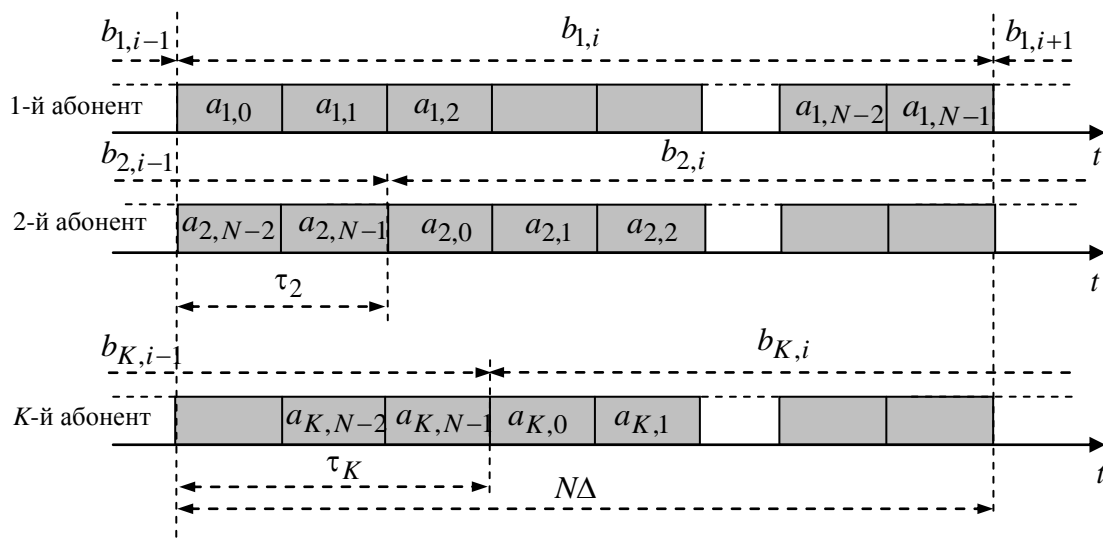


Рис.7.15. Поток данных и сигнатуры в асинхронном CDMA.

Предположим первоначально, что во время приема символа данных k -го пользователя не происходит изменения символов данных всех других пользователей, т.е. $b_{l,i-1} = b_l$, $l = 2, 3, \dots, K$. Тогда отличие ситуации в синхронном варианте от рассматриваемой состоит только во взаимной рассогласованности во времени сигнатур. Начнем рассмотрение с предположения, что период сигнатуры L совпадает с выигрышем от обработки N , который равен числу чипов, приходящихся на длительность одного символа данных, или, что эквивалентно, числу чипов, интегрируемых коррелятором. Если не накладывается никаких ограничений на диапазон возможных взаимных задержек, то l -я сигнатура может быть представлена любой из своих N циклически сдвинутых копий, так что имеется $N(K-1)$ различных N -мерных векторов, каждый из которых является потенциальным источником помех МАИ в k -м приемнике. Если канал подвержен эффектам многолучевости, то любая собственная копия k -го сигнала также может стать помехой

k – му приемнику. Предположим, что может существовать до $N-1$ подобной копии, т.е. задержка многолучевости распространяется на диапазон вплоть до периода сигнатуры. Другой причиной для включения собственных циклических копий в исследуемое множество векторов служит желание иметь низкий уровень боковых лепестков автокорреляции, который важен при поиске сигнала (см. параграф 8.2). При таком добавлении имеем KN векторов, корреляция которых должна быть минимально возможной.

Хорошим инструментом для оценки нижнего предела среднего квадрата корреляции $\overline{\rho^2}$ этих KN векторов снова является граница Велча. Для ее использования достаточно заменить в (7.31) K на KN . Поскольку $KN > K$, то для любого $K \geq 2$ имеем

$$\overline{\rho^2} \geq \frac{K-1}{KN-1}. \quad (7.34)$$

Это неравенство определяет фундаментальный нижний предел, ниже которого средний квадрат корреляции между всеми циклическими копиями всех K сигнатур (включая собственные копии каждой сигнатуры) опуститься никогда не может. При числе пользователей около десяти или более эта версия границы Велча становится особенно простой:

$$\overline{\rho^2} \geq \frac{1}{N}, \quad K \gg 1. \quad (7.35)$$

Предположим теперь, что период сигнатуры в числе L чипов охватывает несколько символов данных $L > N$ и что, как и ранее, в течение k -го текущего символа данных не изменяется информация никакого пользователя¹. Для того, чтобы не определять некоторый чип сигнатуры в качестве первоначального, произведем усреднение нашей оценки по всем начальным чипам, означающее, что вычисление корреляции в приемнике может начинаться с любого из них. Пусть опять диапазон возможных задержек простирается вплоть до периода сигнатуры. Поскольку число чипов на символ данных (интервал интегрирования) остается равным N , то, как и ранее, имеем дело с N -мерными векторами, однако число векторов, корреляции которых контролируется, составляет теперь KL вместо KN , так что из (7.31) вытекает граница

$$\overline{\rho^2} \geq \frac{KL-N}{N(KL-1)}, \quad (7.36)$$

которая при $K \gg 1$ снова обращается в (7.35). Последний результат позволяет продемонстрировать, что модуляция данных никоим образом не может опустить полученные границы. Действительно, любая промодулированная данными сигнатура может трактоваться как новая последовательность некоторого (возможно очень большого) периода L_k . Тогда все промодулированные сигнатуры будут иметь общий период L , равный, по крайней мере, общему кратному всех L_k , и средний квадрат корреляции будет ограничен снизу соотношением (7.36), снова означающим действительность (7.35) для случая многих пользователей.

Только что сделанные заключения устанавливают критерий синтеза множества асинхронных сигнатур: ансамбль из большого количества сигнатур может считаться подходящим, если его средний квадрат корреляции близок границе (7.35). Продемонстрируем, что ансамбли случайных сигнатур удовлетворяют этой границе. Пусть все сигнатуры составляются независимо друг от друга путем случайного независимого выбора элементов каждой из них. Вся процедура может трактоваться как вытаскивание шаров из урны. Остановимся на алфавите M -ичной ФМ, и будем трактовать его как некоторую урну с M

¹ Сохраняя обозначение N для выигрыша от обработки, т.е. число чипов, приходящихся на символ данных, с данного места будем обозначать через L период сигнатуры всякий раз, когда они различны.

различными шарами (кодowymi символами). Вытащим K раз по одному шару, каждый раз запоминая результат и возвращая шар в урну. Данная операция определит первые символы K сигнатур. Каждый следующий символ всех сигнатур генерируется аналогичным образом. Поскольку все M -ичные символы в данной схеме равновероятны, равномерно распределены на плоскости (см. рис. 2.6, с) и независимы друг от друга, то имеем следующие математические ожидания

$$\overline{a_{k,i}} = 0, \quad \overline{a_{k,i} a_{l,j}^*} = \begin{cases} 1, & k=l \text{ и } i=j, \\ 0, & \text{в противном случае} \end{cases} = \delta_{kl} \delta_{ij}, \quad (7.37)$$

где второе соотношение вытекает из того факта, что математическое ожидание произведения независимых величин равно произведению их математических ожиданий. Воспользуемся этими соотношениями при оценке среднего квадрата корреляции сигнатур на интервале интегрирования в N чипов:

$$\overline{|\dot{z}_{kl}(m)|^2} = \overline{\left| \sum_{i=0}^{N-1} a_{k,i} a_{l,i-m}^* \right|^2}. \quad (7.38)$$

Фактически выражение (7.38) представляет собой среднюю мощность помех множественного доступа ($k \neq l$) или среднюю мощность помех многолучевости ($k=l$), создаваемую l -й сигнатурой, задержанной на m чипов, на выходе k -го коррелятора. Возведение в квадрат модуля и изменение очередности суммирования и усреднения (математическое ожидание суммы равняется сумме математических ожиданий) дает

$$\overline{|\dot{z}_{kl}(m)|^2} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \overline{a_{k,i} a_{l,i-m}^* a_{k,j} a_{l,j-m}^*}. \quad (7.39)$$

Для оценки эффекта помех множественного доступа положим $k \neq l$ и разделим слагаемые на произведение независимых случайных переменных:

$$\overline{|\dot{z}_{kl}(m)|^2} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \overline{a_{k,i} a_{k,j}^*} \cdot \overline{a_{l,i-m} a_{l,j-m}^*}.$$

Теперь применение (7.37) в данном соотношении оставляет в сумме, стоящей в правой части выражения, только члены с $i=j$, приводя к следующему результату $\overline{|\dot{z}_{kl}(m)|^2} = N$. Для оценки эффекта помех многолучевости положим $k=l$, $m \neq 0$, что приводит (7.39) к виду

$$\overline{|\dot{z}_{kk}(m)|^2} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \overline{a_{k,i} a_{k,i-m}^* a_{k,j} a_{k,j-m}^*}. \quad (7.40)$$

Среди членов последнего соотношения, имеющих различные индексы i и j , $a_{k,i}$ является независимым как от $a_{k,i-m}$ (поскольку $m \neq 0$), так и от $a_{k,j}$ (поскольку $i \neq j$). Аналогично, $a_{k,j-m}$ независимо как от $a_{k,j}$, так и от $a_{k,i-m}$. Следовательно

$$\overline{a_{k,i} a_{k,j-m}^* a_{k,i-m} a_{k,j}^*} = \overline{a_{k,i} a_{k,j-m}^*} \cdot \overline{a_{k,i-m} a_{k,j}^*}, \quad i \neq j.$$

Вследствие независимости различных символов одной и той же последовательности $\overline{a_{k,i} a_{k,j-m}^*} = \overline{a_{k,i}} \cdot \overline{a_{k,j-m}^*} = 0$ всякий раз, когда $i \neq j-m$, и $\overline{a_{k,i-m} a_{k,j}^*} = \overline{a_{k,i-m}} \cdot \overline{a_{k,j}^*} = 0$, если $i-m \neq j$. Следовательно, слагаемые в сумме (7.40) с различными индексами i, j могут оказаться ненулевыми, если только выполняются оба соотношения: $i = j-m$ и $i-m = j$, что невозможно для любого ненулевого m . Отсюда следует, что в (7.40) только

слагаемые с одинаковыми значениями i, j дают ненулевой вклад, а, значит

$$\overline{|\dot{z}_{kk}(m \neq 0)|^2} = \sum_{i=0}^{N-1} |a_{k,i}|^2 |a_{k,i-m}|^2 = N.$$

Полезный эффект, т.е. мощность, обусловленная несдвинутой k -й сигнатурой на выходе k -го приемника, составит величину

$$|\dot{z}_{kk}(0)|^2 = \sum_{i=0}^{N-1} |a_{k,i}|^2 \sum_{j=0}^{N-1} |a_{k,j}|^2 = N^2.$$

Тогда нормированные нежелательные эффекты, обусловленные либо помехами множественного доступа, либо многолучевыми помехами (нежелательный квадрат корреляции), будут выражаться как

$$\frac{\overline{|\dot{z}_{kl}(m)|^2}}{\overline{|\dot{z}_{kk}(0)|^2}} = \frac{1}{N}, \quad \frac{\overline{|\dot{z}_{kk}(m \neq 0)|^2}}{\overline{|\dot{z}_{kk}(0)|^2}} = \frac{1}{N}. \quad (7.40)$$

Теперь очевидно, что все нежелательные квадраты корреляции в рассматриваемом ансамбле достигают нижней границы (7.35), т.е. множества случайных сигнатур являются оптимальными при числе пользователей порядка десять и более. Чрезвычайно важно подчеркнуть, что модуляция данными случайных последовательностей, удовлетворяющих (7.37) (т.е. умножение последовательностей на независимые от них символы данных), не нарушит (7.37) (см. задачу 7.20). Таким образом, наличие или отсутствие модуляции не затронет всего вышеприведенного вывода, также как и окончательного результата (7.40) и заключения об оптимальности множества.

Представляется, что соотношения (7.37) дают безошибочную инструкцию по синтезу ансамбля сигнатур. Однако на самом деле сигнатуры не могут быть случайными, поскольку приемник а priori должен знать закон модуляции сигнатуры с тем, чтобы сформировать необходимый опорный сигнал коррелятора. Для реализации свойств случайности (7.37) с помощью детерминированных правил кодирования необходимы так называемые *псевдослучайные* последовательности.

Возьмем детерминированную ФМ сигнатуру периода L и будем трактовать ее как одну из равновероятных реализаций стационарной эргодической случайной последовательности $\{a_{k,i}\}$ (дискретный по времени случайный процесс) [14, 66]. Другими реализациями могут быть все циклические сдвиги исходной последовательности. Тогда благодаря свойству эргодичности каждая реализация исчерпывающе представляет весь случайный процесс, и статистическое усреднение $\{a_{k,i}\}$ по всем реализациям эквивалентно усреднению по времени, т.е. оценивание математического ожидания $\overline{a_{k,i}}$ и корреляционного момента $\overline{a_{k,i} a_{k,i-m}^*}$ через постоянную составляющую и периодическую АКФ детерминированной сигнатуры соответственно:

$$\overline{a_{k,i}} = \frac{1}{L} \sum_{i=0}^{L-1} a_{k,i} = \frac{\tilde{a}_{k,0}}{L}, \quad \overline{a_{k,i} a_{k,i-m}^*} = \frac{1}{L} \sum_{i=0}^{L-1} a_{k,i} a_{k,i-m}^* = \rho_{kk}(m). \quad (7.41)$$

Аналогичным образом, трактуя две детерминированные сигнатуры, как реализации двух совместно эргодических случайных последовательностей $\{a_{k,i}\}$ и $\{a_{l,i}\}$, получаем равенство между корреляционным моментом двух случайных последовательностей и взаимной корреляционной функцией двух детерминированных сигнатур:

$$\overline{a_{k,i} a_{l,i-m}^*} = \frac{1}{L} \sum_{i=0}^{L-1} a_{k,i} a_{l,i-m}^* = \rho_{kl}(m). \quad (7.42)$$

Из сравнения (7.41)–(7.42) с (7.37) вытекает критерий псевдослучайности: в качестве сигнатур в асинхронном варианте CDMA с прямым расширением должен использоваться такой ансамбль детерминированных последовательностей, все представители которого в идеале обладают нулевой постоянной составляющей, идеальной периодической АКФ и нулевой периодической ВКФ:

$$\tilde{a}_{k,0} = 0; \quad \rho_{kk}(m) = 0, \quad m \neq 0 \bmod N; \quad \rho_{kl}(m) = 0, \quad k, l = 1, 2, \dots, K. \quad (7.43)$$

В случае отсутствия ограничений на взаимный временной сдвиг (возможно любое m из диапазона $0, 1, \dots, L-1$) последние требования явно противоречат друг другу, делая ансамбли этого сорта гипотетическими для любого конечного значения L . Действительно (см. также задачу 7.21), требования идеальности АКФ и нулевой ВКФ означают ничто иное, как нулевой уровень корреляции между всеми циклическими сдвигами K последовательностей периода L , т.е. нулевого среднего значения нежелательных квадратов корреляций $\overline{\rho^2}$. Как следует из (7.34)–(7.35), данное требование не выполнимо при $K \geq 2$, и, в частности, при многих пользователях $\overline{\rho^2}$ не может опуститься ниже $1/L$.

Заклучение, к которому только что пришли, объясняет причину многих усилий, направленных на поиски ансамблей, характеристики которых с ростом длины L приближаются к характеристикам упомянутых выше гипотетических ансамблей. Широко распространенным критерием подобного приближения является минимаксный критерий, ориентирующий синтез ансамбля на минимизацию максимальное значение на множестве всех нежелательных корреляций. Определим *корреляционный пик* ρ_{\max} , как наибольшее из двух величин: максимума среди всех боковых лепестков автокорреляций последовательностей ρ_{\max}^a и максимума среди значений взаимных корреляций всех пар последовательностей ρ_{\max}^c :

$$\rho_{\max} = \max\{\rho_{\max}^a, \rho_{\max}^c\}, \quad \rho_{\max}^a = \max_{k, m \neq 0} |\rho_{p, kk}(m)|, \quad \rho_{\max}^c = \max_{\substack{k, l, m \\ k \neq l}} |\rho_{p, kl}(m)|. \quad (7.44)$$

Естественно, что для идеального гипотетического ансамбля ρ_{\max} , как и $\overline{\rho^2}$ равны нулю, а для любого реального ансамбля ρ_{\max} может служить адекватной мерой его близости к идеальному.

Поскольку максимальное значение любой переменной никогда не может быть меньше ее среднего значения, то $\rho_{\max}^2 \geq \overline{\rho^2}$, что позволяет распространить границы Велча (7.34)–(7.35) на величину корреляционного пика:

$$\rho_{\max}^2 \geq \frac{K-1}{KL-1} \approx \frac{1}{L}, \quad (7.45)$$

где снова последнее приближение отвечает случаю $K \gg 1$. При дополнительных ограничениях на алфавит фазовой модуляции вышеприведенная граница может оказаться достаточно слабой, особенно при числе последовательностей близким к L . В частности, для достаточно больших ансамблей бинарных $\{\pm 1\}$ последовательностей действует граница Сидельникова [67-67]

$$\rho_{\max}^2 \geq \frac{2}{L}, \quad K \geq \frac{L}{2}. \quad (7.46)$$

Ансамбли со значением ρ_{\max} , достигающие предела, предсказываемого нижними границами, являются, конечно, оптимальными по критерию корреляционного пика, и иногда называются *минимаксными*. Некоторые из них рассматриваются в параграфе 7.5.

7.4. Сигнатуры для асинхронного варианта CDMA, получаемые временным сдвигом.

Во многих реальных ситуациях взаимные временные сдвиги асинхронных сигнатур могут изменяться только в пределах ограниченного диапазона. Конечность канальной задержки распространения с одной стороны и геометрия системы с другой являются наиболее типичными факторами, устанавливающими эти ограничения. Для их детализации обратимся к обратному каналу сотовой системы мобильной связи. Локальное время (тактовая частота) активной МС синхронизировано с принятым сигналом БС и характеризуется задержкой τ относительно времени БС, определяемой расстоянием D между БС и МС, как $\tau_1 = D/c$, где c – скорость распространения света. Поскольку сигнал, передаваемый конкретной МС, достигает приемника БС с той же самой задержкой, то общий временной сдвиг сигнала, дошедшего до БС, относительно времени БС составляет $\tau_2 = 2\tau_1 = 2D/c$. Пусть D_{\max} – максимальное расстояние, при котором интенсивность принимаемого сигнала достаточна для приема БС. Сильное ослабление при распространении (см. параграф 4.6) допускает игнорирование сигналов, которые пришли с расстояний, значительно превышающих радиус соты D_c . Последнее позволяет использовать следующую грубую оценку для максимального расстояния $D_{\max} \approx D_c$. Тогда максимальное значение τ_2 составляет $2D_c/c$, и сигналы мобильных станций, расположенных от БС на расстоянии от нуля до D_c , достигают ее с задержками, принадлежащими диапазону $[0, 2D_c/c]$. Кроме того, существуют многолучевые копии сигналов, так что общая ширина τ_{\max} окна, охватывающего задержки всех многолучевых сигналов, увеличивается на величину канальной задержки распространения τ_{ds} : $\tau_{\max} = 2D_c/c + \tau_{ds}$, где τ_{ds} может быть максимизировано по всем возможным местоположениям МС. Рис. 7.16 служит иллюстрацией к приведенному обсуждению. Сигнал некоторой конкретной МС может оказаться как опережающим, так и задержанным по отношению к некоторому другому сигналу МС, а все многолучевые копии сигнала любой МС потенциально используются приемником БС (RAKE обработка, см. параграф 3.7). Следовательно, полный диапазон возможных взаимных временных сдвигов между любыми многолучевыми копиями любых сигнатур оказывается равным $[-\tau_{\max}, \tau_{\max}]$, где $\tau_{\max} = 2D_c/c + \tau_{ds}$.

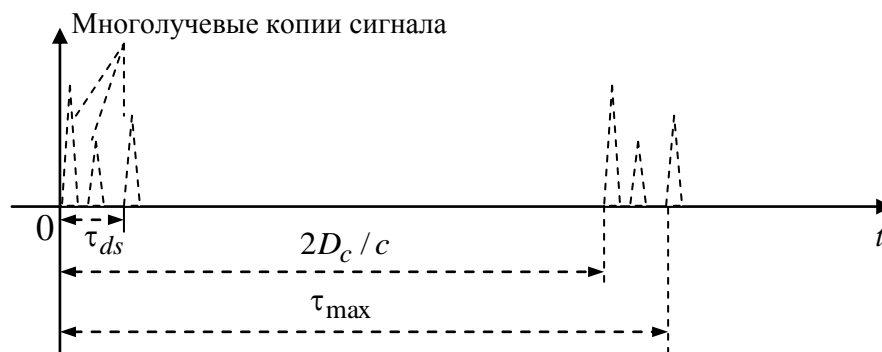


Рис. 7.16. Изменение времени достижения БС сигналом с МС.

Очевидно, что в подобных обстоятельствах следует заботиться о соблюдении второго и третьего условий (7.43) только в пределах диапазона действительно возможных значений m . Выразим величину τ_{\max} в округленном до целого числа m_{\max} чипов:

$m_{\max} = \lceil \tau_{\max} / \Delta \rceil$. Тогда диапазон значений m , где следует выполнять требования (7.43), определяется как $[-m_{\max}, m_{\max}]$. Возьмем теперь последовательность $\{a_{1,i}\}$ периода $L \geq K(m_{\max} + 1)$ и используем в качестве K сигнатур ее циклически сдвинутые одна относительно другой копии на величину $m_{\max} + 1$ позиций:

$$a_{k,i} = a_{1,i-(k-1)(m_{\max}+1)}, \quad k = 1, 2, \dots, K; \quad i = \dots, -1, 0, 1, \dots,$$

как показано на рис. 7.17. Очевидно, что все корреляции между образованными таким образом сигнатурами будут выражены в терминах АКФ $\rho_{11}(m)$ исходной последовательности $\{a_{1,i}\}$. Оценка ВКФ k -й и l -й сигнатур приводит к следующему результату

$$\rho_{kl}(m) = \frac{1}{L} \sum_{i=0}^{N-1} a_{k,i} a_{l,i-m}^* = \frac{1}{L} \sum_{i=0}^{N-1} a_{1,i-(k-1)(m_{\max}+1)} a_{1,i-(l-1)(m_{\max}+1)-m}^*,$$

или

$$\rho_{kl}(m) = \rho_{11}[(k-l)(m_{\max} + 1) + m]. \quad (7.47)$$

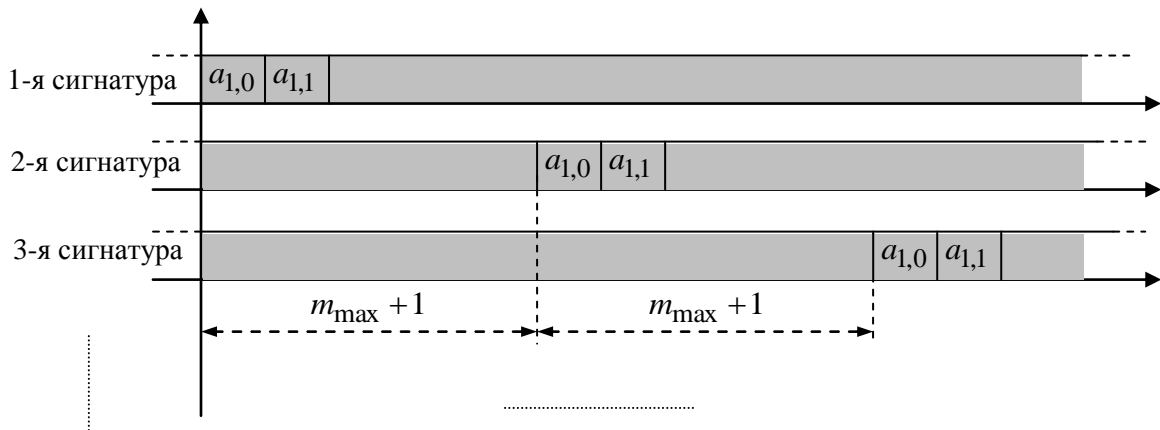


Рис.7.17. Сигнатуры, образованные сдвигом во времени исходной последовательности.

Предположим теперь, что исходная последовательность $\{a_{1,i}\}$ обладает либо идеальной, либо достаточно хорошей периодической АКФ $\rho_{11}(m)$. Первая возможна, например, для троичных или многофазных последовательностей (см. параграф 6.11), тогда как любая минимаксная бинарная последовательность (см. параграфы 6.7, 6.9) может служить в качестве примера последней. Идея заключается в том, что все боковые лепестки $\rho_{11}(m)$ являются пренебрежимо малыми. Тогда при $|m| \leq m_{\max}$ аргумент в квадратных скобках (7.47) обращается в нуль по модулю L только в случае $k = l$ и $m = 0 \pmod L$, что отвечает основному лепестку k -й сигнатуры. Для любой другой комбинации k, l, m правая часть соотношения (7.47) соответствует боковому лепестку $\rho_{11}(m)$, уровень которого полагается несущественным. Таким образом, доказано, что соответствующие сдвиги копий исходной последовательности с хорошей периодической АКФ образуют ансамбль, для которого выполняются условия псевдослучайности (7.43) во всем диапазоне значений взаимных сдвигов сигнатур $|m| \leq m_{\max}$. Отсюда непосредственно следует, что этот ансамбль достигает (при идеальной $\rho_{11}(m)$) или очень близко приближается к наименьшему уровню (7.40) усредненных нежелательных эффектов, обусловленных помехами МАИ и многолучевого распространения, или, что эквивалентно, границам Велча (7.36), (7.35).

Пример 7.4.1. Рассмотрим систему с длительностью чипа $\Delta = 1$ мксек, числом поль-

зователей $K = 60$, канальной задержкой на распространение $\tau_{ds} = 20$ мксек и радиусом соты $D_c = 15$ км. В этом случае $\tau_{\max} = 2D_c/c + \tau_{ds} = 120$ мксек, а $m_{\max} = 120$. Ансамбль сигнатур может быть построен на основе начальной последовательности $\{a_{1,i}\}$, период которой должен быть равным $L \geq K(m_{\max} + 1) = 60 \times 121 = 7260$. Поскольку $\{a_{1,i}\}$ должна иметь хорошую периодическую АКФ, то в роли подходящих кандидатов могут выступать троичная последовательность с идеальной АКФ длины $L = 8011$, бинарная m -последовательность длины $L = 2^{13} - 1 = 8191$ или последовательность Лежандра с $L = 7283$. Тогда 60 сигнатур будут образованы как циклические копии $\{a_{1,i}\}$, сдвинутые друг относительно друга на 121 чипов. Очевидно, что на длину последовательности не накладывается верхний предел, и, возможно, разумнее выбрать ее больше с соответствующим увеличением сдвига между сигнатурами, чтобы гарантировать некоторый резерв безопасности. \square

Обратные каналы стандартов 2-го поколения cdmaOne (IS-95) и 3-го поколения cdma2000 представляют хороший пример реализации данной версии асинхронного варианта CDMA [69]. Бинарная m -последовательность чрезвычайно большой длины $L = 2^{42} - 1$, расширенная еще на один символ, используется в качестве исходной, а конкретные пользовательские сигнатуры всех мобильных станций представляют собой ее соответствующие циклические копии. Псевдослучайные свойства m -последовательности совместно с сигнатурными сдвигами, превосходящими возможные изменения времени прихода сигнала на приемник БС, гарантируют минимальный (см. (7.40)) уровень мощности помех MAI и многолучевости на выходе коррелятора.

7.5. Примеры ансамблей минимаксных сигнатур.

Ансамбли сигнатур, рассмотренные в предшествующем параграфе, могут быть признаны адекватными только в ситуациях, когда взаимные временные сдвиги пользовательских сигналов полностью контролируются системой и могут быть удержаны в рамках предсказанного диапазона. Если же это не выполняется, то асинхронный CDMA, основанный на использовании сдвинутых копий одной и той же последовательности, подвержен возникновению коллизий: сигнал одного из пользователей может приобрести задержку, не позволяющую отличить его от сигнала некоторого другого пользователя. Последнее может служить основанием для использования ансамблей минимаксных сигнатур, т.е. таких ансамблей, корреляционный максимум которых достигает или приближается к границам (7.45) или (7.46). Поскольку корреляционный пик минимаксного ансамбля получен в результате максимизации на всем периоде, то его малое значение (достигнутое за счет достаточно большой длины L) обеспечивает близость корреляционных свойств ансамбля идеальным (7.43), гарантируя псевдослучайность сигнатур.

Обзор всех известных минимаксных ансамблей занял бы достаточно большой объем, так что ограничимся лишь кратким обсуждением тех из них, которые либо имеют широкое практическое использование, либо оказываются наиболее показательными среди других. Читатель, заинтересованный в более подробном изложении, может обратиться к [9, 67, 70].

7.5.1. Бинарные частотно-сдвинутые m -последовательности.

Возьмем бинарную $\{\pm 1\}$ m -последовательность $\{a_{1,i}\}$ периода $L = 2^n - 1$ и используем ее в качестве сигнатуры для первого пользователя. Остальные $K - 1$ сигнатур формируются посимвольным умножением $\{a_{1,i}\}$ на дискретные гармоники частот $(k - 1)/L$, $k = 2, 3, \dots, K$:

$$a_{k,i} = a_{1,i} \exp\left(j \frac{2\pi(k-1)i}{L}\right), \quad i = \dots, -1, 0, 1, \dots; \quad k = 1, 2, \dots, K. \quad (7.48)$$

Тогда квадрат модуля периодической ВКФ k -й и l -й последовательностей запишется как

$$\left|R_{p,kl}(m)\right|^2 = \left|\sum_{i=0}^{L-1} a_{k,i} a_{l,i-m}^*\right|^2 = \left|\sum_{i=0}^{L-1} a_{1,i} a_{1,i-m} \exp\left(j \frac{2\pi(k-l)i}{L}\right)\right|^2. \quad (7.49)$$

Первоначально рассмотрим случай $m = 0 \bmod L$, т.е. $a_{1,i} a_{1,i-m} = |a_{1,i}|^2 = 1$. Тогда, если $k = l$, соотношение (7.49) определяет основной лепесток АКФ k -й сигнатуры, т.е. $\left|R_{p,kk}(0)\right|^2 = L^2$. Если же $k \neq l$, то сумма в (7.49) есть сумма всех корней из единицы степени L , которая равняется нулю (см. параграф 6.11.2). Теперь пусть $m \neq 0 \bmod L$. Тогда, согласно свойству сдвига и сложения m -последовательности (параграф 6.11), $a_{1,i} a_{1,i-m} = a_{1,i-t}$ для некоторого t , и квадрат модуля ВКФ будет определяться как

$$\left|R_{p,kl}(m)\right|^2 = \left|\sum_{i=0}^{L-1} a_{1,i-t} \exp\left(j \frac{2\pi(k-l)i}{L}\right)\right|^2 = \left|\sum_{i=0}^{L-1} a_{1,i} \exp\left(j \frac{2\pi(k-l)i}{L}\right)\right|^2,$$

что представляет собой $(k - l)$ -ю компоненту ДПФ энергетического спектра последовательности $\{a_{1,i}\}$. Поскольку энергетический спектр $\{a_{1,i}\}$ есть ДПФ от ее периодической АКФ, а последняя равняется -1 при всех значениях аргумента кроме нуля, где она равня-

ется L , получаем

$$\left| R_{p,kl}(m) \right|^2 = \sum_{m=0}^{L-1} R_{p,11}(m) \exp\left(j \frac{2\pi(k-l)m}{L} \right) = L+1 - \sum_{m=0}^{L-1} \exp\left(j \frac{2\pi(k-l)m}{L} \right).$$

Последняя сумма отличается от нуля и равна L только при $k=l$, так что, учитывая все полученные результаты и перейдя к нормированным корреляциям, получаем

$$\left| \rho_{p,kl}(m) \right|^2 = \begin{cases} 1, & k=l, m=0 \bmod L, \\ \frac{1}{L}, & k=l, m \neq 0 \bmod L, \\ 0, & k \neq l, m=0 \bmod L, \\ \frac{L+1}{L^2}, & k \neq l, m \neq 0 \bmod L. \end{cases}$$

Откуда видно, что квадрат корреляционного пика ансамбля (7.48)

$$\rho_{\max}^2 = \frac{L+1}{L^2} \approx \frac{1}{L},$$

т.е. практически совпадает с границей Велча (7.45). Таким образом, рассматриваемый ансамбль является минимаксным, оптимальным образом реализует асинхронный вариант CDMA.

Описание вышеприведенного ансамбля можно найти, например, в [71], а еще ранее и независимо он был использован в глобальной спутниковой навигационной системе ГЛОНАСС (см. параграф 11.1) Одним из достоинств данного множества сигнатур в сравнении с другими многофазными ансамблями является возможность формирования сигнатур простым сдвигом несущей частоты. Действительно, увеличение несущей частоты f_0 на $(k-1)/L\Delta$ эквивалентно линейному приращению фазы между соседними чипами на величину $2\pi(k-1)/L$, что в точности совпадает с предписанным правилом (7.48).

Несмотря на наличие множества других минимаксных многофазных ансамблей, бинарные $\{\pm 1\}$ семейства традиционно считаются более привлекательными с реализационной точки зрения, и остальная часть параграфа будет посвящена некоторым важным примерам множеств бинарных сигнатур.

7.5.2. Множества Голда.

Следующие свойства бинарных $\{\pm 1\}$ m -последовательностей служат объяснением метода построения множества, найденного Голдом.

1. Если бинарную m -последовательность $\{u_i\}$ периода $L=2^n-1$ подвергнуть операции *децимирования с индексом децимации d* , где d взаимно простое с L , то результирующая последовательность $\{v_i\}$ снова окажется бинарной m -последовательностью такого же периода. Операция децимирования означает выбор каждого d -го символа последовательности $\{u_i\}$ и запись полученных таким образом элементов друг за другом, так что $v_i = u_{di}$. Назовем полученную в результате указанной процедуры последовательность $\{v_i\}$ децимацией последовательности $\{u_i\}$.

2. Пусть память n бинарной m -последовательности $\{u_i\}$ является нечетным числом, а индекс децимации $d=2^s+1$, где s взаимно просто с n . Тогда d взаимно простое число с длиной $L=2^n-1$ последовательности $\{u_i\}$, децимированная последовательность $\{v_i\}$ является m -последовательностью того же периода L , а ненормированная периодическая ВКФ $R_{p,uv}(m)$ последовательностей $\{u_i\}$ и $\{v_i\}$ принимает только три следующих значения:

$$R_{p,uv}(m) \in \left\{ \pm \sqrt{2(L+1)} - 1, -1 \right\} = \left\{ \pm 2^{\frac{n+1}{2}} - 1, -1 \right\}, m = 0, 1, \dots, L-1. \quad (7.50)$$

3. Пусть память n бинарной m -последовательности $\{u_i\}$ является четным числом, но не кратно четырем, а индекс децимации $d = 2^s + 1$, где s – четное и взаимно простое с $n/2$. Тогда d взаимно просто с длиной $L = 2^n - 1$ последовательности $\{u_i\}$, децимированная последовательность $\{v_i\}$ является m -последовательностью того же периода L , а ненормированная периодическая ВКФ $R_{p,uv}(m)$ последовательностей $\{u_i\}$ и $\{v_i\}$ принимает только три следующих значения:

$$R_{p,uv}(m) \in \left\{ \pm 2\sqrt{L+1} - 1, -1 \right\} = \left\{ \pm 2^{\frac{n+2}{2}} - 1, -1 \right\}, m = 0, 1, \dots, L-1. \quad (7.51)$$

Доказательство этих утверждений достаточно сложно и требует более глубоких знаний об алгебре расширенных конечных полей. Оставляем их доказательство в стороне и отсылаем заинтересованного читателя к оригинальной статье Голда [72] или другим источникам [например, 9, 70].

Возьмем теперь пару m -последовательностей: $\{u_i\}$ и ее децимацию $\{v_i\}$, удовлетворяющую вышеприведенным условиям 2 или 3, и сформируем ансамбль из K сигнатур по следующему правилу

$$\begin{aligned} a_{k,i} &= u_i v_{i-k}, k = 1, 2, \dots, L, \\ a_{L+1,i} &= u_i, \\ a_{L+2,i} &= v_i, \end{aligned} \quad (7.52)$$

где $i = \dots, -1, 0, 1, \dots$. Другими словами, строим L сигнатур путем посимвольного умножения $\{u_i\}$ на циклические копии $\{v_i\}$, а еще две сигнатуры образуют непосредственно сами исходные m -последовательности. В итоге имеем всего $K = L + 2 = 2^n + 1$ сигнатур. На практике традиционно бинарные $\{\pm 1\}$ m -последовательности формируются как двоичные $\{0,1\}$ последовательности, т.е. над полем $GF(2)$ с помощью линейного регистра сдвига, с последующим отображением элементов $GF(2)$ на вещественную пару $\{\pm 1\}$ (см. параграфы 6.6, 6.7). Таким образом, реализация соотношений (7.52) может быть осуществлена с помощью двух n -разрядных LFSR, генерирующих $\{0,1\}$ последовательности $\{u'_i\}$ и $\{v'_i\}$, предшественников $\{u_i\}$ и $\{v_i\}$. Действительно, вместо умножения $\{u_i\}$ на $\{v_{i-k}\}$ может быть реализована операция сложения по модулю 2 их предшественников с последующим отображением результата на множество $\{\pm 1\}$: $u_i v_{i-k} = (-1)^{u'_i + v'_{i-k}}$. Рис. 7.18 ил-

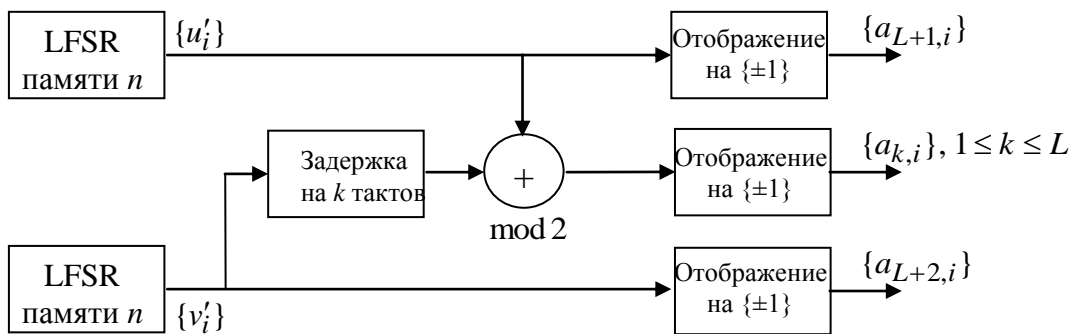


Рис.7.18. Генерирование последовательностей Голда.

люстрирует практическую реализацию алгоритма Голда в соответствие с вышеприведенным обсуждением.

Оценим корреляционный пик ансамбля Голда, начав с вычисления корреляций первых L последовательностей

$$R_{p,kl}(m) = \sum_{i=0}^{L-1} a_{k,i} a_{l,i-m} = \sum_{i=0}^{L-1} u_i u_{i-m} v_{i-k} v_{i-l-m}.$$

Очевидно, что поскольку случай $m=0 \bmod L$ и $k=l$ отвечает основному лепестку АКФ k -й сигнатуры, то следует анализировать ситуацию, когда приведенные равенства одновременно не выполняются. Однако, тогда либо $u_i u_{i-m}$, либо $v_{i-k} v_{i-l-m}$ представляют собой некоторые другие сдвиги исходных последовательностей $\{u_i\}$, $\{v_i\}$, либо только одно из этих произведений есть последовательность, состоящая из одних единиц. В первом случае имеем ВКФ исходных m -последовательностей $\{u_i\}$, $\{v_i\}$, принимающую только три значения, указанных в (7.50) или (7.51), тогда как во втором – боковые лепестки ненормированной АКФ одной из последовательностей $\{u_i\}$, $\{v_i\}$, т.е. -1 .

Рассмотрим теперь ВКФ последовательностей $\{a_{k,i}\}$, $k=1,2,\dots,L$ и $\{a_{l,i}\}$, $l=L+1$:

$$R_{p,kl}(m) = \sum_{i=0}^{L-1} u_i u_{i-m} v_{i-k}.$$

Если $m=0 \bmod L$, то $u_i u_{i-m} = 1$, и величина ВКФ есть просто постоянная составляющая $\{v_i\}$, т.е. -1 . В противном случае, $u_i u_{i-m} = u_{i-s}$ при некотором s , и имеем ВКФ исходных m -последовательностей, подчиняющуюся ограничениям (7.50) или (7.51). Аналогичный результат имеет место для ВКФ последовательностей $\{a_{k,i}\}$, $k=1,2,\dots,L$ и $\{a_{l,i}\}$, $l=L+2$.

Наконец ВКФ $\{a_{L+1,i}\}$ и $\{a_{L+2,i}\}$ непосредственно представляет собой ВКФ исходных m -последовательностей, тогда как их автокорреляционные функции, как АКФ m -последовательностей, обладают боковыми лепестками ненормированной АКФ, равными -1 . Учитывая совместно все полученные результаты, видно, что корреляционный пик (7.44) множества Голда определяется максимальным по модулю значением ВКФ, устанавливаемым соотношениями (7.50) или (7.51). После нормировки к длине L приходим к следующей оценке

$$\rho_{\max}^2 \leq \begin{cases} \frac{(\sqrt{2(L+1)}+1)^2}{L^2}, & n \neq 0 \bmod 2, \\ \frac{(2\sqrt{(L+1)}+1)^2}{L^2}, & n = 2 \bmod 4 \end{cases} \approx \begin{cases} \frac{2}{L}, & n \neq 0 \bmod 2, \\ \frac{4}{L}, & n = 2 \bmod 4, \end{cases} \quad (7.53)$$

с последним приближением, отвечающим большой длине $L \gg 1$. Как видно, при любом нечетном значении памяти n ансамбли сигнатур Голда асимптотически ($L \gg 1$) удовлетворяют нижней границе Сидельникова (7.46), тогда как при четном n , не кратном четырем, их проигрыш в уровне ρ_{\max} по отношению к граничному составляет порядка 3 дБ.¹

Пример 7.5.1. Построим последовательности Голда длины $L = 2^3 - 1 = 7$. Ансамбль такой малой длины не имеет практического значения, однако служит хорошей иллюстра-

¹ При $n = 0 \bmod 4$ также существует ансамбль Голда с тем же значением корреляционного пика, как и в случае $n = 2 \bmod 4$, но с числом последовательностей на одну меньше [67, 70].

цией идеи построения. Начнем с бинарной $\{0,1\}$ m -последовательности, впервые встретившейся в примере 6.6.1: $\{u'_i\} = \{1,0,0,1,0,1,1\}$. Индекс децимации $d = 3$ удовлетворяет вышеприведенному ограничению в пункте 2. Тогда децимированная последовательность $\{v_i\} = \{1,1,1,0,1,0,0\}$. Посимвольное суммирование $\{u'_i\}$ и $\{v_i\}$ по модулю два дает последовательность $\{0,1,1,1,1,1,1\}$, которая после отображения на алфавит $\{\pm 1\}$ дает первую последовательность Голда $\{a_{1,i}\} = \{+-----\}$. Сдвиг $\{v'_i\}$ вправо на одну позицию и сложение по модулю 2 с $\{u'_i\}$ дает последовательность $\{1,1,1,0,0,0,1\}$, которая после перехода к символам $\{\pm 1\}$ дает вторую последовательность Голда $\{---+++-\}$. Еще шесть последовательностей Голда получаются в результате дальнейших сдвигов $\{v'_i\}$, сложения по модулю 2 с $\{u'_i\}$ и изменения символов на $\{\pm 1\}$. Совместно с $\{u'_i\}$ и $\{v'_i\}$, преобразованных в $\{\pm 1\}$ последовательности, получаем всего $K = 2^3 + 1 = 9$ последовательностей. Проверка значения пика корреляции в данном простейшем случае не имеет особого значения, поскольку при $L = 7$ ненормированная периодическая корреляция не может превосходить значения 5, предсказываемого соотношением (7.53). Построения ансамблей Голда больших длин и проверка их оптимальности является предметом задачи 7.40. \square

Ансамбли Голда пользуются большой популярностью в современных CDMA системах. Достаточно сказать, что они используются в глобальной спутниковой навигационной системе GPS NAVSTAR для разделения сигналов спутников, в системе мобильной связи 3-го поколения стандарта UMTS для скремблирования CDMA кодов и т.п.

7.5.3. Множества Касами и их расширения.

Принцип построения множеств Касами очень близок к описанной выше схеме конструирования последовательностей Голда. Осуществим операцию децимирования над бинарной $\{\pm 1\}$ m -последовательностью $\{u_i\}$ четной памяти $n = 2h$ с индексом децимации $d = 2^h + 1$. Очевидно, что d не взаимно простое число с периодом $L = 2^n - 1 = (2^h - 1)(2^h + 1)$ последовательности $\{u_i\}$, что приводит к децимированной последовательности $\{v_i\} = \{u_{di}\}$ с периодом, являющимся множителем L . Можно показать, что если $\{u_i\}$ инициализирована так, что $u_0 = -1$, то «короткая» последовательность $\{v_i\}$ будет бинарной m -последовательностью периода $L_1 = 2^h - 1$, ненормированная периодическая ВКФ которой с $\{u_i\}$ на периоде L принимает только два значения [9,70,73]:

$$R_{p,uv}(m) = \pm 2^h - 1 = \pm \sqrt{L+1} - 1, \quad m = 0, 1, \dots, L-1. \quad (7.54)$$

Тогда L_1 сигнатур Касами длины L образуются путем посимвольного произведения исходной m -последовательности $\{u_i\}$ с L_1 различными циклическими копиями $\{v_i\}$, а еще одной сигатурой является сама «длинная» последовательность, так что:

$$\begin{aligned} a_{k,i} &= u_i v_{i-k}, \quad k = 1, 2, \dots, L_1, \\ a_{L_1+1,i} &= u_i. \end{aligned} \quad (7.55)$$

где $i = \dots, -1, 0, 1, \dots$. Таким образом, всего имеется $K = L_1 + 1 = 2^h = \sqrt{L+1}$ подобных сигнатур периода L . Очевидно, что снова умножение $\{\pm 1\}$ последовательностей $\{u_i\}$, $\{v_i\}$ может быть реализовано как сложение по модулю 2 их $\{0,1\}$ предшественников $\{u'_i\}$, $\{v'_i\}$, но в отличие от множества Голда для формирования «короткой» последовательности $\{v'_i\}$ необходим регистр сдвига LFSR в два раза меньшей длины $h = n/2$ (см. рис. 7.19).

Доказательство минимаксных свойств множества Касами

$$\rho_{\max}^2 = \frac{(\sqrt{L+1}+1)^2}{L^2} \approx \frac{1}{L}, \quad L \gg 1 \quad (7.56)$$

осуществляется на основе (7.54) аналогично доказательству свойств множества Голда и оставляется читателю в качестве упражнения (задача 7.28). Сравнение двух бинарных ансамблей показывает значительный (6 дБ) выигрыш множеств Касами в уровне корреляционного пика у ансамблей Голда одинаковой длины за счет значительно меньшего (в $(L+2)/\sqrt{L+1} \approx \sqrt{L}$ раз) числа последовательностей K .¹

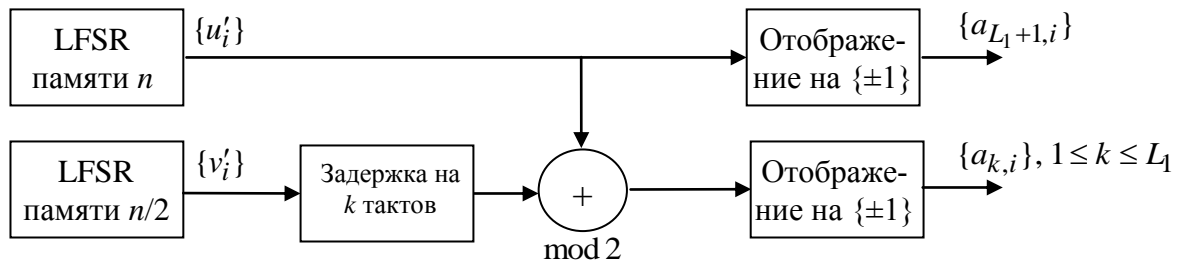


Рис.7.19 Формирование последовательностей Касами.

Пример 7.5.2. Сконструируем множество Касами длины $L = 2^4 - 1$ ($h = 2, K = \sqrt{L+1} = 4$). Начнем с построения бинарной $\{0,1\}$ m -последовательности $\{u'_i\}$ длины $L = 2^4 - 1 = 15$ на основе примитивного полинома $f(x) = x^4 + x + 1$ и начального состояния регистра $u'_0 = 1, u'_1 = u'_2 = u'_3 = 0$. Получаем $\{u'_i\} = \{1, 0, 0, 0, 1, 0, 0, 1, 1, 0, 1, 0, 1, 1, 1\}$. Децимирование этой последовательности с индексом $d = 2^h + 1 = 5$ дает m -последовательность периода три $\{v_i\} = \{1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1\}$. Сумма по модулю 2 последовательности $\{u'_i\}$ с тремя сдвинутыми копиями $\{v'_i\}$ после перехода к алфавиту $\{\pm 1\}$ образует первые три последовательности Касами:

$$\begin{aligned} \{a_{1,i}\} &= \{++-----+----+-\}, \\ \{a_{2,i}\} &= \{+-+--+--+---++++\}, \\ \{a_{3,i}\} &= \{----++-+++++--++\}. \end{aligned}$$

Четвертым представителем является $\{u'_i\}$, символы который преобразованы во множество $\{\pm 1\}$: $\{a_{4,i}\} = \{-++++-+-+---\}$. Непосредственное вычисление показывает, что все их ненормированные ВКФ, также как и боковые лепестки ненормированных АКФ первых трех последовательностей принимают только значения вида -5 и 3 , так что $\rho_{\max}^2 = 1/9$ в полном соответствии с (7.56). Построение произвольных множеств Касами и проверка их корреляционных свойств на языке программирования Matlab является предметом задачи 7.41. \square

Относительно малое число последовательностей Касами подчеркивает важность способа, позволяющего расширить множество Касами почти в два раза без изменения величины корреляционного пика, который был предложен Камалетдиновым [74]. Пусть n

¹ Для бинарных множеств границы (7.45), (7.46) могут быть незначительно улучшены, если учесть, что ненормированные корреляции принимают только целые значения. В результате оказывается, что как множества Голда с нечетной памятью, так и множества Касами строго (а не только асимптотически) оптимальны по уровню корреляционного пика среди всех бинарных множеств [67, 70].

кратно 4, т.е. $n = 4r$, где r – целое, так что $L = 2^{4r} - 1 = 16^r - 1 = 15, 255, 4095, \dots$. Тогда при данных параметрах в дополнение к множеству Касами существует и другой бинарный ансамбль длины L и объема $\sqrt{L+1}$, называемый ансамблем бент–последовательностей [9, 75], который обладает тем же минимаксным свойством $\rho_{\max}^2 = (\sqrt{L+1} + 1)^2 / L^2 \approx 1/L$. В самых общих выражениях построение бент–последовательностей снова состоит в посимвольном умножении двух исходных последовательностей: «длинной» m – последовательности периода $L = 2^{4r} - 1$ и некоторой специальной последовательности, основанной на т.н. бент–функциях. Детали построения указанных последовательностей достаточно сложны и здесь могут быть опущены, однако основным является тот факт, что значение нормированной ВКФ любой бент–последовательности с любой из первых L_1 последовательностей Касами (7.55) не превосходит по модулю корреляционный пик как ансамбля Касами, так и бент–последовательностей. Следовательно, возможно образовать составной ансамбль, включающий $L_1 = 2^h - 1 = 2^{2r} - 1 = \sqrt{L+1} - 1$ последовательностей Касами и $\sqrt{L+1}$ бент–последовательностей и обладающий прежним значением корреляционного пика $\rho_{\max}^2 = (L+1)/L^2 \approx 1/L$. Полученный таким образом ансамбль является уникальным в том смысле, что среди всех известных бинарных ансамблей со значением корреляционного пика $\rho_{\max}^2 \approx 1/L$ только он один обладает наибольшим числом сигнатур $K = 2\sqrt{L+1} - 1$.

7.5.4. Ансамбли Камалетдинова.

Существуют и другие бинарные минимаксные ансамбли [9, 67], отличающиеся от рассмотренных только тонкой структурой последовательностей, но не значениями длины L , объема K и корреляционного максимума ρ_{\max} . На этом фоне особый интерес представляют ансамбли, описанные Камалетдиновым [76], поскольку охватывают диапазон длин, отличный от установленных для множеств Голда и Касами.

Для более ясного восприятия идеи рассмотрим отчасти упрощенную версию множеств Камалетдинова, не приводящую, тем не менее, к потерям как в диапазоне охвата длин, так и достижимых значений параметров. Для обрисования в общих чертах первой схемы Камалетдинова возьмем простое нечетное $p > 3$ вида $p = 4h + 3 = 3 \pmod{4}$ и распространим определение двузначного характера $\psi(x)$, данное в параграфе 6.8, на нулевой элемент $GF(p)$, полагая $\psi(0) = 1$ (альтернативный вариант $\psi(0) = -1$ приводит к аналогичному результату). Будем трактовать позицию символа последовательности с номером i как элемент поля $GF(p)$, т.е. приведенным по модулю p , и образуем $p+1$ p –ичных последовательностей $d_{k,i}$ над $GF(p)$ (т.е. с элементами из этого поля) по правилу:

$$d_{k,i} = \begin{cases} i + \alpha^{i+k} + \alpha^{-i}, & k = 1, 2, \dots, p-1, \\ i + \alpha^i, & k = p, \\ i + \alpha^{-i}, & k = p+1, \end{cases} \quad (7.57)$$

где все арифметические операции осуществляются по правилам конечных полей, α – примитивный элемент $GF(p)$ и $i = \dots, -1, 0, 1, \dots$. Можно заметить, что каждая последовательность в (7.57) образована как сумма последовательностей с взаимно простыми периодами p и $p-1$ ($\alpha^{p-1} = \alpha^0 = 1$) и, следовательно, обладает периодом $L = p(p-1)$. Осуществим теперь отображение последовательностей (7.57) на бинарный алфавит $\{\pm 1\}$, исполь-

зую понятие расширенного двузначного характера

$$a_{k,i} = \psi(d_{k,i}), \quad k = 1, 2, \dots, p+1; \quad i = \dots, -1, 0, 1, \dots \quad (7.58)$$

Сформированное подобным образом бинарное множество обладает следующими параметрами

$$L = p(p-1), \quad K = p+1, \quad \rho_{\max}^2 = \frac{(p+3)^2}{L^2}. \quad (7.59)$$

Длина L может быть сделана достаточно большой только за счет выбора p ($p \gg 1$), при котором $(p+3)^2/L = (p+3)^2/(p^2-p) \approx 1$ и $\rho_{\max}^2 \rightarrow 1/L$, доказывая после сравнения с (7.45), по крайней мере, асимптотическую оптимальность ансамбля по уровню пика корреляции.

Пример 7.5.3. Пусть $p = 7$. Прямая проверка подтверждает, что $\alpha = 3$ является примитивным элементом поля $GF(7)$. Тогда последовательности $\{\alpha^i\}$ и $\{\alpha^{-i}\}$ обладают периодом $p-1 = 6$ и имеют вид: $\{\dots, 1, 3, 2, 6, 4, 5, 1, 3, \dots\}$ и $\{\dots, 1, 5, 4, 6, 2, 3, 1, 5, \dots\}$ соответственно. Комбинирование их по модулю 7 с последовательностью $\{i\} = \{\dots, 0, 1, 2, 3, 4, 5, 6, 0, 1, \dots\}$ периода 7, как это предписано (7.57), дает $K = p+1 = 8$ 7-значных последовательностей периода $L = p(p-1) = 42$. Например, первой из них является $\{d_{1,i}\} = \{221136110025006614665503554462443351332240\}$. Замена 7-значных элементов расширенными характерами согласно правилу $\psi(0) = \psi(1) = \psi(2) = \psi(4) = 1$ и $\psi(3) = \psi(5) = \psi(6) = -1$ превращает последовательности в их 8 бинарных аналогов, например, $\{a_{1,i}\} = \{++++--+++++---+-----+-----+---++++-+---++++\}$. Достаточно утомительно вычислять их АКФ и ВКФ «в ручную», поэтому задача 7.42 оказывает программную поддержку в решении данной проблемы. \square

Вторая конструкция Камалетдинова использует p -ичную ($p = 4h+3 = 3 \pmod{4}$) линейную последовательность $\{c_i\}$, полученную децимацией с индексом $d = p-1$ $p-1$ сдвига $\{d_{i+k}\}, k = 1, 2, \dots, K = p-1$ p -ичной m -последовательности $\{d_i\}$ памяти $n = 2$, т.е. длины $p^2 - 1$. Поскольку d делит $p^2 - 1$, то последовательность $\{c_i\} = \{d_{di+k}\}$ обладает периодом $(p^2 - 1)/(p-1) = p+1$. Теперь построим $p-1$ последовательность над $GF(p)$

$$d_{k,i} = i + d_{di+k}, \quad k = 1, 2, \dots, p-1; \quad i = \dots, -1, 0, 1, \dots, \quad (7.60)$$

и отобразим их на бинарный алфавит $\{\pm 1\}$ согласно (7.58). В результате имеем ансамбль с параметрами

$$L = p(p+1), \quad K = p-1, \quad \rho_{\max}^2 = \frac{(p+1)^2}{L^2} = \frac{1}{p^2}. \quad (7.61)$$

Снова для больших длин ($p \gg 1$) отношение $(p+1)^2/L = (p+1)^2/(p^2+p) \rightarrow 1$ и $\rho_{\max}^2 \rightarrow 1/L$, демонстрируя, по крайней мере, асимптотическую оптимальность ансамбля.

Пример 7.5.4. В данном случае отсутствует исключение для $p = 3$ и p -ичная m -последовательность $\{d_i\}$ памяти $n = 2$ и длины $p^2 - 1 = 8$ может быть сформирована с помощью примитивного полинома над $GF(3)$ второй степени $f(x) = x^2 + x + 2$, или, что эквивалентно, с помощью рекурсии $d_i = 2d_{i-1} + d_{i-2}$. При начальном состоянии регистра $d_0 = 1, d_1 = 0$ получаем последовательность $\{d_i\} = \{\dots, 1, 0, 1, 2, 2, 0, 2, 1, 1, 0, \dots\}$. Ее сдвиги, децимированные с индексом $d = p-1 = 2$, трансформируются в две последовательности с

периодом 4: $\{\dots, 1, 1, 2, 2, 1, 1, 2, 2, \dots\}$ и $\{\dots, 0, 2, 0, 1, 0, 2, 0, 1, \dots\}$. После посимвольного сложения с последовательностью $\{i\} = \{\dots, 0, 1, 2, 0, 1, 2, \dots\}$ они дают $K = p - 1 = 2$ последовательностей периода $L = p(p + 1) = 12$: $\{1, 2, 1, 2, 2, 0, 2, 0, 0, 1, 0, 1\}$ и $\{0, 0, 2, 1, 1, 1, 0, 2, 2, 2, 1, 0\}$. Последний шаг, состоящий в замене их элементов расширенными характеристиками $\psi(0) = \psi(1) = +1$, $\psi(2) = -1$, приводит к образованию множества Камалетдинова из двух бинарных последовательностей длины $L = 12$: $\{a_{1,i}\} = \{+-+--+-++++\}$ и $\{a_{2,i}\} = \{+++++---++\}$. Вычисление значений их АКФ и ВКФ не составляет труда выполнить вручную (или с помощью программы из задачи 7.43), которое приводит к $\rho_{\max}^2 = 1/9$ в полном соответствии с (7.61). \square

Для доказательства утверждений о величине пика корреляции в вышеприведенных ансамблях необходимо привлечение теории квадратичных уравнений в конечных полях. Оставляя эту сложную задачу вне нашего внимания, рекомендуем обратиться заинтересованному читателю к оригинальной статье [76].

В таблице 7.1 представлены итоговые результаты рассмотрения бинарных минимаксных ансамблей, в которой приведены значения длины (перечень всех длин существующих ансамблей в диапазоне $7 \leq L \leq 1023$), числа сигнатур и квадрата максимума корреляции. Таблица достаточно выразительно подчеркивает вклад множеств Камалетдинова: в рассматриваемом интервале число их длин составляет 11 против 6 для Голда и 4 для Касами.

Таблица 7.1. Примеры ансамблей бинарных минимаксных сигнатур.

Ансамбль	Длина L	Объем K	Квадрат максимума корреляции ρ_{\max}^2
Голд	$2^n - 1, n \neq 0 \pmod{4}$ 7, 31, 63, 127, 511, 1023	$L + 2 = 2^n + 1$	$\frac{(\sqrt{2(L+1)} + 1)^2}{L^2}$ $\rightarrow \frac{2}{L}, n$ нечетное, $\frac{(2\sqrt{(L+1)} + 1)^2}{L^2}$ $\rightarrow \frac{4}{L}, n$ четное
Касами	$2^n - 1, n - \text{четное}$ 15, 63, 255, 1023	$\sqrt{L+1}$	$\frac{(\sqrt{L+1} + 1)^2}{L^2} \rightarrow \frac{1}{L}$
Объединение Касами и бент последовательностей	$2^n - 1, n = 0 \pmod{4}$ 15, 255	$2\sqrt{L+1} - 1$	$\frac{(\sqrt{L+1} + 1)^2}{L^2} \rightarrow \frac{1}{L}$
Камалетдинов 1	$p(p-1), p - \text{простое}$ 42, 110, 342, 506, 930	$p+1 = \frac{\sqrt{4L+1} + 3}{2}$ $\rightarrow \sqrt{L}$	$\frac{(p+3)^2}{L^2} \rightarrow \frac{1}{L}$
Камалетдинов 2	$p(p+1), p - \text{простое}$ 12, 56, 132, 380, 552, 992	$p-1 = \frac{\sqrt{4L+1} - 3}{2}$ $\rightarrow \sqrt{L}$	$\frac{(p+1)^2}{L^2} \rightarrow \frac{1}{L}$

Задачи.

7.1. В CDMA системе с прямым расширением спектра, основанной на периодических бинарных сигнатурах и передаче данных с бинарной ФМ, пользователь передает сигнал вида $\{++----+----++++----+-\}$, занимающий интервал более двух бит данных. Какой код сигнатуры использует этот абонент (общий знак всех символов является несущественным), если длительность бита данных равна периоду сигнатуры?

7.2. В CDMA системе с прямым расширением спектра, основанной на периодических бинарных сигнатурах и передаче данных с бинарной ФМ, некоторый абонент использует сигнатурный код вида $\{+++--+-\}$ с длительностью бита данных, равной периоду сигнатуры 7Δ . Вследствие неудачи в синхронизации по времени сужающий опорный сигнал приемника отстает от принятого расширенного сигнала на один чип. Каким будет результат демодуляции данных в случае передачи нулевого потока бит?

7.3. В какой степени наличие амплитудной модуляции в АФМ сигнатуре затронет структуру приемника с прямым расширением спектра? Возвратит ли в этом случае процедура снятия расширения символ данных к виду, отвечающему передаче без расширения спектра?

7.4. CDMA система с прямым расширением спектра использует квадратурную ФМ для передачи данных со скоростью 64 кбит/сек и расширяющий код с чиповой скоростью $1.28 \cdot 10^6$ чип/сек. Определить коэффициент расширения и полосу, занимаемую системой.

7.5. CDMA–система с прыгающей частотой использует 4-х частотный расширяющий сигнал длины $N = 4: \{1, 4, 2, 3\}$ и передачу данных с ЧМ-4 (каждая пара бит передается на одной из 4-х частот). Передается битовый поток вида 00101101. Изобразить возможную частотно–временную решетку переданного сигнала, если один бит данных охватывает две длительности чипа. Какой тип прыгающей частоты используется: быстрый или медленный?

7.6. CDMA система с прыгающей частотой использует 4-х частотный расширяющий сигнал длины $N = 4: \{1, 4, 2, 3\}$ и передачу данных с ЧМ-4 (каждая пара бит передается на одной из четырех частот). Передается битовый поток вида 10110100. Изобразить возможную частотно–временную решетку переданного сигнала, если один сигнатурный чип охватывает два бита данных. Какой тип прыгающей частоты используется: быстрый или медленный?

7.7. CDMA–система с быстрой перестройкой частоты использует 16–ти частотный расширяющий сигнал и модуляцию данных с ЧМ-4. Длительность чипа составляет 10 мксек. Оценить минимальные полосы расширяющего и переданного сигнала, если чипы различных частот будут ортогональны.

7.8. В выделенной полосе $W_f = 76.8$ КГц следует организовать синхронную CDMA систему, передающую данные с помощью бинарной ФМ со скоростью $R = 9.6$ кбит/сек. Сколько пользователей она может обслуживать при сохранении оптимальности однопользовательского приемника? Построить соответствующее множество бинарных сигнатур. На сколько изменится число пользователей, если передачу данных с бинарной ФМ заменить на квадратурную ФМ, ФМ-8 или КАМ-16? Если любой из перечисленных методов модуляции увеличивает число пользователей, то какой ценой это достигается?

7.9. Синхронная CDMA система обслуживает 36 пользователей, используя ортогональные сигнатуры с одинаковой энергией на бит. Сколько новых сигнатур с такой же энергией на бит можно добавить к существующим, не ухудшая минимального расстояния между различными групповыми сигналами?

7.10. Какова минимальная длина синхронных сигнатур, допускающая увеличение не менее чем на 33% числа пользователей в схеме перенасыщения (7.23)?

7.11. Добавить дополнительную сигнатуру к четырем функциям Уолша длины $N = 4$. Является ли дополнительная сигнатура бинарной? Если нет, то можно ли модифицировать исходные сигнатуры так, чтобы дополнительная сигнатура была бы бинарной?

7.12. Построено $K = (4N - 1)/3$ синхронных сигнатур согласно схеме перенасыщения (7.23). Рационально ли их использовать в CDMA системе с K пользователями, если допустим только обычный однопользовательский приемник?

7.13. Определить минимальную длину, потенциально допускающую мощность помех MAI на сигнатуру в однопользовательском приемнике при синхронном перенасыщенном CDMA не выше -30 дБ относительно мощности полезного сигнала, если число пользователей равно 101.

7.14. Доказать, что три и более бинарных последовательностей длины N не могут быть ортогональными друг другу, если их длина не кратна четырем.

7.15. Существует ли перенасыщенное множество из $K = 21$ бинарной сигнатуры, удовлетворяющее границе Велча? Что можно сказать для случая $K = 22, 23$ или 32 ?

7.16. Обрисовать процедуру построения множества из $K = 256$ бинарных последовательностей длины $N = 100$, лежащих на границе Велча.

7.17. (Каристинос–Падос [64]). Доказать, что для перенасыщенного множества нечетного числа K бинарных сигнатур граница Велча (7.30) возрастает до

$$TSC \geq \frac{K^2}{N} + \frac{N-1}{N}.$$

7.18. Построить ансамбль из $K = 15$ бинарных сигнатур длины $N = 12$, достигающих границы, приведенной в предыдущей задаче. Обобщить процедуру на случай $K = 2^m - 1$ сигнатур ($K > N$).

7.19. Каков минимальный период $K = 11$ асинхронных сигнатур, которые не отрицают получение среднего квадрата корреляции между всеми их циклическими копиями в пределах -20 дБ?

7.20. Рассмотрим случайные сигнатуры, удовлетворяющие (7.37). Доказать, что умножение сигнатур на символы данных (модуляция данных) не нарушает соотношения (7.37), обеспеченного независимостью символов данных от символов сигнатур.

7.21. Доказать, что если две последовательности одного и того же наименьшего периода L обладают идеальной периодической АКФ, то их периодическая ВКФ не может равняться нулю при всех взаимных сдвигах.

7.22. Определить максимальное число асинхронных сигнатур периода $L = 100$, которые позволяют удержать максимум корреляции ниже -23 дБ.

7.23. В пределах зоны с радиусом $D_c = 15$ км свободно могут передвигаться $K = 50$ пользователей. Максимальная задержка распространения в канале между пользователем и центральной станцией составляет $\tau_{ds} = 20$ мксек. Полоса, занимаемая системой, составляет 2 МГц. Определить минимальные длины m -последовательности и последовательности Лежандра, позволяющие использовать сдвинутые по времени сигнатуры для канала «пользователь – центральная станция». Определить минимальную длину троичной последовательности с идеальной ПАКФ памяти 3, удовлетворяющую данной задаче.

7.24. В параграфе 7.5.1 для формирования множества частотно-сдвинутых сигнатур используется m -последовательность. Может ли любая другая бинарная минимаксная последовательность (например, последовательность Лежандра) позволить использование этого же метода получения множества сигнатур с квадратом корреляции около $1/L$? Если нет, то почему?

7.25. Система с CDMA использует несущее колебание с длиной волны 4 см и длительностью чипа сигнатуры 1 мксек. Длина сигнатур должна быть $L = 2^{10} - 1 = 1023$. Каково максимальное число частотно-сдвинутых сигнатур может быть образовано, если диапазон скоростей пользователей простирается до 144 км/час?

7.26. Найти все индексы децимации, подходящие алгоритму Голда для длин 63, 127, 511 и 1023.

7.27. Необходим ансамбль сигнатур для обслуживания $K = 100$ пользователей с максимумом корреляции не более 0.064. Какова минимальная длина ансамбля Голда, удовлетворяющего этим требованиям?

7.28. Доказать минимаксные свойства (7.56) множеств Касами.

7.29. Необходим ансамбль сигнатур мощности не менее 31 с максимумом корреляции ниже -23 дБ. Среди известных бинарных множеств найти ансамбль минимальной длины, удовлетворяющий данным требованиям.

7.30. Необходим ансамбль сигнатур мощности не менее 24 с максимумом корреляции ниже -25 дБ. Среди известных бинарных множеств найти ансамбль минимальной длины, удовлетворяющий данным требованиям.

8. Поиск и слежение за сигналом с прямым расширением спектра.

8.1. Процедуры поиска и слежения.

Одной из наиболее характерных задач в технологии распределенного спектра является измерение времени прихода и частоты принятого сигнала. В системах, где широкополосные сигналы служат инструментом местоопределения и измерения параметров движущегося объекта (радиолокация, радионавигация, гидролокация), частотно–временная оценка является основной операцией. В широкополосной связи указанное действие составляет основу процедуры установления временной шкалы. Действительно, для корректной демодуляции передаваемых данных приемник любой цифровой системы связи должен с достаточной точностью определить границы символов, кадров и т. п. в принятом потоке данных. Другими словами, генератор локального приемника должен быть соответствующим образом синхронизирован с принятым потоком данных. В широкополосных системах достаточно точная синхронизация является необходимым условием, поскольку временное рассогласование между принятым расширяющим сигналом и его местной сужающей копией – опорой, превосходящее или равное длительности чипа, полностью нарушит процедуру снятия расширения и последующую демодуляцию данных (см. параграф 7.1). Таким образом, связанные с синхронизацией действия приемника включают предварительное (до начала этапа восстановления данных) согласование собственного сужающего образца с расширяющим кодом принятого сигнала и поддержание достаточно точной синхронизации между ними в течение всего последующего времени приема данных. Очевидно, что с теоретической точки зрения в процедуре синхронизации нет ничего нового: для согласования локальной опоры с принятым сигналом следует только измерить частотно–временной сдвиг принятого сигнала относительно местного генератора. Затем, если это необходимо, генератор приемника может быть скорректирован по времени и частоте и, значит, синхронизирован с принятым сигналом.

Оптимальные (максимально правдоподобные) стратегии измерения временного запаздывания и частоты основательно были рассмотрены в параграфах 2.12–2.14. Однако на практике их «чистая» реализация очень часто сталкивается с серьезными препятствиями. Начальное (например, при первичной инициализации приемника) смещение локального генератора по времени и частоте относительно принятого сигнала может оказаться достаточно большим. К числу факторов, обуславливающих подобное рассогласование, относятся автономное функционирование передающего и приемного генераторов, широкий диапазон изменения длины пути распространения между передатчиком и приемником, доплеровский сдвиг частоты, вызванный их относительным движением, и т. п. В этих обстоятельствах непосредственное применение МП правила иногда может оказаться чрезмерно затруднительным или даже нереализуемым с точки зрения необходимых ресурсов, что демонстрирует следующий пример.

Пример 8.1.1. Дальномерный C/A–сигнал системы GPS Navstar (см. параграф 11.1) обладает периодом $L = 1023$ чипов или, в реальном времени, равным $L\Delta = 1$ мсек. Для решения навигационной задачи необходимо измерять временное положение сигнала с точностью не хуже долей микросекунды, например, 0.1 длительности чипа Δ . Если приемник инициализируется без предварительных сведений о расстройке локального генератора относительно сигнала, то неопределенность о временной задержке сигнала составляет один период, т.е. $L\Delta$. Обращаясь к структуре приемника, содержащей набор корреляторов (см. рис. 2.18), можно увидеть, что ее реализация требует применения 10230 параллельных корреляторов. Переход к структуре с использованием согласованного фильтра (рис. 2.19) не делает проблему проще: цифровой вариант подобного фильтра с памятью $L\Delta$ должен оперировать, по крайней мере, с десятью отсчетами на чип, выполняя 1023

суммирования в течение одного интервала между выборками (менее 100 нсек). Привлечение таких обширных аппаратных или программных ресурсов для решения только одной из многих задач приема не выглядит экономически оправданным, по крайней мере, учитывая существующие технологические тенденции. □

Для избежания трудностей реализационного плана на практике процедуры частотно–временного оценивания в широкой области неопределенности часто реализуются в виде двух последовательных этапов. Первый из них, называемый *поиском (acquisition)* (поиском кода), производит грубое измерение необходимых параметров и обеспечивает предварительными оценками, используемыми на втором этапе, называемом *слежением (tracking)*. Этот второй этап, обычно выполняемый специальными схемами слежения за кодом и частотой, обеспечивает точными частотно–временными оценками, которые в дальнейшем непосредственно используются местным опорным генератором для согласования сжимающего сигнала с принятым расширяющим кодом. Однако для вхождения в синхронизм и этап слежения схемы слежения нуждаются в первоначальных целеуказаниях, например, в знании временного положения принятого сигнала с точностью до длительности одного чипа или около. Достижение данного условия, как уже указывалось ранее, является задачей этапа поиска, решение которой уменьшает начальную неопределенность сигнальных параметров до диапазона, требуемого схемой слежения. Сравнительно мягкие требования относительно точности оценивания на этапе поиска позволяют сократить количество вычисляемых статистик и упростить структуру их реализующую. Если возвратиться к условиям примера 8.1.1, то ослабление требований к необходимой точности временного измерения до длительности одного чипа означает в десять раз меньшее число корреляторов в схеме на рис. 2.18 или в десять раз меньшую скорость обработки в структуре с согласованным фильтром. В итоге, основной ресурсо–сберегающий технологией, используемой при поиске, является частичная или полная замена параллельных вычислений решающих статистик их последовательным выполнением.

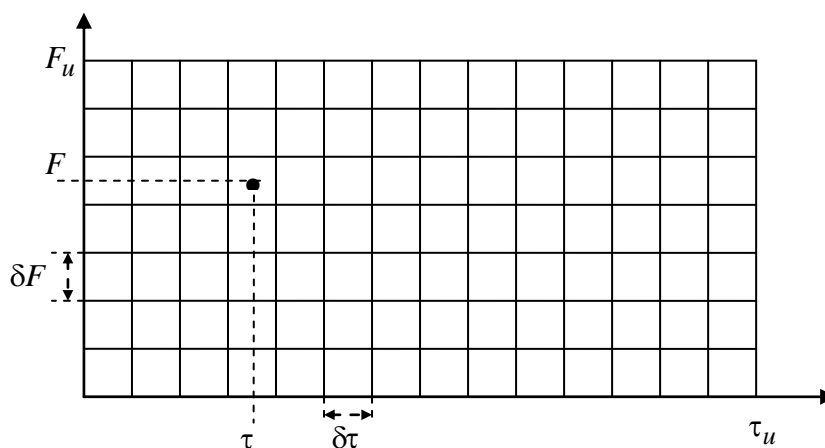


Рис. 8.1. Зона поиска и позиция сигнала на частотно–временной плоскости.

Для объяснения данного утверждения будем трактовать неизвестные задержку τ и частотный сдвиг F сигнала как сигнальные координаты на частотно–временной плоскости. Предположим, что диапазоны начальной неопределенности по параметрам τ и F составляют τ_u и F_u соответственно, и что в результате поиска они сокращаются до $\delta\tau$ и δF . Тогда, как показывает рис. 8.1, позиция сигнала находится в пределах одной из M прямоугольных ячеек (cells) размера $\delta\tau \times \delta F$, где $M = (F_u \tau_u) / (\delta F \delta \tau)$. Процедура поиска должна определить в какой из M ячеек содержится сигнал, т.е. осуществить проверку истинности M конкурирующих гипотез (см. параграф 2.8). Если бы использовалась оптимальная процедура указанной проверки, то параллельно были бы вычислены M корреля-

ций типа (2.74) для значений τ и F , которые отвечают центрам ячеек, и решение было бы принято в пользу тех τ и F , которым отвечало бы максимальное значение корреляции. На практике, однако, процедуры поиска используют факт длительного нахождения сигнала на входе приемника, что позволяет одновременно вычислять только некоторые (а не все M) корреляции. Если ни одна из них не является достаточно большой, то принимается решение, что среди тестируемых ячеек отсутствует истинная (т.е. содержащая сигнал), и поиск продолжается путем проверки другой группы ячеек. Данная процедура продолжается до тех пор, пока некоторая корреляция не будет признана достаточно большой для принятия решения об истинности соответствующей ей ячейки. Принятие решения завершает процедуру поиска, после чего начинают работать схемы слежения, используя в качестве начальных условий полученные оценки. Были осуществлены значительные исследования, касающиеся алгоритмов и стратегий поиска (см., например, библиографию в [77]). Ниже ограничимся только очень кратким обсуждением, начав с простейшей версии этой процедуры.

8.2. Процедура последовательного поиска.

8.2.1. Модель алгоритма.

При *последовательном поиске* в каждый момент времени анализируется только одна ячейка, т.е. вычисляется единственная корреляция наблюдения и локальной сигнальной копии, характеризующейся некоторыми конкретными значениями сдвига по времени и частоте. Затем величина корреляции анализируется с целью принятия решения об истинности или ложности данной ячейки. При принятии решения могут использоваться различные критерии. Например, поиск может продолжаться до тех пор, пока все ячейки внутри области неопределенности (см. рис. 8.1) не будут проанализированы, причем в течение анализа запоминаются максимальное значение корреляции на текущий момент времени и соответствующие ей значения τ и F . Затем, после анализа последней ячейки, автоматически становится известной ячейка, признанная истинной, поскольку ее координаты хранятся в памяти, а процедура поиска завершается простым их считыванием. Данная стратегия, хотя и реализующая правило МП оценивания, производит необходимые вычисления корреляций не одновременно, а последовательно во времени для следующих друг за другом сегментов принятого сигнала.

Более типичной для реальных приемников является другая версия последовательного поиска, когда найденное в текущий момент значение корреляции только сравнивается с порогом [6, 9, 77]. Если значение корреляции превышает порог, то принимается решение, что текущая ячейка является истинной и поиск заканчивается. В противном случае поисковая система анализирует следующую ячейку и так далее.

С точки зрения анализа выполнения неважно сколько параметров неизвестно и должно быть оценено в ходе поиска: одновременно время и частота (или еще что-нибудь), либо что-то одно из них. Существенным является только общее число анализируемых ячеек. Для большей прозрачности дальнейших обсуждений будем полагать, что процедура поиска заключается только в измерении временного запаздывания принятого сигнала, тогда как значение частоты а priori известно с достаточной точностью. На рис 8.2 представлена структура устройства, осуществляющего последовательный поиск в рассматриваемых условиях. В случае периодичности расширяющего кода максимальная область неопределенности по времени может простираться только на один период, а все большие задержки приводятся к одному периоду. В свете этого совершенно оправдано использование термина «фазы» как синонима задержки периодического кода [2, 6, 9], а переход от одной ячейки к следующей в области неопределенности означает просто изменение фазы местной копии кода. Если текущее значение корреляции ниже порога, то управляющая поиском логика предписывает местному генератору увеличить фазу кодовой копии

$s(t - \tau)$ на его выходе на один чип или часть чипа, а устройство переходит к анализу следующей ячейки. Если же текущая корреляция превышает порог, то логика управления сигнализирует об окончании поиска, а генератор сохраняет фазу кода, отвечающую ячейке, которая признана истинной. В следующих параграфах обсудим характеристики вышеупомянутого алгоритма, базирующиеся на анализе, приведенном в [78], а читателя, заин-

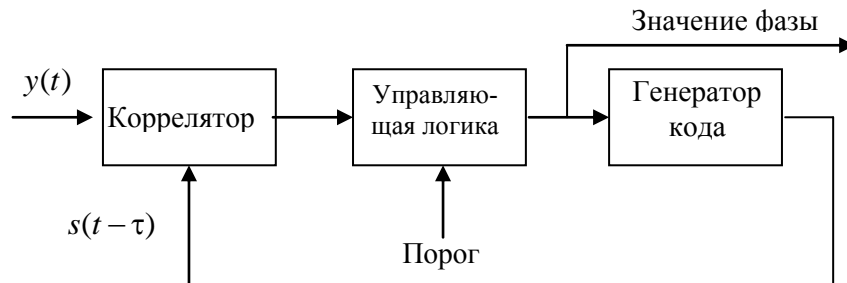


Рис. 8.2. Последовательный поиск фазы расширяющего кода.

тересованного в изучении подробностей, отсылаем к [6, 9, 77, 79-81].

8.2.2. Вероятность правильного завершения поиска и среднее число шагов.

Не нарушая общности, для упрощения анализа будем полагать, что местная копия кода сдвинута относительно принятого сигнала на целое число длительностей чипа. Тогда при периоде кода L всего возможно L значений фазы кода, одно из которых является истинным. В предположении различия всех фаз может существовать не более L ячеек поиска, а каждый переход от одной ячейки к следующей означает увеличение фазы кода на длительность одного чипа. В ходе поиска очень часто оказывается излишним анализ всех этих L ячеек, поскольку благодаря достоверной априорной информации возможно сокращение области поиска до M из общего числа L ячеек. Первоначально будем считать, что поиск стартует с наименее благоприятной ячейки области неопределенности, т.е. с наиболее удаленной от истинной ячейки. Иллюстрацией данного предположения служит рис. 8.3: первое достижение истинной ячейки (обозначенной закрашенным кружком) имеет место только после благополучного прохождения через $M - 1$ пустых ячеек (отмечен-

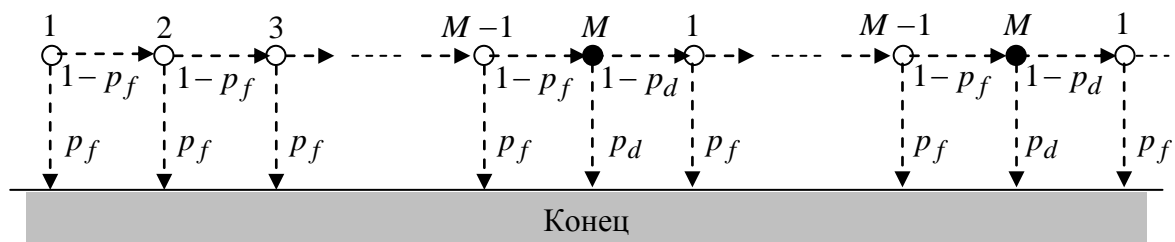


Рис. 8.3. Иллюстрация последовательного поиска кода в области неопределенности из M ячеек.

ных не закрашенным кружком).

Для достижения надежного отличия между уровнями корреляции в пустой и истинной ячейках система поиска в каждой ячейке должна вычислять корреляцию на достаточном временном интервале: *времени анализа (dwell time)*. При любом времени анализа и пороге решения об истинности или пустоте ячейки не могут быть вынесены абсолютно

безупречно. Одной из возможных ошибок является ложная тревога (см. параграф 3.2), т.е. принятие пустой ячейки за истинную. В этом случае процедура поиска завершится в ложном месте и совершенно естественно стремление удержать вероятность подобного события на достаточно низком уровне. С другой стороны, с ненулевой вероятностью анализ в истинной ячейке также может завершиться неправильным решением: пропуском сигнала и переходом в следующую (пустую) ячейку, которая уже была проанализирована ранее. Указанное событие делает процедуру поиска циклической: если поиск не заканчивается при первом прохождении области неопределенности, то производится второе прохождение и т.д., причем каждый из последовательных повторов инициирует новый цикл. Далее полагаем, что разыскиваемый сигнал постоянно присутствует на входе приемника, так что число возможных циклов не ограничено сверху. Из рис. 8.3 следует, что если система поиска безошибочно доходит до истинной ячейки, то возможны два варианта выхода: принятие правильного решения (и окончание поиска) с вероятностью обнаружения p_d или пропуск сигнала с вероятностью $1 - p_d$. Последнее событие влечет за собой только продолжение поиска на следующем цикле проверки и не имеет отрицательных последствий кроме увеличения времени до момента завершения поиска. Анализ пустой ячейки также может иметь два возможных исхода: правильное решение об отсутствии сигнала, сопровождаемое переходом в следующую ячейку с вероятностью $1 - p_f$, или признание ее истинной с вероятностью ложной тревоги p_f . В отличие от пропуска сигнала последствия от второй ошибки оказываются значительно катастрофичнее, поскольку неправильная фаза кода, найденная в результате поиска, означает, что все последующие операции, осуществляемые приемником, будут бесполезны. Для обеспечения низкой вероятности данного события каждое решение о нахождении фазы кода, как правило, подвергается дополнительной проверке за счет увеличения времени анализа в подозрительной ячейке (см. следующий параграф), но, тем не менее, вероятность окончания поиска неправильным решением остается ненулевой.

Назовем шагом поиска любой анализ в некоторой ячейке, завершающийся решением либо о продолжении, либо об окончании поиска. Из рис. 8.3 непосредственно видно, что когда поиск начинается с ячейки номер один (наиболее удаленной от истинной), он может закончиться в t -й ложной ячейке ($t = 1, 2, \dots, M - 1$) после $mM + t$ шагов и в истинной ячейке после $mM + M = (m + 1)M$ шагов, где $m = 0, 1, \dots$ – число полных «лишних» поисковых циклов, т.е. проходов по всей области неопределенности, предшествующих конечному $(m + 1 - \text{му})$ циклу, на котором поиск завершается либо в ложной, либо истинной ячейках. Тогда рассмотрение рис. 8.3, как диаграммы переходов, позволяет записать следующее выражение для вероятности $p(s)$ окончания поиска после s шагов:

$$p(s) = \begin{cases} p_f (1 - p_f)^{t-1} \left[(1 - p_d)(1 - p_f)^{M-1} \right]^m, & s = mM + t, \quad t = 1, 2, \dots, M - 1, \\ p_d (1 - p_f)^{M-1} \left[(1 - p_d)(1 - p_f)^{M-1} \right]^m, & s = mM + M, \end{cases} \quad (8.1)$$

где $m = 0, 1, \dots$. Все события (и только они), вероятность которых определяется второй строкой (8.1), подразумевают завершение поиска правильной оценкой фазы кода. Следовательно, полная вероятность P_{c1} (индекс «1» указывает, что поиск стартует с первой ячейки) правильного завершения поиска (правильного захвата) есть просто сумма всех этих вероятностей по всем возможным значениям m , т.е. сумма геометрической прогрессии с показателем $(1 - p_d)(1 - p_f)^{M-1}$:

$$P_{c1} = p_d (1 - p_f)^{M-1} \sum_{m=0}^{\infty} \left[(1 - p_d)(1 - p_f)^{M-1} \right]^m = \frac{p_d (1 - p_f)^{M-1}}{1 - (1 - p_d)(1 - p_f)^{M-1}}. \quad (8.2)$$

Другим важным параметром является среднее число шагов \bar{s}_1 поиска:

$$\bar{s}_1 = \sum_{s=1}^{\infty} sp(s) = \bar{m}M + \bar{t}, \quad (8.3)$$

где снова индекс «1» указывает на стартовую ячейку с номером один. В соотношении (8.3) величина \bar{m} отвечает среднему числу бесполезных циклов поиска, осуществленных до наступления конечного, тогда как \bar{t} обозначает среднее число шагов в последнем цикле, заканчивающемся либо в ложной, либо в истинной ячейке. Вероятность того, что отдельный цикл поиска окажется бесполезным, определяется как произведение вероятностей не окончания цикла во всех $M-1$ пустых ячейках и единственной истинной ячейке, т.е. $(1-p_d)(1-p_f)^{M-1}$. Следовательно, вероятность $p(m)$ прохождения до окончания поиска ровно m бесполезных циклов составит

$$p(m) = \left[1 - (1-p_d)(1-p_f)^{M-1}\right]^m, \quad m = 0, 1, \dots \quad (8.4)$$

Как видно, распределение вероятности $p(m)$ подчиняется геометрическому закону, математическое ожидание которого хорошо известно, однако, учитывая дальнейшие потребности, приведем полный вывод его с использованием производящей (генераторной) функции [14, 66]

$$g_m(z) = z^{\bar{m}} = \sum_{m=0}^{\infty} z^m p(m).$$

Производные от генераторной функции в точке $z=1$ позволяют вычислить моменты связанной случайной переменной. В частности

$$\left. \frac{dg_m(z)}{dz} \right|_{z=1} = \sum_{m=0}^{\infty} mp(m) = \bar{m}.$$

Для обобщенного геометрического распределения $p(l) = a^l(1-a), l = 0, 1, \dots; 0 < a < 1$, производящая функция получается в виде суммы геометрической прогрессии

$$g_l(z) = \sum_{l=0}^{\infty} z^l a^l (1-a) = \frac{1-a}{1-za},$$

так что после дифференцирования получаем

$$\bar{l} = \left. \frac{a(1-a)}{(1-za)^2} \right|_{z=1} = \frac{a}{1-a}. \quad (8.5)$$

Сравнение обобщенного геометрического закона с распределением (8.4) явно показывает, что подстановка $a = (1-p_d)(1-p_f)^{M-1}$ в (8.5) дает искомое выражение для математического ожидания \bar{m} :

$$\bar{m} = \frac{(1-p_d)(1-p_f)^{M-1}}{1 - (1-p_d)(1-p_f)^{M-1}}. \quad (8.6)$$

При нахождении второго члена \bar{t} в (8.3) отметим, что вероятность окончания финального цикла в пустой ячейке с номером t вне зависимости от числа предшествующих бесполезных циклов может быть найдена путем суммирования по m всех вероятностей, описываемых первой строкой в (8.1) (см. также рис. 8.3), тогда как вероятность достижения единственной (M -й) истинной ячейки и остановки в ней (снова вне зависимости от числа предшествующих бесполезных циклов) в точности определяется (8.2). Таким образом, распре-

деление вероятности $p(t)$ числа шагов t в финальном цикле будет

$$p(t) = \begin{cases} p_f(1-p_f)^{t-1} \sum_{m=0}^{\infty} [(1-p_d)(1-p_f)^{M-1}]^m = \frac{p_f(1-p_f)^{t-1}}{1-(1-p_d)(1-p_f)^{M-1}}, & t=1,2,\dots,M-1, \\ P_{c1} = \frac{p_d(1-p_f)^{M-1}}{1-(1-p_d)(1-p_f)^{M-1}}, & t=M. \end{cases}$$

Производящая функция данного распределения вероятностей имеет вид

$$\begin{aligned} g_t(z) &= \sum_{t=1}^M z^t p(t) = \frac{p_f}{1-(1-p_d)(1-p_f)^{M-1}} \sum_{t=1}^{M-1} z^t (1-p_f)^{t-1} + z^M P_{c1} = \\ &= \frac{p_f}{1-(1-p_d)(1-p_f)^{M-1}} \cdot \frac{z-z^M(1-p_f)^{M-1}}{1-z(1-p_f)} + z^M P_{c1}. \end{aligned}$$

Дифференцирование $g_t(z)$ в точке $z=1$ дает после несложных алгебраических преобразований следующее соотношение

$$\bar{t} = \frac{1-(1-p_f)^M - Mp_f(1-p_d)(1-p_f)^{M-1}}{p_f[1-(1-p_d)(1-p_f)^{M-1}]}. \quad (8.7)$$

Тогда, используя (8.6) и (8.7) в (8.3), окончательно получаем выражение для среднего числа шагов

$$\bar{s}_1 = \frac{1-(1-p_f)^M}{p_f[1-(1-p_d)(1-p_f)^{M-1}]}. \quad (8.8)$$

Отклоним теперь наше исходное предположение о стартовой ячейке и рассмотрим, каким образом повлияет на полученный результат начало поиска с произвольной ячейки номера r . В этом случае возникает частичный цикл (указателем которого служит индекс нуль), охватывающий $M-r$ пустых плюс одну истинную ячейку. В ходе этого цикла возможны следующие события: окончание поиска в ложной t -й ячейке с вероятностью $p_0(t|r) = p_f(1-p_f)^{t-r}$, $t=r, r+1, \dots, M-1$, завершение поиска в истинной ячейке с вероятностью $p_0(t=M|r) = p_d(1-p_f)^{M-r}$ и, наконец, пропуск сигнала и продолжение поиска с первой ячейки с вероятностью $P_{0m}(r) = (1-p_d)(1-p_f)^{M-r}$. Тогда полная вероятность правильного завершения поиска при старте с r -й ячейки будет

$$P_{cr} = p_0(t=M|r) + P_{0m}(r)P_{c1} = \frac{p_d(1-p_f)^{M-r}}{1-(1-p_d)(1-p_f)^{M-1}}. \quad (8.9)$$

Аналогичным образом пересчитывается и среднее число шагов. Если поиск завершается в t -й ячейке нулевого цикла, то число пройденных шагов составит величину $t-r+1$, но если сигнал пропущен, то к уже пройденным $M-r+1$ шагам в среднем дополнительно будет добавляться \bar{s}_1 шагов. Таким образом, при старте с произвольной (r -й) ячейки среднее число шагов составит величину

$$\bar{s}_r = \sum_{t=r}^M (t-r+1)p_0(t|r) + (M-r+1+\bar{s}_1)P_{0m}(r). \quad (8.10)$$

Первая сумма последнего выражения после замены индекса суммирования на $i = t - r + 1$ становится

$$p_f \sum_{i=1}^{M-r} i(1-p_f)^{i-1} + (M-r+1)p_0(t=M|r), \quad (8.11)$$

где первое слагаемое легко оценивается с помощью производящей функции, как это было выполнено при выводе (8.7)

$$\sum_{i=1}^{M-r} i(1-p_f)^{i-1} = \frac{1-(1-p_f)^{M-r+1} - (M-r+1)p_f(1-p_f)^{M-r}}{p_f^2}.$$

Использование последнего соотношения в (8.11), (8.10) совместно с равенством $p_0(t=M|r) + P_{0m}(r) = (1-p_f)^{M-r}$ после простых алгебраических преобразований дает

$$\bar{s}_r = \frac{1-(1-p_f)^{M-r+1}}{p_f} + \bar{s}_1 P_{0m}(r) = \frac{1-P_{cr}}{p_f} + \frac{P_{cr}}{p_d}. \quad (8.12)$$

Очевидно, что при подстановке $r=1$ соотношения (8.9) и (8.12) обращаются в (8.2) и (8.8) соответственно.

Теперь, зная априорное распределение вероятности $p_0(r)$ начальной ячейки с номером r , можно усреднить (8.9) и (8.12) по всем возможным начальным ячейкам из области поиска. Для равномерного априорного распределения $p_0(r) = 1/M$, $r=1, 2, \dots, M$ эта операция приводит к следующей полной усредненной вероятности \bar{P}_c правильного завершения поиска и среднему числу шагов \bar{s} поиска

$$\bar{P}_c = \frac{1}{M} \sum_{r=1}^M P_{cr} = \frac{p_d [1 - (1-p_f)^M]}{M p_f [1 - (1-p_d)(1-p_f)^{M-1}]}, \quad (8.13)$$

$$\bar{s} = \frac{1 - \bar{P}_c}{p_f} + \frac{\bar{P}_c}{p_d}. \quad (8.14)$$

8.2.3. Минимизация среднего времени поиска.

Вычисление и сравнение с порогом корреляции для каждого значения фазы кода означает анализ в течение некоторого конечного времени каждой анализируемой ячейки. В общем случае это время может быть случайным и, более того, зависит от истинности или ложности текущей ячейки. Однако в данный момент ограничимся рассмотрением только простейшей версии последовательного поиска, полагая постоянным время анализа T_d . Некоторые другие варианты будут кратко обсуждены в следующем параграфе. Тогда среднее время \bar{T}_s , затрачиваемое системой поиска, представляет собой просто произведение среднего числа шагов на время анализа: $\bar{T}_s = \bar{s} T_d$. Очевидно, что при фиксированной мощности сигнала более достоверное решение об истинности или ложности ячейки может быть вынесено лишь при увеличении времени анализа T_d , т.е. может быть обеспечено меньшее значение вероятностей ложной тревоги (p_f) и пропуска сигнала ($1-p_d$) в каждой ячейке.

Очевидно, что надежность поиска, характеристикой которой служит вероятность правильного завершения поиска \bar{P}_c , должна быть не хуже некоторой определенной вели-

чины. Как следует из (8.13), одно и то же значение \bar{P}_c может быть достигнуто при различных комбинациях вероятностей p_f, p_d для ячейки. Данный факт указывает на возможность минимизации среднего времени поиска за счет вариации одного параметров p_f или p_d при фиксировании постоянной величины вероятности \bar{P}_c правильного завершения поиска. Практическая возможность подобной оптимизации достаточно прозрачна. Предположим, что на вероятность правильного обнаружения накладываются жесткие ограничения, требующие близости ее к единице. Это означает, что поиск, почти несомненно, будет завершен за один единственный (начальный) цикл, однако для обеспечения высокой вероятности обнаружения анализ в каждой ячейке должен быть продолжительным, так что смещение в направлении к истинной ячейке будет медленным, а среднее время поиска \bar{T}_s – большим. В противоположность рассмотренному варианту, можно согласиться с высокой вероятностью пропуска сигнала с целью уменьшения T_d , однако это приведет к высокой вероятности повторения циклов, а, значит, увеличению среднего числа шагов в сравнении с предыдущим случаем, вследствие чего среднее время поиска \bar{T}_s снова может оказаться большим. Очевидно, что существует некоторый промежуточный оптимум вероятности обнаружения на ячейку p_d , минимизирующий величину \bar{T}_s .

Для решения задачи $\bar{T}_s = \min$ при $\bar{P}_c = \text{const}$ необходимо точно определить зависимость времени анализа от вероятностей p_f, p_d , т.е. эквивалентную модель канала. Полагая канал с АБГШ, вспомним, что модуль корреляции фактически представляет собой вещественную огибающую на выходе согласованного фильтра, которая для пустой ячейки отвечает огибающей гауссовского шума, а в случае истинной ячейки – огибающей смеси сигнала и шума. Хорошо известно, а также может быть найдено во многих учебниках по связи [например, 2, 4, 7. 8], что плотность распределения вероятности огибающей гауссовского шума подчиняется Релеевскому закону (см. параграф 3.2 или (3.12)), тогда как огибающая суммы сигнала и гауссовского шума имеет распределение Релея–Райса. В нормализованной форме, удобной в данном случае, последнее может быть представлено как

$$W(Y) = \begin{cases} Y \exp\left(-\frac{Y^2 + q_d^2}{2}\right) I_0(q_d Y), & Y \geq 0, \\ 0, & Y < 0, \end{cases} \quad (8.15)$$

где Y – значение огибающей, нормированное к среднеквадратическому значению шума, q_d – отношение сигнал-шум по напряжению, накопленному за время анализа T_d , а $I_0(\cdot)$ – модифицированная функция Бесселя нулевого порядка первого рода. Очевидно, что подстановка $q_d = 0$, означающая отсутствие сигнала, превращает (8.15) в распределение Релея (отметим, что $I_0(0) = 1$).

Напомним, что решение о содержании ячейки выносится на основе сравнения Y с порогом. Если порог, нормированный к среднеквадратическому значению шума, есть Y_t , то решение об истинности ячейки принимается всякий раз, когда $Y \geq Y_t$. Тогда для вероятностей p_f, p_d можно записать следующие соотношения

$$p_f = \int_{Y_t}^{\infty} W(Y|H_0) dY, \quad p_d = \int_{Y_t}^{\infty} W(Y|H_1) dY,$$

где плотность вероятности $W(Y|H_0)$ и $W(Y|H_1)$ представляют собой варианты (8.15) для

гипотез H_0 (пустая ячейка, $q_d = 0$) и H_1 (истинная ячейка, $q_d > 0$) соответственно. Тогда вышеприведенные интегралы сводятся к виду

$$p_f = \exp\left(-\frac{Y_t^2}{2}\right), \quad p_d = Q_M(q_d, Y_t), \quad (8.16)$$

где $Q_M(\cdot, \cdot)$ есть просто обозначение интеграла от распределения Релея–Райса, называемое также Q – функцией Маркума [7, 8].

Решение первого из соотношений (8.16) относительно Y_t при заданной p_f дает величину порога, необходимую для удержания требуемого уровня вероятности ложной тревоги: $Y_t = \sqrt{2 \ln(1/p_f)}$. Подстановка его во второе соотношение (8.16) непосредственным образом связывает величины p_d и p_f для заданного отношения сигнал-шум q_d :

$$p_d = Q_M\left(q_d, \sqrt{2 \ln \frac{1}{p_f}}\right). \quad (8.17)$$

В свою очередь, отношение сигнал-шум, обеспечиваемое за время анализа T_d , определяется, как и ранее (см. параграф 3.2): $q_d = \sqrt{2PT_d/N_0}$, где P – мощность сигнала, а N_0 – спектральная плотность мощности одностороннего белого шума. Теперь, пусть $T_d(p_f, p_d)$ и $q_d(p_f, p_d)$ – время анализа и отношение сигнал-шум, необходимые для обеспечения фиксированных вероятностей p_d и p_f . Тогда

$$T_d(p_f, p_d) = \frac{q_d^2(p_f, p_d)}{2P/N_0}. \quad (8.18)$$

В итоге, согласно (8.14), среднее время поиска будет

$$\bar{T}_s = \bar{s}T_d = \left(\frac{1-\bar{P}_c}{p_f} + \frac{\bar{P}_c}{p_d}\right)T_d(p_f, p_d)$$

или в нормированном виде

$$\bar{T}_s\left(\frac{2P}{N_0}\right) = \left(\frac{1-\bar{P}_c}{p_f} + \frac{\bar{P}_c}{p_d}\right)q_d^2(p_f, p_d). \quad (8.19)$$

Теперь процедура оптимизации становится очевидной. Пусть заданы вероятность правильного завершения поиска \bar{P}_c и размер диапазона поиска M .

1. Выберем некоторое значение вероятности ложной тревоги на ячейку p_f из диапазона $0 < p_f < 2(1-\bar{P}_c)/M$.

2. Решение уравнения (8.13) относительно неизвестной p_d дает ее величину, обеспечивающую требуемые \bar{P}_c и p_f :

$$p_d = \frac{M\bar{P}_c p_f [1 - (1-p_f)^{M-1}]}{1 - (1-p_f)^{M-1} [1 + p_f(M\bar{P}_c - 1)]}.$$

3. Решение (8.17) относительно неизвестного q_d позволяет найти его значение, га-

рантирующее вероятности p_d и p_f .

4. Подставляем найденное значение в (8.19).

5. Для различных значений p_f строим зависимости среднего времени поиска от p_d и определяем пару величин p_d и p_f , доставляющих минимум \bar{T}_s .

На рис. 8.4 представлены зависимости нормированного времени поиска $\bar{T}_s \left(\frac{2P}{N_0} \right) M^{-1}$ от вероятности обнаружения на ячейку p_d для двух значений вероятности успешного завершения поиска $\bar{P}_c = 0,99; 0,999$ и четырех значений ширины зоны неопределенности $M = 50, 200, 10^3$ и 10^4 . Как видно, минимум среднего времени поиска прихо-

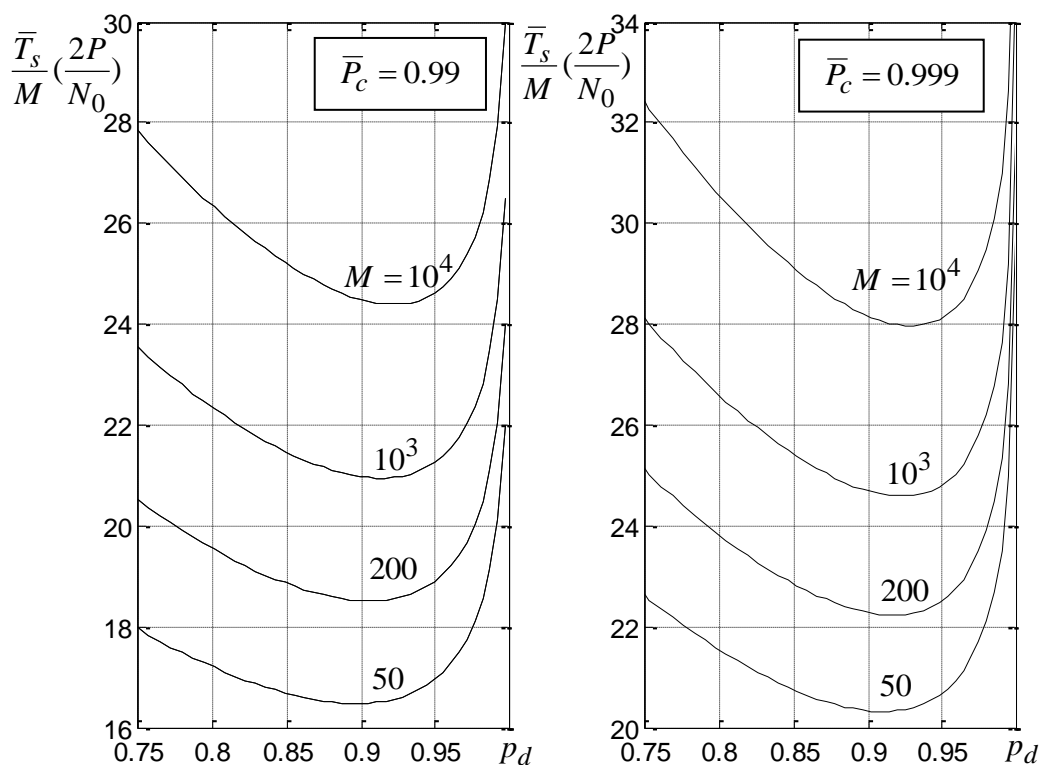


Рис.8.4. Кривые нормированного среднего времени поиска от вероятности обнаружения.

дится на интервал значений $p_d \in [0.9, 0.95]$. Учитывая достаточно плоский характер кривых в районе минимума, величина $p_d = 0.9$ может быть принята в качестве универсального значения оптимальной вероятности обнаружения на ячейку не зависимо от диапазона M и вероятности правильного поиска \bar{P}_c .

8.3. Методы ускорения поиска.

8.3.1. Постановка задачи.

Естественно, что среднее время завершения поиска растет с увеличением числа тестируемых ячеек M , т.е. с расширением области неопределенности при сохранении всех остальных параметров неизменными. Тогда становится очевидной тенденция к уменьшению M в случае возможности обеспечения системы поиска априорной информацией о фазе кода, частоте сигнала или других неизвестных параметрах. Например, в GPS Navstar каждый из 24 спутников передает данные о текущем и предсказуемом положении всего космического созвездия, которое запоминается в приемнике пользователя. Благодаря этой информации пользователь после захвата сигнала любого спутника и зная его приблизительное положение, может заранее вычислить с некоторой точностью фазы кодов других видимых спутников, существенно сократив область неопределенности их поиска. Нечто подобное имеет место и в системе мобильного телефона cdmaOne (IS-95), в которой жесткая синхронизация всех базовых станций способствует поиску сигналов новой БС пользовательским приемником в ходе эстафетной передачи.

Однако неизбежны сценарии, когда область неопределенности настолько обширна, что может потребовать недопустимо большого времени завершения поиска при проведении обычного последовательного поиска, рассмотренного выше. Среди прочих подобная ситуация характерна для этапа инициализации приемника (его первого включения), когда собственный эталонный генератор имеет произвольный сдвиг по отношению к системному времени, а для уменьшения области поиска не могут быть использованы априорные сведения. Проиллюстрируем указанную ситуацию с помощью примера.

Пример 8.3.1. Рассмотрим процедуру поиска фазы кода длины $L = 2^{15}$ («короткого» кода системы cdmaOne) в отсутствие априорной информации, позволяющей уменьшить область неопределенности. Полагая $M = L = 2^{15}$ и вероятность правильного окончания не менее чем 0.99, из экстраполяции рис.8.4 можно увидеть, что $\bar{T}_s \geq 25 \cdot 2^{15} / (2P/N_0)$. Если принять в качестве соответствующего значение $2P/N_0 = 40$ дБ·Гц, то среднее время поиска превысит 80 с. Однако, эта оценка может оказаться слишком оптимистичной, поскольку адекватна гипотетической процедуре поиска с гарантированной синхронизацией чипов, допускающей в ходе поиска увеличение фазы кода на один чип (см. предположение в начале параграфа 8.2.2). На практике подобная чиповая синхронизация часто не достижима, и смещение на один чип грозит попаданием на склон основного лепестка АКФ сигнала, а не на его пик (см., например, рис. 6.9), что увеличивает вероятность пропуска истинной ячейки. Для избежания подобных проблем необходимо использовать более мелкий шаг приращения, как правило, равный половине длительности чипа, тем самым, увеличивая число ячеек M в области неопределенности, а значит, и время захвата. □

Проблема временных затрат может стать приоритетной в приложениях, использующих очень длинные широкополосные коды, таких как, системы измерения дальности и слежения за удаленными космическими объектами. Опишем кратко, не прибегая к математическим тонкостям, основные методы ускорения операции поиска.

8.3.2. Последовательный просмотр ячеек.

Стратегия постоянного времени анализа вне зависимости от истинности или ложности ячейки, которая была принята ранее, может быть скорректирована, учитывая тот факт, что большинство ячеек являются пустыми. В самом деле, потенциальный метод, предполагающий быстрое распознавание пустой ячейки в обмен на продолжительное рас-

смотрение истинной, интуитивно должен привести к экономии времени поиска. Такие методы действительно существуют и объединяются под общим названием *последовательного анализа* (*sequential analysis*). Простейшей последовательной процедурой является двухэтапная процедура анализа [9,78,82]. Ее основная идея состоит в разделении анализа на два этапа. На первом из них сравнительно низкий порог даже при достаточно малом времени анализа T_{d1} обеспечивает низкую вероятность пропуска сигнала. В тоже время вероятность ложной тревоги оказывается много большей той, которую следовало бы признать допустимой в предшествующем (одноэтапном) методе. Благодаря короткому времени анализа T_{d1} ложные ячейки в среднем просматриваются быстро, однако значительная их часть (вплоть до 10% и даже больше) считаются истинными. Для отсеивания ложных ячеек, которые на первом этапе были ошибочно признаны в качестве истинных, применяется второй этап, который характеризуется гораздо большей достоверностью по сравнению с первым. Последнее достигается выбором соответствующих параметров: более длительным временем анализа T_{d2} и высоким порогом, гарантирующих полную (включая также и первый этап) вероятность ошибки на ячейку, которая удовлетворяет требуемой вероятности правильного завершения поиска (8.13). Если этапы осуществляются независимо, т.е. корреляция на втором этапе вычисляется без учета результатов, полученных на первом этапе, то полные вероятности ложной тревоги и правильного обнаружения на ячейку определяются, как $p_f = p_{f1}p_{f2}$ и $p_d = p_{d1}p_{d2}$, где второй индекс отвечает номеру этапа. Теперь полное время анализа на ячейку оказывается случайным, поскольку просмотр любой ячейки заканчивается либо на первом этапе, либо с некоторой вероятностью передается второму этапу. Вероятность ложной тревоги на первом этапе p_{f1} , будучи во много раз выше полной вероятности p_f , все же значительно меньше вероятности обнаружения p_{d1} . В результате среднее время анализа, приходящееся на ложную ячейку $\bar{T}_{df} = T_{d1} + p_{f1}T_{d2}$ меньше аналогичной характеристики для истинной ячейки $\bar{T}_{dt} = T_{d1} + p_{d1}T_{d2}$, что благоприятствует сокращению времени завершения поиска. При оптимальном выборе p_{d1}, p_{d2} среднее время поиска может быть уменьшено в два раза или более по сравнению со случаем поиска с фиксированным временем анализа [9,78].

Дальнейшее усовершенствование возможно на пути увеличения числа этапов, когда каждый из них проверяет решения, принятые на предыдущем [9,78,83-86]. Предельным вариантом данной стратегии множества этапов служит последовательный анализ Вальда [87], в котором попытки решения совершаются непрерывно по мере обработки каждого последующего чипа. В данном варианте используются два порога, и ячейка признается пустой, как только накопленное значение корреляции упадет ниже меньшего из порогов, тогда как решение об истинности ячейки принимается в случае превышения верхнего порога. Если решающая статистика (величина корреляции) остается между порогами, анализ ячейки продолжается путем дальнейшего интегрирования большего числа чипов [77,78].

Вследствие случайности времени анализа в ячейке вычисление времени завершения поиска в многоэтапных (или в общем случае последовательных) стратегиях поиска может показаться более сложным по сравнению со случаем фиксированного времени анализа. Эффективный путь упрощения данной задачи указан в [78], где показано, что среднее время поиска, начинаемого с наименее благоприятной ячейки, может быть найдено как произведение среднего числа циклов на среднюю длительность одного цикла.

8.3.3. Последовательно–параллельный поиск.

Очевидным ресурсом ускорения поиска является применение нескольких параллельных корреляторов, каждый из которых работает автономно и сканирует отдельную

часть области неопределенности. В этом случае исходная область неопределенности просто разбивается на n_c подобластей, каждая из которых охватывает M/n_c ячеек, где n_c – число параллельных каналов, и время поиска соответственно уменьшается в n_c раз. В предельном случае, когда $n_c = M$, поиск становится полностью параллельным и не требует последовательных шагов. Данная возможность находит широкое применение в реальном оборудовании и особенно эффективна при использовании дискретных компонентов, которые обязательно присутствуют в приемнике, но, с другой стороны, обычно свободны во время процедуры поиска. Например, любой современный GPS приемник содержит множество корреляционных каналов, необходимых для параллельного отслеживания сигналов всех (или, по крайней мере, четырех) видимых спутников. Во время поисковой процедуры эти каналы свободны от другой нагрузки и могут быть использованы для поиска сигналов.

8.3.4. Последовательности быстрого поиска.

Возможны ситуации, когда ускорение поиска приобретает первостепенную значимость. Представим, например, систему позиционирования объектов в удаленном пространстве. С целью исключений неоднозначности измерения расстояния в очень широком диапазоне, оцениваемом может быть сотнями тысяч километров или более, необходимо использовать широкополосный сигнал соответственно большого периода (например, сотни тысяч или миллионы чипов). Излишне говорить, что традиционные стратегии поиска, рассмотренные выше, окажутся чрезмерно медленными, если только число корреляторов не достигает многих сотен. В подобных ситуациях эффективным решением могут служить специальные кодовые последовательности, оптимизированные согласно критерию минимума времени поиска.

Почему последовательный поиск оказывается слишком медленным при использовании сигналов с хорошей АКФ (т.е. имеющей малые боковые лепестки)? Ответ очевиден: просмотр и отбраковка любой текущей пустой ячейки сокращает область неопределенности только на одну ячейку, тогда как необходимо осуществить не менее чем M проверок, чтобы пройти всю зону, стартуя с наименее благоприятной ячейки. Закономерно возникает другой вопрос: возможно ли построить последовательность, которая позволила бы сократить наполовину первоначальную зону неопределенности после вычисления одной корреляции вместо отбрасывания только одной ячейки? Стиффлер нашел строгое решение данной задачи [86], хотя некоторые эффективные коды, сокращающие время поиска, были предложены ранее [87].

Последовательности быстрого поиска (rapid acquisition sequences) Стиффлера имеют четкую структуру, представляя собой просто сумму строго синхронизированных n компонент. Первым из них является последовательность чипов чередующейся полярности (...+-+--+--+...), т.е. имеет период $L_1 = 2$ чипам. Вторым является меандр (...++---+---...) с периодом $L_2 = 4$ чипам и т.д. вплоть до n -го меандра периода $L_n = 2^n$ чипов. Можно трактовать k -й компонент, как последовательность «длинных» чипов чередующейся полярности длительности $2^{k-1}\Delta$, где Δ – длительность чипа первого меандра. Поиск с помощью одного коррелятора начинается с определения фазы первого компонента. При условии существования чиповой синхронизации возможны только два ее значения, и вычисление корреляции между наблюдаемым колебанием и местной опорой, которая представляет собой копию первого меандра, устраняет эту неопределенность: значение корреляции положительно, если опорный образец находится в фазе с принятым первым меандром, и отрицательно, если они противоположны. После завершения этого этапа осуществляется чиповая синхронизация со вторым принятым меандром (при этом длительность чипов равна 2Δ), и снова возможны только два значения фазы второго ме-

андра. Данная неопределенность разрешается аналогичным образом, т.е. вычислением корреляции с опорным образцом, являющимся копией второго меандра, и т.д. до установления на n -м этапе фазы «самого медленного» (n -го) меандра в результате вычисления корреляции с местным образцом. Таким образом, все процедура поиска состоит только из n этапов, каждый из которых уменьшает в два раза первоначальную зону неопределенности $M = 2^n$.

Процедура разделения зоны неопределенности на две равные половины называется дихотомией. Последовательности быстрого захвата Стиффлера наилучшим образом согласуются с этой процедурой.

Прямое обобщение этой идеи на случай поискового оборудования с n_c корреляторами состоит в замене меандров на n компонентов периода $L_1 = n_c + 1$ чипов, в которых чип каждого компонента в L_1 раз длиннее предшествующего. Полный период суммарной последовательности составляет $L = L_1^n$ первоначальных чипов. Каждый компонент должен иметь по возможности хорошую периодическую АКФ, и в классе бинарных компонентов наилучшими являются минимаксные последовательности (см. параграфы 6.7, 6.9). На первом шаге неопределенность из L_1 возможных фаз «наиболее быстрого» компонента разрешается после параллельного вычисления корреляции принятого сигнала с n_c локальными опорными образцами, представляющими собой $L_1 - 1$ различных циклических копий первого компонента со сдвигом L_1 . Если среди n_c корреляций некоторые являются положительными, то опора, соответствующая максимальному значению корреляции, определяет фазу, которая принимается в качестве синхронизированной с первым компонентом принятого сигнала. В противном случае в качестве истинной принимается L_1 циклическая копия. После данного этапа становятся известными границы чипов второго компонента, и осуществляется следующий этап, повторяющий те же операции, что и на предыдущем, но со вторым компонентом и т.д. Процедура завершается после n -го шага, на котором неопределенность по фазе наиболее медленного компонента разрешается аналогично предшествующим этапам. Как видно, каждый шаг в этом случае сопровождается проверкой $L_1 = n_c + 1$ гипотез и сокращает в L_1 раз первоначальную область неопределенности $M = L = L_1^n$. Опять, только что описанные последовательности в наибольшей степени пригодны для данного L_1 -альтернативного тестирования: для приемника любой сложности (числа параллельных корреляторов n_c) существует единственный код, который минимизирует время поиска. Вследствие этого подобные коды принято называть согласованными [70]. Последовательности быстрого поиска Стиффлера являются одними из них, отвечающими приемнику с единственным коррелятором.

Численные оценки показывают, что выигрыш во времени поиска, сопровождающий использование данных кодов, может составить более сотни раз [70, 86] (см. также задачу 8.14).

Обычная сумма бинарных компонентов чаще является многоуровневой, чем бинарной последовательностью, что влечет амплитудную модуляцию сигнала и часто рассматривается нежелательной. Ограничение суммы и удержание только ее знака превращает согласованный код в бинарный, сохраняя процедуру поиска неизменной. Единственной платой за данную трансформацию является незначительное (в пределах 1.5–2 раз) сокращение выигрыша во времени поиска.

Отметим, что двухэтапный синхронизирующий код стандарта UMTS (см. параграф 11.3) служит примером подобного подхода: на первом этапе ищется первичный код, период которого в 15 раз меньше вторичного. Тогда второй этап поиска решает задачу неопределенности о том, какая из $M = 15$ фаз вторичного кода является истинной.

8.4. Слежение за кодом.

8.4.1. Оценка запаздывания в результате слежения.

В беспроводных приемниках, как правило, используется замкнутая петля слежения в качестве инструмента непрерывного и точного измерения параметров. В зависимости от характера измеряемого параметра ее примерами являются автоматическая регулировка частоты, петля фазового слежения, автоматическая регулировка усиления и др. Строгое математическое обоснование, доказывающее оптимальность петель слежения для изменяющихся во времени параметров сигнала, базируется на теории нелинейного оценивания [88], однако их основная идея непосредственно следует из правила максимального правдоподобия (МП), рассмотренного в главе 2. Специфический характер приемника широкополосного сигнала проявляет себя главным образом в операции снятия расширения (см. параграф 7.1), требующей точного синхронизма местной сужающей опоры с принятым сигналом. Учитывая данный факт, сосредоточим основное внимание на точном измерении задержки (фазы кода) принимаемого сигнала.

Для быстрой понимания упростим задачу, сведя ее к оцениванию задержки модулирующего сигнала, не учитывая влияние случайности фазы. Предположим, что τ является неизвестной задержкой модулирующего сигнала $s(t)$. Тогда в силу правила МП (2.55) оптимальный измеритель должен формировать оценку $\hat{\tau}$ этого параметра как его значение, максимизирующее корреляцию $z(\tau)$ между опорной сигнальной копией $s(t - \tau)$ и наблюдением $y(t)$. Одним из путей аппаратной реализации данного алгоритма служит набор корреляторов (см. рис. 2.18), непосредственно вычисляющих функцию $z(\tau)$ в M выборочных точках; другим является структура с согласованным фильтром, воспроизводящая $z(\tau)$ в реальном времени (рис. 2.19, где детектор огибающей излишен для модулирующего сигнала). Однако обе эти схемы могут оказаться неосуществимыми в случае применения широкополосного сигнала большой длины: первая из них вследствие множества корреляторов, а вторая – в связи со сложностью реализации согласованного фильтра (см. пример 8.1.1).

Структура со следящей петлей представляет собой еще одну альтернативу. Отметим, что в точке максимума $z(\tau)$ ее производная обнуляется:

$$z'(\hat{\tau}) = \left. \frac{dz(\tau)}{d\tau} \right|_{\tau=\hat{\tau}} = - \int_0^T y(t) s'(t - \tau) dt = 0. \quad (8.20)$$

Назовем $e(\hat{\tau}) = z'(\hat{\tau})$ *сигналом ошибки (error signal)*, причины для которого вскоре станут ясны. Как видно, можно осуществлять поиск такого $\hat{\tau}$, который обратит сигнал ошибки в нуль. Предположим, что истинная задержка сигнала равна τ , предварительной оценкой является $\hat{\tau}$ и вычислим среднее сигнала ошибки по всем шумовым реализациям в $y(t)$:

$$\overline{e(\hat{\tau})} = \overline{z'(\hat{\tau})} = - \int_0^T \overline{y(t) s'(t - \hat{\tau})} dt = - \int_0^T \overline{s(t - \tau) s'(t - \hat{\tau})} dt = \int_0^T \overline{s'(t - \tau) s(t - \hat{\tau})} dt. \quad (8.21)$$

Последнее равенство следует после интегрирования по частям при учете, что интервал интегрирования простирается на все «тело» сигнала вне зависимости от τ : $s(-\tau) = s(T - \tau) = 0$. При аналогичных условиях среднее сигнала ошибки при $\hat{\tau} = \tau$ (предварительная оценка совпадает с истинным значением параметра) равняется нулю:

$$\overline{e(\hat{\tau} = \tau)} = - \int_0^T \overline{s(t - \tau) s'(t - \tau)} dt = - \left. \frac{[s'(t - \tau)]^2}{2} \right|_0^T = 0,$$

однако в той же точке согласно последней части (8.21) производная от $\overline{e(\hat{\tau})}$

$$\left. \frac{de(\hat{\tau})}{d\hat{\tau}} \right|_{\hat{\tau}=\tau} = z''(\tau) = -\int_0^T [s'(t-\tau)]^2 dt$$

равна энергии производной сигнала со знаком минус, т.е. отрицательна. Последнее означает, что если $\hat{\tau}$ лежит в достаточной близости слева от τ , то $e(\hat{\tau})$ положительно, если же $\hat{\tau} > \tau$, то $e(\hat{\tau})$ отрицательно. Данное положение подсказывает структуру *петли захвата по задержке (delay-lock loop (DLL))*, решающую соотношение (8.20) посредством ряда последовательных итераций и представленную на рис. 8.5. Местный генератор опоры формирует сдвинутую по времени копию производной сигнала $s'(t - \hat{\tau})$, которая используется

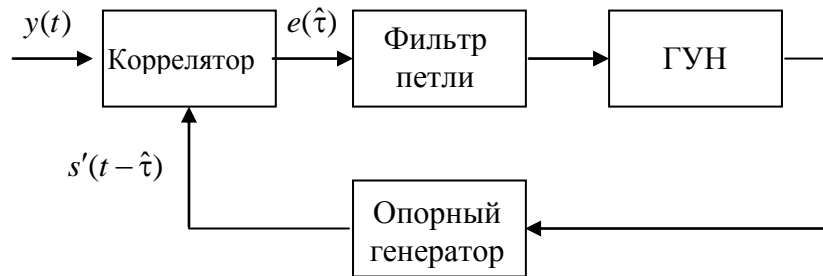


Рис. 8.5. Обобщенная структура DLL.

для вычисления корреляции с наблюдением $y(t)$. Затем результирующий сигнал ошибки очищается от шумов посредством фильтра петли для приближения к среднему в (8.21). Когда сглаженный сигнал ошибки положителен, то это с большой вероятностью свидетельствует об опережении сигнала локальной опорой $s'(t - \hat{\tau})$ и вынуждает *генератор управляемый напряжением (ГУН) (voltage controlled oscillator (VCO))* снизить свою частоту, т.е. увеличить задержку (оценочную) опоры $\hat{\tau}$. В противном случае сглаженный отрицательный сигнал ошибки приводит ГУН¹ к необходимости увеличения частоты, т.е. уменьшает задержку опоры. Очевидно, что в устойчивом состоянии DLL удерживает сигнал ошибки в районе нуля, обеспечивая синхронизм между локальной опорой и принятым сигналом.

Очевидно, что для адекватной работы DLL нуждается в начальном целеуказании, т.е. стартовом значении $\hat{\tau}$, которое достаточно близко к истинной задержке сигнала. С одной стороны, данное условие накладывает определенные требования к точности процедуры поиска. С другой стороны, опорное колебание, получаемое на основе МП правила, обычно не осуществимо (например, оно может содержать дельта-функции, если сигнал состоит из прямоугольных чипов). Как правило, оно заменяется некоторым другим сигналом, который, будучи более подходящим для практической реализации, сохраняет основные свойства: отчетливую нечетную зависимость сигнала ошибки на выходе коррелятора от ошибки оценивания $\hat{\tau} - \tau$. При конструировании подобной опоры решающую роль может играть желание обеспечить втягивание (достижение синхронизма) при более мягких требованиях к начальному целеуказанию. Более подробно это будет рассмотрено в 8.4.3.

8.4.2. Ранне-поздние дискриминаторы DLL.

Первым элементом в структуре DLL является дискриминатор, т.е. комбинация из

¹ В реальном исполнении входная величина, управляющая частотой генератора, необязательно является напряжением, например, в цифровой реализации петли входное число играет аналогичную роль. Однако, для избежания излишней множественности терминов будем использовать название ГУН в качестве универсального.

коррелятора и опоры, формирующая сигнал ошибки $e(\hat{\tau})$. В одной из классических схем дискриминатора DLL для модулирующего сигнала $s(t)$ в качестве опоры $s_r(t)$ берется разность двух сдвинутых по времени копий сигнала: запаздывающей $s(t-\delta/2)$ и опережающей $s(t+\delta/2)$, где δ определяет их временное разнесение. Тогда полезный компонент сигнала ошибки, обусловленный рассогласованием $\varepsilon = \hat{\tau} - \tau$ опоры и принятого сигнала, может быть записан в виде

$$\begin{aligned} \overline{e(\hat{\tau})} &= A \int_0^T s(t-\tau) s_r(t-\hat{\tau}) dt = A \int_0^T s(t-\tau) \left[s\left(t-\hat{\tau}-\frac{\delta}{2}\right) - s\left(t-\hat{\tau}+\frac{\delta}{2}\right) \right] dt = \\ &= AE[\rho(\varepsilon+\delta/2) - \rho(\varepsilon-\delta/2)], \end{aligned} \quad (8.22)$$

где E – энергия стандартного сигнала $s(t)$ за время интегрирования T , A – амплитуда принятого сигнала, взвешенная относительно $s(t)$, а $\rho(\tau)$ – нормированная АКФ сигнала, вычисленная на временном интервале $[0, T]$. Возьмем, например, периодический дискретный сигнал (широкополосный код) длины N с длительностью чипов Δ , и пусть время интегрирования простирается на целое число l периодов сигнала: $T = lN\Delta$, тогда $\rho(\tau)$ – периодическая АКФ. Если $s(t)$ является минимаксной бинарной последовательностью (m -последовательностью, последовательностью Лежандра и т.п.) с прямоугольными чипами, то ее нормированная периодическая АКФ $\rho(\tau)$ имеет вид, представленный на рис. 8.6, *a*. Опережающая и запаздывающая копии $\rho(\tau)$, входящие в соотношение (8.22), показаны на

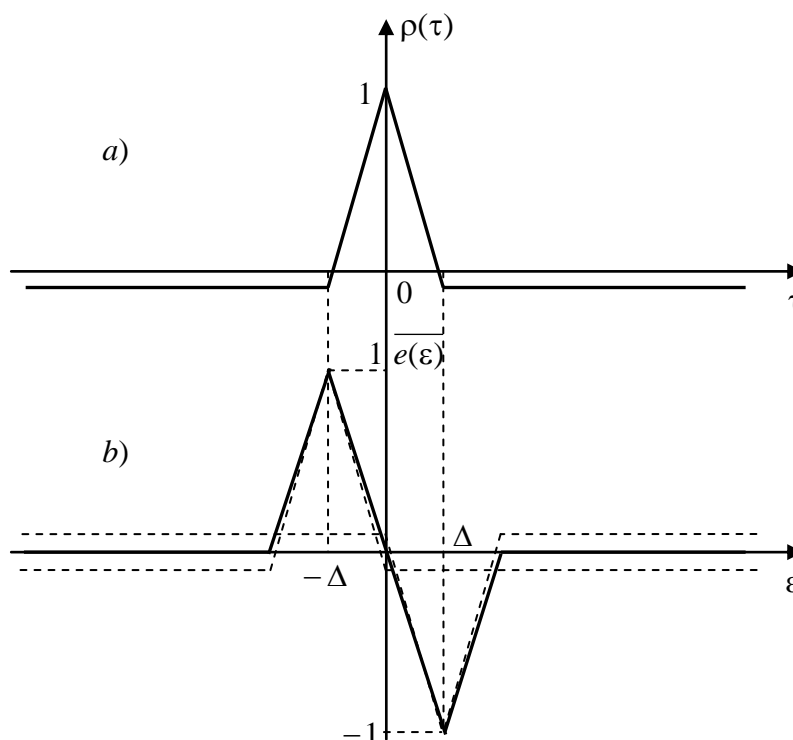


Рис.8.6. Дискриминационная кривая ранне-поздней DLL.

рис. 8.6, *b* пунктирными линиями для примерного разнесения $\delta = 2\Delta$, а их разность $e(\varepsilon) = e(\hat{\tau} = \tau + \varepsilon)$, называемая *дискриминационной кривой*, также представлена на этом рисунке сплошной линией. Легко заметить, что рассматриваемый дискриминатор полностью отвечает выдвинутым требованиям: опережающая и запаздывающая опорные версии входного сигнала соответственно порождают положительный или отрицательный сигнал

ошибки, предоставляя возможность ГУНу изменить собственную частоту и, тем самым, сдвинуть опору в соответствующем направлении.

На рис. 8.7 представлена одна из возможных структур ранне–позднего дискриминатора для модулирующего сигнала. Кодовый генератор, управляемый ГУНом, формирует опережающую сигнальную копию $s(t + \delta/2)$, которая, задержанная на δ , дает запаздывающую копию. Их разность подается на коррелятор в качестве опорного сигнала. Кодовая копия $s(t)$, синхронизированная с входным сигналом и используемая для снятия расширения спектра, может быть получена в результате задержки на $\delta/2$ опережающей ко-

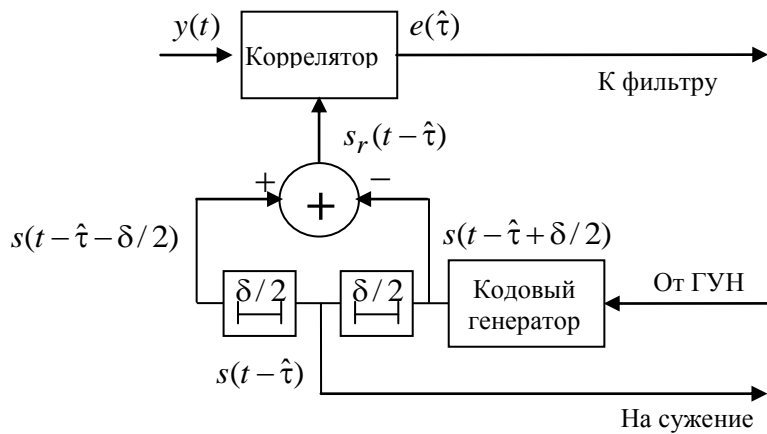


Рис.8.7. Ранне–поздний дискриминатор модулирующего сигнала

пии. Если кодовый генератор построен на основе регистра сдвига и $\delta = 2\Delta$, то отсутствует необходимость во введении дополнительных элементов задержки, поскольку все три копии могут быть считаны с трех последовательных разрядов регистра.

Альтернативная схема такого же дискриминатора включает два коррелятора, которые первоначально отдельно вычисляют корреляции наблюдения с опережающей и запаздывающей копиями сигнала, а затем осуществляется вычитание полученных результатов с целью получения сигнала ошибки $e(\hat{\tau})$ [18, 77].

В случае полосного сигнала рассмотренный дискриминатор может быть применен, если петля восстановления фазы приемника предварительно синхронизирована с несущей частотой принимаемого сигнала, так что полосный сигнал может быть трансформирован в его действительный низкочастотный эквивалент. По этой причине данный дискриминатор часто называют *когерентным*, также как и содержащую его схему DLL. Если предварительная фазовая синхронизация не возможна, то используется *некогерентный* дискриминатор, основанный на сравнении квадратов модулей двух корреляций. Они вычисляются между наблюдаемой комплексной огибающей $\dot{Y}(t)$ и комплексными огибающими двух опор, формируемых в виде опережающей и запаздывающей копий кода сигнала. Прежде всего, данный дискриминатор оказывается эффективным несмотря на наличие модуляции данными. На практике он часто реализуется структурой, изображенной на рис. 8.8, где умножение комплексных огибающих осуществляется посредством гетеродинирования (см. комментарии в конце параграфа 7.1.2). Рассмотрим работу данной структуры, игнорируя входной шум и, не нарушая общности, полагая $\tau = 0$. Пусть опоры представляют собой сдвинутые во времени копии расширяющего сигнала, обладающего комплексной огибающей $\dot{S}(t)$ и несущей частотой f_1 , которая отличается от несущей принятого сигнала f_0 :

$$s_r\left(t - \varepsilon \pm \frac{\delta}{2}\right) = \operatorname{Re}\left[\dot{S}\left(t - \varepsilon \pm \frac{\delta}{2}\right) \exp(j2\pi f_1 t + j\vartheta)\right],$$

где начальная фаза ϑ в окончательном результате не будет играть никакой роли. Комплексная огибающая приходящего сигнала $A\dot{B}(t)\dot{S}(t)\exp(j\phi)$ включает в себя наряду с расширяющим кодом также амплитуду сигнала A , компонент $\dot{B}(t)$ модуляции данными и неизвестную начальную фазу ϕ . После умножения входного сигнала на опорные колебания и выделения низкочастотного компонента два результирующих полосных сигнала разностной несущей частоты $f_0 - f_1$ будут обладать комплексными огибающими вида

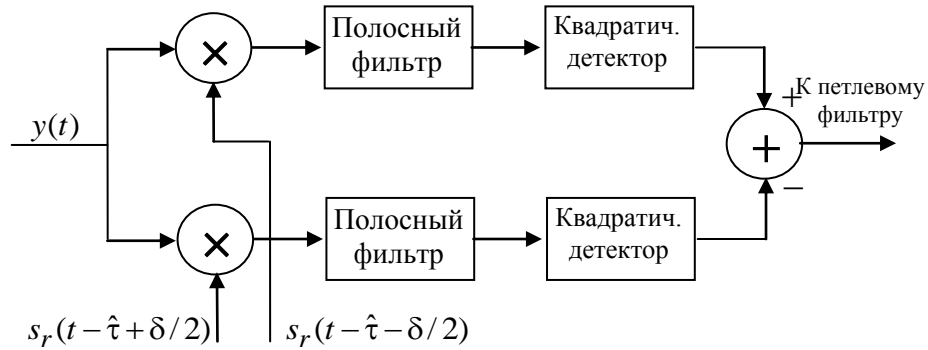


Рис.8.8. Ранне–поздний некогерентный дискриминатор

$A\dot{B}(t)\dot{S}(t)\dot{S}^*(t - \varepsilon \pm \delta/2)\exp[j(\phi - \vartheta)]$. Предположим, что полосные фильтры после умножителей обладают импульсной характеристикой с комплексной огибающей $\dot{H}(t)$. Тогда вещественные огибающие на выходах фильтров, вычисленные в терминах интеграла свертки (см. параграф 2.12.1), имеют вид

$$\left| \frac{A}{2} \int_{-\infty}^{\infty} \dot{B}(\theta)\dot{S}(\theta)\dot{S}^*\left(\theta - \varepsilon \pm \frac{\delta}{2}\right)\dot{H}(t - \theta)d\theta \right|.$$

Если импульсная характеристика фильтра представляет собой прямоугольник длительности T , равной длительности символа данных, а модуляция данных – фазовая, тогда выборки выходной вещественной огибающей в моменты $t = T$

$$\left| \frac{A}{2} \int_0^T \dot{B}(t)\dot{S}(t)\dot{S}^*\left(t - \varepsilon \pm \frac{\delta}{2}\right)dt \right| = \left| \frac{A}{2} \int_0^T \dot{S}(t)\dot{S}^*\left(t - \varepsilon \pm \frac{\delta}{2}\right)dt \right| = AE \left| \dot{\rho}\left(\varepsilon \mp \frac{\delta}{2}\right) \right|,$$

пропорциональны модулю соответствующего значения АКФ $\dot{\rho}(\tau)$ расширяющей комплексной огибающей. Разность квадратов модулей снова дает дискриминационную кривую формы, подобную изображенной на рис. 8.6 (см. задачи 8.7, 8.15).

Реализация схемы на рис. 8.8 может столкнуться с трудностями, обусловленными дисбалансом параметров опережающей и запаздывающей ветвей. Известны различные решения по их преодолению [9,18,77], включая петлю с τ -джиттером (*tau-dither loop*) (другое название – с временным разделением – «time-shared»), где используется только одна ветвь, переключающаяся поочередно между опережающим и запаздывающим образцами.

8.4.3. Шумовые характеристики петли захвата по задержке (DLL).

Петля с захватом по задержке (DLL) является частным случаем схемы фазовой автоподстройки (phase-lock loop) и проявляет такие же трудности в части анализа своего поведения, которые присущи нелинейным системам с обратной связью [89,90]. Однако одна из наиболее важных характеристик работы DLL – шумовая ошибка устойчивого состояния устройства слежения за задержкой – легко вычисляется при условии применимости

линейной аппроксимации.

В реальной практике, как правило, желательна достаточно малая шумовая ошибка, свидетельствующая о хорошей способности петли к фильтрации шума. Флюктуации выходного сигнала петли могут считаться малыми, если ошибка, т.е. разность между истинным текущим значением задержки сигнала τ и ее оценкой $\hat{\tau}$, вырабатываемой DLL, удерживается в рамках линейной зоны дискриминационной кривой с вероятностью, близкой к единице. Если это условие выполняется, то можно полагать, что дискриминационная кривая линейна в широком диапазоне значений ошибки $\varepsilon = \hat{\tau} - \tau$. Последнее позволяет линеаризовать модель системы следующим образом.

Ограничимся узкополосным (или, что эквивалентно, когерентным) дискриминатором DLL и вычислим спектральную плотность мощности шума $\tilde{N}_d(f)$ на его выходе. Поскольку коррелятор подобного дискриминатора вычисляет корреляцию наблюдения с опорным сигналом $s_r(t) = s(t - \delta/2) - s(t + \delta/2)$, то, согласно (2.15), выходная дисперсия шума определяется как $\sigma^2 = N_0 E_r / 2$, где E_r – энергия опорного сигнала за время интегрирования T . Если, как и ранее, E – энергии стандартного сигнала, то $E_r = 2E[1 - \rho(\delta)]$ и $\sigma^2 = N_0 E[1 - \rho(\delta)]$. Интегрирование на интервале T может трактоваться как низкочастотная фильтрация с полосой $W_f = 1/T$, означающая, что найденная мощность шума распределена в данной полосе и, следовательно, односторонняя спектральная плотность мощности шума на выходе дискриминатора составит

$$\tilde{N}_d(f) = \frac{\sigma^2}{W_f} = N_0 E T [1 - \rho(\delta)]. \quad (8.23)$$

При прямоугольной форме чипа и пренебрежимо малом уровне боковых лепестков $\rho(\tau)$ представляет собой треугольник с единичной высотой и основанием Δ , так что $\tilde{N}_d(f) = N_0 E T$ при $\delta \geq \Delta$.

Замена истинной дискриминационной кривой ее воображаемой линейной аппроксимацией просто означает бесконечное продолжение линейного участка в окрестностях нуля с тем же наклоном S_d . Последняя величина может быть найдена из (8.22), учитывая четность АКФ

$$S_d = \left. \frac{de(\varepsilon)}{d\varepsilon} \right|_{\varepsilon=0} = AE \left[\rho' \left(\frac{\delta}{2} \right) - \rho' \left(-\frac{\delta}{2} \right) \right].$$

Повторное обращение к треугольной форме АКФ дает

$$S_d = \begin{cases} -\frac{2AE}{\Delta} = -2AEW, & \delta < 2\Delta, \\ -\frac{AE}{\Delta} = -AEW, & \delta = 2\Delta, \end{cases} \quad (8.24)$$

где подставлена оценка полосы дискретного сигнала $W = 1/\Delta$.

Представим теперь, что вместо истинного шума $n(t)$, который добавляется к принятому сигналу, непосредственно к измеряемому параметру τ прибавляется эквивалентный шум $n_\tau(t)$ со спектральной плотностью мощности $\tilde{N}_\tau(f) = \tilde{N}_d(f) / S_d^2$. На выходе линейного дискриминатора, заменяющего реальный аналог, этот воображаемый шум будет не отличим от истинного шума на выходе реального дискриминатора, поскольку его спектральная мощность $S_d^2 \tilde{N}_\tau = \tilde{N}_d(f)$ является абсолютно такой же. В итоге приходим к модели системы, изображенной на рис. 8.9, на вход которой поступает не сигнал, искаженный шумом, а параметр τ в смеси с воображаемым аддитивным шумом $n_\tau(t)$. Данная смесь обрабатывается линейной замкнутой петлей, в которой оценка $\hat{\tau}$ вычитается из

входной величины, образуя выходной сигнал ошибки ε . Затем фильтр с передаточной функцией $\tilde{h}(f)$ сглаживает сигнал ошибки ε , умноженный на крутизну дискриминатора S_d , выработывая выходную оценку $\hat{\tau}$. В данном варианте фильтрация включает в себя операции, выполняемые петлевым фильтром и ГУНом по превращению сигнала ошибки

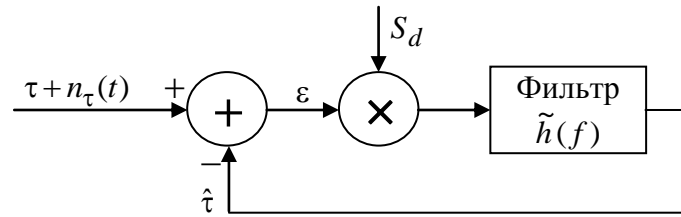


Рис.8.9. Линеаризованная модель DLL.

реального дискриминатора $e(t)$ в соответствующий сдвиг опорного сигнала. Данная модель является полностью линейной, и дисперсия $D\{\tau\}$ случайных флюктуаций на ее выходе может быть найдена на основе принципа суперпозиции не зависимо от сигнального компонента как

$$D\{\hat{\tau}\} = \int_0^{\infty} \tilde{N}_\tau(f) |\tilde{h}_l(f)|^2 df, \quad (8.25)$$

где $\tilde{h}_l(f)$ – передаточная функция замкнутой петли. Для нахождения последней характеристики достаточно подать на вход петли дельта-функцию. Тогда выходной спектр, являющийся в точности $\tilde{h}_l(f)$, подчиняется соотношению $\tilde{h}_l(f) = [1 - \tilde{h}_l(f)] S_d \tilde{h}(f)$, приводящему к правилу, которое хорошо известно из теории линейных систем с обратной связью [2,7]

$$\tilde{h}_l(f) = \frac{S_d \tilde{h}(f)}{1 + S_d \tilde{h}(f)}. \quad (8.26)$$

Эквивалентный спектр флюктуаций задержки $N_\tau(f)$ распределен в полосе $W_f = 1/T$, которая, как правило, значительно шире полосы замкнутой петли; в противном случае последняя не сможет эффективно сгладить флюктуации шума. Это допускает следующую версию (8.25)

$$D\{\hat{\tau}\} = \tilde{N}_\tau(f) B_N = \frac{\tilde{N}_d(f) B_N}{S_d^2}, \quad (8.27)$$

в которой шумовая полоса петли B_N определяется как

$$B_N = \int_0^{\infty} |\tilde{h}_l(f)|^2 df = \int_0^{\infty} \left| \frac{S_d \tilde{h}(f)}{1 + S_d \tilde{h}(f)} \right|^2 df.$$

Теперь, возвращаясь к (8.23) и (8.24) и отмечая, что $\rho(\delta) = 1 - \delta/\Delta$, $0 < \delta < \Delta$, можно представить (8.27) в следующей форме

$$D\{\hat{\tau}\} = \begin{cases} \frac{\delta/\Delta}{4W^2 q_l^2}, & 0 < \delta < \Delta, \\ \frac{1}{4W^2 q_l^2}, & \Delta \leq \delta < 2\Delta, \\ \frac{1}{W^2 q_l^2}, & \delta = 2\Delta, \end{cases} \quad (8.28)$$

где

$$q_l^2 = \frac{A^2 E}{N_0 B_N T} = \frac{A^2 P}{N_0 B_N}, \quad (8.29)$$

называется отношением сигнал-шум в петле. Причина для такого названия очевидна: числитель (8.29) содержит действительную энергию сигнала, поскольку $P = E/T$ – мощность стандартного (имеющего амплитуду $A=1$) сигнала. В то же время знаменатель представляет собой мощность шума в пределах шумовой полосы петли.

Соотношение (8.28) полностью подобно формуле Вудворда для потенциально достижимой точности временного измерения, представленной в параграфе 2.12.2:

$$D\{\hat{\tau}\} \approx \frac{1}{(2\pi W_{rms})^2 q^2}, \quad q \gg 1,$$

подчеркивающее, что следящая петля является адекватным способом временного измерения.

Имеются три параметра, затрагивающие точность DLL в устойчивом состоянии: уровень сигнала относительно шума $A^2 P/N_0$, шумовая полоса петли B_N и расстояние δ между опережающим и запаздывающим стробами. Первый из них представляет собой «грубый» ресурс повышения точности и не нуждается в специальных пояснениях. Оптимизация по второму параметру не является простой задачей, поскольку механическое уменьшение шумовой полосы без тщательного проектирования петлевого фильтра может драматически ухудшить такие динамические свойства системы, как длительность захвата и способность отслеживания сигнала с изменяющимся временем задержки. Выбор разнесения δ также является предметом компромисса: уменьшение δ по сравнению с длительностью чипа Δ обеспечивает более высокую точность оценки благодаря положительной корреляции шума на выходе опережающей и запаздывающей ветвей. В то же время, чем меньше разнесение δ , тем уже сама дискриминационная характеристика (см. рис. 8.6). Это предполагает более жесткие требования к точности поиска, поскольку последний должен гарантировать попадание временного рассогласования между локальной опорой и принятым сигналом внутрь активной (не нулевой) зоны дискриминационной характеристики. Другим фактором, которым нельзя пренебрегать, является опасность потери синхронизма, увеличивающаяся при сужении дискриминационной характеристики. Общепринятым путем достижения согласия между конфликтующими требованиями к параметрам следящей петли служит процедура адаптации: на начальном этапе ввода в слежение могут использоваться более широкая шумовая полоса и большее разнесение, которые после окончания переходных процессов уменьшаются для достижения более высокой точности в устойчивом состоянии.

Задачи.

8.1. Необходимо организовать последовательный поиск с постоянным временем анализа $t_d = 2$ мс. Дискретный сигнал, который необходимо отыскать, занимает полосу 1 МГц и обладает кодом длиной $L=1000$. Отсутствует априорная информация о фазе кода, а начальное частотное рассогласование местного генератора и несущей частотой сигнала лежит в диапазоне ± 10 КГц. Грубо оценить минимальное число анализируемых ячеек.

8.2. Найти асимптотические приближения полной средней вероятности и среднего числа шагов последовательного поиска, если вероятность ложной тревоги очень мала ($Mp_f \ll 1$). Дать практическое толкование полученным результатам.

8.3. Используется последовательный поиск с фиксированным временем анализа и порогом, оптимизированным для некоторой мощности сигнала P . Что произойдет с полной средней вероятностью поиска и средним числом шагов в двух предельных случаях: $P \rightarrow 0$ и $P \rightarrow \infty$, если не производить перенастройку времени анализа и порога? Дать обоснование полученных результатов.

8.4. Найти выражения для вероятностей ложной тревоги и обнаружения, а затем для времени анализа на ячейку, необходимого для обеспечения заданных p_f, p_d , если флюктуации амплитуды сигнала подчиняются закону Релея (3.12), а случайная начальная фаза равномерно распределена на интервале $[-\pi, \pi]$. (Подсказка: простейшим путем является трактование сигнала в качестве гауссовского процесса, независимого от шума.)

8.5. Найти выражение для вероятности правильного завершения и времени поиска для последовательности быстрого поиска Стиффлера периода $L = 2^n$.

8.6. Построить дискриминационную кривую когерентной DLL для случаев $\delta = \Delta$, $0 < \delta < \Delta$ и $\Delta < \delta < 2\Delta$. Определить протяженность зоны максимальной крутизны. Почему значения $\delta > 2\Delta$ неуместны?

8.7. Построить дискриминационную кривую некогерентной схемы DLL при разнесении $\delta = \Delta$. Почему значение $\delta = 2\Delta$ неуместно?

8.8. Показать, что генератор, управляемый напряжением, схемы DLL функционирует, как интегратор сигнала ошибки. Предполагается отсутствие шума на входе и постоянство задержки сигнала. Доказать, что в установившемся режиме ошибка на выходе DLL равна нулю. Будет ли данное утверждение верно, если задержка сигнала линейно изменится, а DLL не содержит дополнительных интеграторов?

8.9. Рассмотрим схему DLL без дополнительного петлевого фильтра. Предполагается, что крутизна дискриминатора равна 0.5 В/мкс, а ГУН изменяет свою частоту на 100 КГц на 1 В. Первоначальная разность частот ГУНа и принятого сигнала составляет 10 КГц. Определить величину ошибки в установившемся состоянии DLL при отсутствии шума.

8.10. Определить шумовую полосу и дисперсию выходной ошибки в терминах крутизны дискриминатора и усиления ГУНа схемы DLL, не содержащей петлевой фильтр. Дать обоснование зависимости этих характеристик от параметров системы.

9. Канальное кодирование в широкополосных системах.

9.1. Предварительные замечания и терминология.

В ходе передачи, хранения и обработки данные необходимо представлять в некоторой подходящей форме. В цифровой связи первичное сообщение, вырабатываемое источником, может рассматриваться, как последовательность бит данных или битовый поток (*bitstream*). Битовый поток отображается в последовательность символов некоторого определенного алфавита, что традиционно называется *кодированием*. Цели кодирования могут быть различными. Например, термины кодирования источника или компрессии данных означает ликвидацию избыточности битового потока с целью представления данных источника в наиболее экономичной форме. Другим вариантом кодирования является шифрование, осуществляемое для защиты данных от непреднамеренного перехвата или фальсификации. Предметом данной главы является *канальное кодирование*, направленное на обеспечение передачи данных по каналу связи, максимально невосприимчивой к вредным эффектам неустранимых канальных помех. Частные ситуации, рассмотренные в параграфах 2.3, 2.5–2.7, показывают насколько важно найти соответствующий сигнальный вариант для преодоления разрушительного влияния канального шума. Наряду модуляцией канальное кодирование предоставляет инструментарий для эффективной сигнализации.

В течение более чем пятидесятилетнего существования теория канального кодирования руководствовалась и мотивировалась фундаментальной теоремой Шеннона о пропускной способности канала, упомянутой в главе 1. Согласно этой теореме, любой канал характеризуется некоторой константой C (выраженной в бит/сек), называемой пропускной способностью, которая определяет верхнюю границу достижимой скорости R передачи информации по каналу. Если $R > C$, то не существует способа построения сигнала, обеспечивающего произвольную достоверность передачи данных. Наоборот, когда $R < C$ всегда можно найти код, гарантирующий любую желаемую вероятность перепутывания одного сообщения с другим на приемной стороне (см. рис 1.1). Теорема о пропускной способности Шеннона, давая строго математическое обоснование существования кода, не указывает никакого конкретного кодового алгоритма достижения устанавливаемого ею качества. Более того, ее доказательство, базирующееся на усредненной по всем возможным канальным кодам вероятности ошибки, показывает, что почти все коды достаточно большой длины будут хороши с этой точки зрения. Более того, поиски конкретных кодовых правил, позволяющих достичь шенноновского предела, оказались тщетными вплоть до момента открытия турбо-кодов в 1993, хотя множество важных и широко используемых результатов было получено в ходе стремления к этой цели.

Очевидно, что современная теория кодирования слишком сложна, чтобы позволить изложить даже ее начальные основы в небольшой главе. Все это оказывается еще более печальным на фоне ключевой роли теории кодирования в глобальной информационной технологии, в которой широкополосная связь является только одним из направлений. Кроме того, роль канального кодирования в широкополосных системах чрезвычайно высока, поскольку большинство из них проектируются для работы в сложной помеховой обстановке и, более того, многие из них сами создают сильные внутрисистемные помехи (помехи множественного доступа (МАИ) в асинхронном варианте CDMA). Влияние помех МАИ, в отличие от естественного (теплового) шума, не может быть преодолено только посредством «грубой силы», т.е. увеличением мощности сигнала, поскольку все пользователи обладают равными правами и выигрыш в отношении сигнал-помеха (SIR) одного из них, достигнутый этим способом, оборачивается потерями для других (см. параграфы 4.5–4.6). Последнее оставляет проектировщику только два способа преодоления МАИ: увеличение показателя расширения или привлечение мощных канальных кодов. Руководствуясь принципом рациональности, ограничимся рассмотрением в этой главы только проблем кодирования, связанных с коммерческими 2G и 3G широкополосными стандартами

cdmaOne, UMTS и cdma2000. По этой причине приведенный ниже математический инструментарий, обозначения и способ описания в максимально степени приспособлены соответствовать этой частной задаче наиболее экономичным и легким путем. Заинтересованного в более глубоком охвате этого вопроса читателя отсылаем к книгам по теории кодирования (например, [31,33,91]).

Начнем с некоторой основной классификации канальных кодов. Первой характерной чертой их различия является размер алфавита, согласно которому можно говорить о бинарных, троичных и т.п. кодах. Хотя область применения не двоичных (например, Рида-Соломона или Унгербоека) кодов в настоящее время достаточно широка, сконцентрируем наше внимание только на двоичных кодах, используемых в вышеупомянутых спецификациях. Другим направлением группирования кодов служит способ отображения информационных данных в *кодовые слова* или *кодовые вектора* (т.е. последовательности кодовых символов, переносящих передаваемое сообщение). Дело заключается в том, что любое канальное кодирование состоит во введении определенной избыточности в сообщение, которое увеличивает разнесение излученных сигналов друг относительно друга и, следовательно, делает их менее подверженными перепутыванию. В зависимости от способа добавления этой избыточности все канальные коды подразделяются на *блоковые* и *древовидные*. Характерной особенностью блоковых кодов является сегментация битового потока источника, которая состоит в делении на блоки из k информационных бит и представлении каждого блока $n > k$ двоичными символами. В результате избыточные $n - k$ символов служат для защиты только k бит источника данного кодового слова. Кодовые слова древовидных (например, сверточного) кодов обладают отличной структурой: непрерывный битовый поток источника кодируется в бесконечный поток кодовых символов (*code-stream*) без фрагментации (см. детали в параграфе 9.3).

Достигнув приемной стороны, кодовое слово необходимо обратно отобразить в переданные биты данных. Данная операция называется *декодированием*. Фактически, благодаря модуляции, любое кодовое слово передается в канал в виде некоторого сигнала. При прохождении АБГШ (либо другого стационарного) канала сигнал подвергается искажению шумом, мгновенные отсчеты которого непрерывны. В случае гауссовского шума оптимальная (максимально правдоподобная) стратегия решения приемника эквивалентна правилу решения по минимуму евклидова расстояния (см. параграф 2.1), которое означает принятие за истинный сигнала, ближайшего к полученному наблюдению. Такая непосредственная процедура заканчивается получением декодированных бит данных и носит название (наряду с многочисленными приближениями) *мягкого* декодирования. Сложность алгоритмов мягкого декодирования послужила причиной частого применения альтернативного варианта, называемого *жестким* декодированием. Данный метод декодирования включает два этапа: на первом из них принимаются решения по всем индивидуальным кодовым символам и, в результате, наблюдение оказывается демодулированным в вектор, состоящий из символов, принадлежащих кодовому алфавиту (в рассматриваемом случае – двоичному). Некоторые символы, полученного таким образом *двоичного наблюдения*, могут оказаться ошибочными, и тогда весь демодулированный вектор, возможно, будет отличаться от всех кодовых слов. Второй этап заключается в отыскании среди всех допустимых кодовых слов одного максимально правдоподобного, т.е. отвечающего наибольшей вероятности трансформации каналом в текущее демодулированное двоичное наблюдение. Подобная процедура декодирования, заканчивающаяся, как правило, принятием решения об истинности некоторого конкретного кодового вектора, называется *исправлением ошибок*. В качестве альтернативы декодирование может ограничиться только проверкой текущего двоичного наблюдения: является ли оно подлинным кодовым словом, либо произошли какие-то ошибки вследствие разрушительного действия канала. Тогда, если двоичное наблюдение принадлежит множеству разрешенных кодовых слов, оно обратно отображается соответствующей последовательностью бит данных. В противном случае регистрируется факт неудачной передачи, и приемник либо запрашивает передатчик о повторе

сообщения (как в случае систем с активным запросом), либо пытается восстановить его путем интерполяции предыдущих и последующих сообщений. Этот тип декодирования, называемый *обнаружением ошибок*, характерен при применении блочных кодов в современных коммерческих беспроводных широкополосных системах, и следующий параграф сфокусирован на форматах кодирования, связанных с работой по обнаружению ошибок.

9.2. Блочные коды, обнаруживающие ошибки.

9.2.1. Двоичные блочные коды и способность к обнаружению ошибок.

Предположим, что блок b_0, b_1, \dots, b_{k-1} представляет собой k бит источника, которые необходимо закодировать двоичным кодовым словом $\mathbf{u} = (u_0, u_1, \dots, u_{n-1})$ длины $n > k$. Все 2^k комбинаций из k бит источника полагаются возможными, что означает существование $M = 2^k$ кодовых слов. В каждом кодовом слове $n - k$ двоичных символов являются избыточными в том смысле, что только k символов необходимо для однозначного отображения M сообщений источника в двоичные вектора. Эти избыточные символы наделяют кодовые слова большей различимостью друг от друга, обеспечивая тем самым лучшую сопротивляемость воздействию шума. Множество всех $M = 2^k$ слов длины n называется (n, k) блочным кодом. Жесткое декодирование, т.е. демодуляция непрерывного наблюдения в двоичное, отвечает модели *двоичного симметричного канала* (ДСК) – *binary symmetrical channel* (BSC), трансформирующего входные $u = 0, 1$ в противоположные выходные двоичные символы $y \neq u$ с *переходной* (*crossover*) (или *символьной ошибкой* (*symbol error*)) вероятностью p . Прилагательное «симметричный» подчеркивает одинаковые вероятности переходов «0» в «1» и наоборот (см. рис.9.1).

Предположим, что $\mathbf{y} = (y_0, y_1, \dots, y_{n-1})$ – двоичное наблюдение на выходе ДСК. Ес-

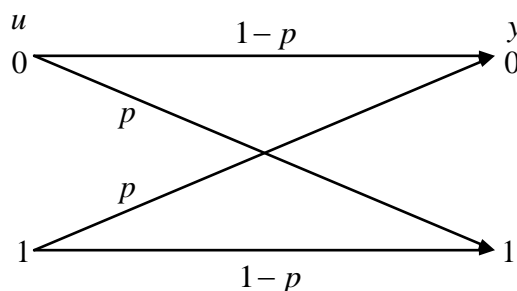


Рис.9.1. Модель двоичного симметричного канала

ли вектор \mathbf{y} не совпадает ни с одним из M кодовых векторов \mathbf{u} , то приемник считает, что имеются ошибочные символы (обнаружение ошибок). В противном случае выделяются k бит данных, отвечающих оцененному кодовому слову $\hat{\mathbf{u}} = \mathbf{y}$. Очевидно, что если переданным является слово \mathbf{u}_0 , а двоичное наблюдение совпадает с другим кодовым словом, т.е. $\hat{\mathbf{u}} = \mathbf{y} \neq \mathbf{u}_0$, то имеет место не обнаруживаемая ошибка, а выделенные биты не отвечают переданным.

Введем еще несколько определений. Расстоянием Хэмминга $d_H(\mathbf{f}, \mathbf{g})$ между двумя векторами $\mathbf{f} = (f_0, f_1, \dots, f_{n-1})$ и $\mathbf{g} = (g_0, g_1, \dots, g_{n-1})$ одинаковой длины n называется число позиций, на которых вектора имеют отличные элементы $f_i \neq g_i$. Весом Хэмминга $w_H(\mathbf{f})$ вектора \mathbf{f} называется число его ненулевых компонент. Если, например,

$\mathbf{f} = (01011)$, $\mathbf{g} = (11001)$, то $d_H(\mathbf{f}, \mathbf{g}) = 2$, а $w_H(\mathbf{f}) = 3$. Непосредственно можно убедиться, что $d_H(\mathbf{f}, \mathbf{g}) = w_H(\mathbf{f} - \mathbf{g})$ и $w_H(\mathbf{f}) = d_H(\mathbf{f}, \mathbf{0})$, где $\mathbf{0}$ – нулевой вектор.

Предположим, что передается кодовое слово \mathbf{u} . ДСК преобразует его в отличное фиксированное кодовое слово \mathbf{v} (вызывая, тем самым, не обнаруживаемую ошибку и выдавая неправильные биты данных, соответствующие \mathbf{v}), если ошибки в символах произойдут на всех $d_H(\mathbf{u}, \mathbf{v})$ позициях, где \mathbf{u} и \mathbf{v} отличаются, и не будут повреждены на $n - d_H(\mathbf{u}, \mathbf{v})$ позициях, где символы отличаются. Тогда для ДСК без памяти, т.е. канала, в котором все символьные ошибки независимы, можно считать, что вероятность $P(\mathbf{y} = \mathbf{v} | \mathbf{u})$ вышеупомянутого события будет

$$P(\mathbf{y} = \mathbf{v} | \mathbf{u}) = p^{d_H(\mathbf{u}, \mathbf{v})} (1 - p)^{n - d_H(\mathbf{u}, \mathbf{v})}. \quad (9.1)$$

Поскольку $p < 1/2$ (в противном случае можно было бы просто переобозначить выходные символы «0» и «1»), то для уменьшения вероятности перепутывания кодовых слов \mathbf{u} и \mathbf{v} расстояние Хэмминга между ними должно быть максимально возможным. Рассмотрим теперь расстояния Хэмминга $d_H(\mathbf{u}, \mathbf{v})$ между всеми различными кодовыми векторами кода U . Обозначим наименьшее среди них как d_H и назовем его (минимальным) кодовым расстоянием кода U :

$$d_H = \min_{\substack{\mathbf{u} \neq \mathbf{v} \\ \mathbf{u}, \mathbf{v} \in U}} d_H(\mathbf{u}, \mathbf{v}). \quad (9.2)$$

Передаваемым может оказать любое кодовое слово и для минимизации риска перепутывания двух ближайших кодовых векторов расстояние между ними, т.е. кодовое расстояние d_H , должно быть максимально возможным. В итоге приходим к следующему утверждению:

Утверждение 9.2.1. Код U способен обнаруживать любые t_d и менее ошибки в символах (вплоть до t_d ошибок), если и только если кодовое расстояние $d_H \geq t_d + 1$.

Действительно, возьмем код с $d_H \leq t_d$, и выберем пару ближайших кодовых векторов \mathbf{u}, \mathbf{v} . Если d_H символов \mathbf{u} , отличных от аналогичных вектора \mathbf{v} , будут искажены, то \mathbf{u} превратится в \mathbf{v} , означая, что существует образец с не более, чем t_d ошибками, который не обнаруживается. И наоборот, если расстояние между любыми кодовыми векторами превосходит t_d , то не существует образца с t_d или менее ошибками, который трансформирует кодовое слово в любое другое.

Аналогичным образом можно доказать следующее положение (задача 9.4).

Утверждение 9.2.2. Код U способен исправлять t_c и менее ошибочных символов, если и только если его кодовое расстояние $d_H \geq 2t_c + 1$.

9.2.2. Линейные коды и их полиномиальное представление.

Будем трактовать двоичные кодовые символы $\{0,1\}$ как элементы двоичного конечного поля $GF(2)$ (см. параграф 6.6) и рассмотрим посимвольные линейные операции над кодовыми словами кода U , подчиняющиеся арифметическим правилам $GF(2)$. Очевидно, что существует только одна не тривиальная операция этого типа, а именно, посимвольное сложение в $GF(2)$: $\mathbf{u} = (u_0, u_1, \dots, u_{n-1})$, $\mathbf{v} = (v_0, v_1, \dots, v_{n-1}) \Rightarrow \mathbf{u} + \mathbf{v} = (u_0 + v_0, u_1 + v_1, \dots, u_{n-1} + v_{n-1})$. Например, если $\mathbf{u} = (100111)$, $\mathbf{v} = (010110)$, то $\mathbf{u} + \mathbf{v} = (110001)$. Посимвольное вычитание не играет самостоятельной роли и просто повторяет сложение, поскольку в $GF(2)$ отрицание элемента есть сам элемент. Аналогично,

посимвольное умножение на скаляр из $GF(2)$ (т.е. на 0 или 1) любого кодового слова превращает его либо в нулевой вектор, либо не изменяет его совсем.

Двоичный код U называется линейным, если сумма любого числа кодовых векторов снова дает кодовый вектор, принадлежащий U . Данное название следует из того факта, что подобный код образует векторное (линейное) пространство над полем $GF(2)$ [31,33,91], хотя эта концепция и не находит серьезного применения в дальнейшем изложении. Любой линейный код U длины n содержит нулевой вектор (т.е. с n нулевыми компонентами) в качестве кодового, поскольку сумма произвольного кодового вектора, входящего в U , с самим собой в точности дает нулевой вектор: $\mathbf{u} + \mathbf{u} = 0$. Следующее положение служит объяснением одной из причин особенного интереса к линейным кодам.

Утверждение 9.2.3. Кодовое расстояние линейного кода U равно минимальному весу Хэмминга среди всех ненулевых кодовых слов:

$$d_H = \min_{\substack{\mathbf{u} \in U \\ \mathbf{u} \neq 0}} w_H(\mathbf{u}). \quad (9.3)$$

Для доказательства этого положения достаточно произвести подстановку $d_H(\mathbf{u}, \mathbf{v}) = w_H(\mathbf{u} - \mathbf{v})$ в (9.2) и заметить, что разность $\mathbf{u} - \mathbf{v} = \mathbf{u}'$ вновь кодовое слово U . Как показывает соотношение (9.3), отсутствует необходимость проверки всех $M(M-1)/2$ различных векторных пар для определения кодового расстояния линейного кода, содержащего M слов. Достаточно «взвесить» $M-1$ ненулевых кодовых векторов, т.е. осуществить в $M/2$ раз меньшее число тестов, что демонстрирует, учитывая, как правило, значительное число слов M , чрезвычайно значительный выигрыш.

Для более быстрого понимания идеи конструирования обнаруживающих ошибки кодов для 2G и 3G беспроводных стандартов значительную помощь может оказать полиномиальное описание линейных кодов.

Сопоставим кодовому слову $\mathbf{u} = (u_0, u_1, \dots, u_{n-1})$ кодовый полином $u(z)$ относительно переменной z , упорядоченный как

$$u(z) = u_{n-1}z^{n-1} + u_{n-2}z^{n-2} + \dots + u_0,$$

где по соглашению принимается, что $z^0 = 1$. Указанное полиномиальное представление, широко используемое в теории кодирования, представляет собой просто форму z -преобразования, лежащего в основе анализа дискретных линейных систем, обработки дискретных сигналов, цифровой фильтрации и т.д. [2, 7]. Однозначное соответствие между множествами кодовых слов и кодовых полиномов означает, что сумма двух кодовых полиномов $u(z), v(z)$ линейного кода U снова является кодовым полиномом того же кода.

В частности, если $u(z) = \sum_{i=0}^{n-1} u_i z^i$, $v(z) = \sum_{i=0}^{n-1} v_i z^i$ – кодовые полиномы слов \mathbf{u}, \mathbf{v} линейного кода U , то $u(z) + v(z) = \sum_{i=0}^{n-1} (u_i + v_i) z^i$ кодовый полином слова $\mathbf{u} + \mathbf{v} \in U$, где сложение коэффициентов подчиняется правилам поля, которому они принадлежат, т.е. в нашем случае полю $GF(2)$.

Полиномиальная арифметика, используемая в анализе и синтезе кодов, включает еще две операции: умножение и деление с *остатком*. Правила этих операций универсальны независимо от полей, которым принадлежат полиномиальные коэффициенты, но, имея дело с двоичными кодами, будем пользоваться двоичной арифметикой. Рассмотрим произвольный (не обязательно кодовый) двоичный (т.е. с коэффициентами из $GF(2)$) полином $a(z)$. Наибольшая степень z в этом полиноме, содержащая ненулевой коэффициент, называется *степенью* $a(z)$ с обозначением $\deg a(z)$. Пусть $a(z), b(z)$ – два двоичных полинома, причем $\deg a(z) = m$, $\deg b(z) = n$. Тогда их произведение $a(z)b(z)$ является полиномом степени $m+n$, полученным в результате распространения коммутативного ($z^i a = a z^i$) и дистрибутивного законов на операции, включающие формальную переменную

ную z и объединяющие вместе коэффициенты одинаковых степеней z :

$$\begin{aligned} a(z)b(z) &= (a_m z^m + a_{m-1} z^{m-1} + \dots + a_0)(b_n z^n + b_{n-1} z^{n-1} + \dots + b_0) = \\ &= a_m b_n z^{m+n} + (a_m b_{n-1} + a_{m-1} b_n) z^{m+n-1} + \dots + (a_1 b_0 + a_0 b_1) z + a_0 b_0 = \sum_{k=0}^{m+n} \left(\sum_{i=0}^k a_i b_{k-i} \right) z^k. \end{aligned}$$

Очевидно, что $z^m z^n = z^{m+n}$, все операции над a_i, b_i выполняются в $GF(2)$, а в последней внутренней сумме коэффициенты a_i, b_i , индексы которых становятся отрицательными или превосходят степень полинома должны быть положены равными нулю. Возьмем, например, двоичные полиномы $a(z) = z^4 + z^3 + 1$, $b(z) = z^2 + z + 1$, тогда их произведение $a(z)b(z) = z^6 + z^3 + z^2 + z + 1$.

Алгоритм деления делимого $a(z)$ на делитель $b(z)$ с остатком осуществляется следующим образом:

$$a(z) = q(z)b(z) + r(z), \quad (9.4)$$

где $q(z)$ является частным, а $r(z)$ – остатком от деления. Единственность $q(z), r(z)$ гарантируется соотношением $\deg a(z) = \deg q(z) + \deg b(z)$ и $\deg r(z) < \deg b(z)$. Алгоритм (9.4) подобен «школьному» правилу деления целых чисел с остатком, в котором степени полиномов заменяют целые величины. Одним из вариантов реализации этой операции является «длинное деление», т.е. последовательное вычисление остатка и деление его на делитель до тех пор, пока степень остатка станет меньше степени делителя. На первой итерации $b(z)$ умножается на z , возведенную в степень, уравнивающую степень произведения со степенью делимого $a(z)$. Вычитание (эквивалентно сложению в $GF(2)$) полученного произведения из $a(z)$ приводит к первому остатку, играющему на второй итерации роль делимого, и т.д. Проиллюстрируем это на примере.

Пример 9.2.1. Пусть $a(z) = z^4 + z^3 + 1$, $b(z) = z^2 + z + 1$. Применим алгоритм длинного деления:

$$\begin{array}{r} z^2 + 1 \\ z^2 + z + 1 \overline{) z^4 + z^3 + 1} \\ \underline{+} \\ z^4 + z^3 + z^2 \\ \underline{+} \\ z^2 + 1 \\ \underline{+} \\ z^2 + z + 1 \\ \underline{+} \\ z \end{array}$$

После двух итераций имеем $q(z) = z^2 + 1$, $r(z) = z$, так что деление с остатком приводит к следующему результату $z^4 + z^3 + 1 = (z^2 + 1)(z^2 + z + 1) + z$. \square

Как и в случае целых чисел, можно сказать, что $a(z)$ делится на $b(z)$ (или $b(z)$ делит $a(z)$), если остаток равен нулю, т.е. $a(z) = q(z)b(z)$.

Рассмотрим теперь линейный код U длины n со всеми кодовыми полиномами, делящимися на фиксированный полином $g(z)$ степени $r > 1$. Тогда любой кодовый полином U представим в виде $u(z) = b(z)g(z)$, и, поскольку имеется 2^{n-r} различных множителей $b(z)$, обеспечивающих степень произведения не больше $n-1$, то такой код может включать не более чем 2^{n-r} кодовых слов. Действительно, всегда возможно с одинаковыми

затратами пойти другим путем и построить код с этим максимальным числом слов, т.е. при заданном полиноме $g(z)$ фиксированной степени r передать $k = n - r$ информационных бит. Для этого достаточно использовать $k = n - r$ бит данных b_0, b_1, \dots, b_{k-1} в качестве коэффициентов полинома данных $b(z) = b_{k-1}z^{k-1} + b_{k-2}z^{k-2} + \dots + b_0$ и построить соответствующий кодовый полином как произведение $u(z) = b(z)g(z)$. Тогда 2^k различным k -битовым блокам данных однозначно соответствуют такое же число кодовых полиномов степени не выше чем $n - 1$. Линейность подобного кода легко проверяется (задача 9.11). В подобной конструкции полином $g(z)$ называется *порождающим полиномом* (*generator polynomial*) кода U . Отметим, что число избыточных или *проверочных* символов такого кода всегда равно r , т.е. степени порождающего полинома.

Когда полином данных $b(z)$ непосредственно умножается на порождающий полином $g(z)$, тогда кодовое слово, отвечающее полиному $u(z) = b(z)g(z)$, является несистематическим, т.е. биты данных в нем не выделены явным образом. Для построения *систематического* кода, в котором, например, последние k двоичных символов являются информационными битами, а $r = n - k$ первых – проверочными символами, необходимо осуществить определенную перегруппировку в кодовых словах. После подобной операции полное множество кодовых слов останется тем же и только отображение бит данных в кодовые слова поменяется. Умножим полином данных $b(z)$ на $z^r = z^{n-k}$, придя к полиному $z^{n-k}b(z)$ степени не выше $n - 1$. Если остаток $r(z)$ от деления произведения на $g(z)$ отбрасывается (или добавляется к $z^{n-k}b(z)$ в двоичном случае), то оно становится делимым на порождающий полином $g(z)$, т.е. превращается в кодовый полином. Последний

$$u(z) = z^{n-k}b(z) + r(z)$$

отвечает систематическому кодовому слову, поскольку биты данных в точности являются последними k коэффициентами $z^{n-k}b(z)$, а $r(z)$ степени меньшей, чем $r = n - k$ не может их затронуть.

Пример 9.2.2. Построим кодовое слово линейного (5,2) кода с порождающим полиномом $g(z) = z^3 + z^2 + 1$ при битах данных вида $b_0 = 1, b_1 = 1$. Тогда $b(z) = z + 1$, $z^{n-k}b(z) = z^4 + z^3$, а после деления на $g(z)$ получаем остаток $r(z) = z$. Суммирование $z^{n-k}b(z)$ с остатком приводит к кодовому полиному $u(z) = z^4 + z^3 + z = zg(z)$, соответствующего систематическому кодовому слову, последними двумя символами которого являются биты данных. \square

9.2.3. Вычисление синдрома и обнаружение ошибок.

Предположим, что по ДСК передается слово \mathbf{u} линейного кода U . На выходе элементы двоичного наблюдения \mathbf{y} , искаженные каналом, будут отличаться от переданных символов, что может быть записано как

$$\mathbf{y} = \mathbf{u} + \mathbf{e}, \tag{9.5}$$

где \mathbf{e} – *вектор ошибок*, состоящий из нулей и единиц на местах, которые отвечают не искаженным и искаженным символам соответственно. Например, если слово $\mathbf{u} = (01011)$ из примера 9.2.2 трансформировано ДСК в наблюдение $\mathbf{y} = (11110)$, то вектор ошибок $\mathbf{e} = (10101)$. Аналогично кодовым словам наблюдение $\mathbf{y} = (y_0, y_1, \dots, y_{n-1})$ и вектор ошибок $\mathbf{e} = (e_0, e_1, \dots, e_{n-1})$ могут быть представлены в полиномиальной форме

$y(z) = y_{n-1}z^{n-1} + y_{n-2}z^{n-2} + \dots + y_0$ и $e(z) = e_{n-1}z^{n-1} + e_{n-2}z^{n-2} + \dots + e_0$. Тогда (9.5) примет вид

$$y(z) = u(z) + e(z).$$

Пусть $g(z)$ – порождающий полином кода U . Остаток $s(z)$ от деления полинома наблюдения $y(z)$ на $g(z)$ называется *синдромом*. Поскольку любой кодовый полином делится на $g(z)$, то синдром повторяет остаток от деления полинома ошибок $e(z)$ на $g(z)$. Следовательно, ненулевой синдром всегда свидетельствует о наличии ошибок в наблюдении y и обнаружение ошибок может быть реализовано как вычисление синдрома наблюдения y и принятие решения о наличии ошибок при ненулевом синдроме.

Очевидно, что не все образцы ошибок обнаруживаемы, и любой не обнаруживаемый вектор ошибок всегда является некоторым кодовым вектором. Действительно, если $e(z)$ является кодовым полиномом, то он делится на $g(z)$ и синдром равен нулю. С другой стороны, нулевой остаток говорит только о том, что $e(z)$ делится на $g(z)$, но любой полином степени не выше, чем $n-1$, делящийся на $g(z)$, является кодовым полиномом.

Пример 9.2.3. Предположим, что передаваемое кодовое слово $\mathbf{u} = (01011)$ (5,2) линейного кода трансформировано ДСК в $\mathbf{y} = (11001)$, т.е. произошла двукратная ошибка. Деление $y(z) = z^4 + z + 1$ на $g(z) = z^3 + z^2 + 1$ приводит к ненулевому синдрому $s(z) = z^2$, сигнализирующему о наличии ошибок. С другой стороны, если бы наблюдение было бы $\mathbf{y} = (11101)$, что соответствовало бы трехкратной ошибке, синдром оказался бы равным нулю: $y(z) = z^4 + z^2 + z + 1 = (z+1)g(z)$, т.е. неспособным обнаружить такой образец ошибки. \square

Иногда способность линейного кода обнаруживать ошибки характеризуется долей не обнаруживаемых образцов ошибок в их общем числе. Поскольку всего возможно 2^n векторов ошибок и только 2^k , повторяющие кодовые вектора, не обнаруживаемы, то указанная доля составляет $2^{-(n-k)} = 2^{-r}$.

Только что рассмотренные линейные коды, построенные на основе порождающих полиномов, известны как циклические коды или укороченные циклические коды. Когда они используются только для обнаружения (а не для исправления) ошибок, они часто называются *циклическими избыточными кодами (cyclic redundancy codes (CRC))*.

9.2.4. Выбор порождающих полиномов для CRC.

Поскольку доля необнаруженных ошибок экспоненциально падает с ростом числа проверочных символов, то это может стимулировать использование порождающих полиномов высоких степеней. Однако следует напомнить, что проверочные символы являются избыточными и увеличение их числа сверх рационального уровня может оказаться расточительным. Как правило, CRC используются на более высоких уровнях системного протокола для проверки качества кадров данных, восстановленных на физическом уровне, т.е. после использования ресурсов значительно более мощных исправляющих ошибки кодов. Вследствие этого символьные ошибки, встречаемые CRC, достаточно редки и совершенно маловероятно, чтобы в кодовом слове с сотнями символов произошло более нескольких ошибок. Таким образом, обнаружение трехкратных ошибок на кодовое слово часто считается удовлетворительным. Опишем процедуру синтеза CRC, удовлетворяющего данному требованию.

Возьмем двоичный примитивный полином $g_1(z)$ степени m (см. параграф 6.6). Важным фактом, доказанным в алгебре расширенных полей, является то, что примитивный полином степени m никогда не делит бином $z^l + 1$ ненулевой степени $l < 2^m - 1$

[30,32]. Тогда можно доказать следующее положение.

Утверждение 9.2.4. Линейный код U с порождающим полиномом $g(z) = (z+1)g_1(z)$ степени m обнаруживает все ошибки кратности до трех, если длина кода $n \leq 2^m - 1$.

Вследствие линейности кода U (см. (9.3)) и утверждения 9.2.1 необходимо только доказать, что минимальный вес ненулевого слова U не меньше четырех. Любой кодовый полином $u(z)$, делящийся на $z+1$, может быть представлен в виде $u(z) = q(z)(z+1)$. Хотя z является формальной переменной, последнее равенство сохранится после подстановки $z=1$ в обе его части, приводя к результату $u(1) = u_{n-1} + u_{n-2} + \dots + u_0 = 0$, означающему четность числа ненулевых элементов слова, т.е. четность его веса. Полагая существующим слово веса два с ненулевыми i -м и j -м символами ($j > i$), имеем кодовый полином $u(z) = z^j + z^i = z^i(1 + z^{j-i})$, который не делится на $g_1(z)$, поскольку последний, являясь неприводимым, не может содержать z в качестве множителя, и не делит $1 + z^{j-i}$ ($j-i < n \leq 2^m - 1$) как примитивный. Следовательно, наименьший вес ненулевого слова U будет четыре. \square

Теперь выбор соответствующего порождающего CRC полинома очевиден. Если желателен код длины n , тогда просто следует найти примитивный полином $g_1(z)$ степени $m \geq \lceil \log_2(n+1) \rceil$ и построить порождающий полином в виде $g(z) = (z+1)g_1(z)$.

Коды CRC этого типа есть ничто иное, как популярные коды Хэмминга (либо укороченные при $n < 2^m - 1$) с исключенными словами нечетного веса. В альтернативном применении они используются для исправления любой однократной и обнаружения любой двукратной ошибок.

Пример 9.2.3. Несколько CRC кодов используются в 2G и 3G спецификациях [18,69,92]. Показательным примером может служить код, используемый всеми тремя стандартами (cdmaOne, UMTS и cdma2000), который обладает порождающим полиномом вида

$$g(z) = z^{16} + z^{12} + z^5 + 1 = (z+1)g_1(z), \quad \text{где}$$

$g_1(z) = z^{15} + z^{14} + z^{13} + z^{12} + z^4 + z^3 + z^2 + z + 1$ – примитивный полином. Аналогично, порождающие полиномы других CRC кодов этих стандартов (степеней 30, 24 и т.п.) представляют собой произведение бинорма $z+1$ и примитивного полинома.

\square

9.3. Сверточные коды.

Сверточные коды находят широкое применение в современных телекоммуникациях как эффективное средство обеспечения надежной передачи данных по каналам с шумами. Входя в состав более общего класса древовидных кодов, в рамках его они отличаются линейностью алгоритмов кодирования. Различие между сверточными и блоковыми кодами до некоторой степени размыто: любой сверточный код может трактоваться как блоковый код соответственно большой длины. Более разумно рассмотреть особенности и причины чрезвычайной популярности сверточных кодов в их рекуррентной форме, позволяющей реализовать значительно более практичные по сравнению с другими процедуры декодирования с исправлением ошибок (алгоритм Витерби).

9.3.1. Сверточный кодер.

Идея сверточного кодирования в общих словах может быть описана следующим образом. Возьмем блок (вектор) из v_c последовательных бит источника и линейно преобразуем его в $n > 1$ выходных двоичных кодовых символов, занимающих временной интервал, равный длительности одного бита источника. Линейность по отношению к векторам с компонентами из поля $GF(2)$ означает просто суммирование по модулю два выбранных компонентов. После этого обновим блок бит источника, включив в него один новый и исключив один наиболее старый. Снова имеем блок из v_c бит источника, задержанный относительно исходного блока на один бит (и содержащий $v_c - 1$ предшествующих бит и один новый), который кодируется n новыми кодовыми символами. Данная процедура непрерывно повторяется раз за разом, каждый момент времени вовлекая один новый бит и отбрасывая старейший. Рис. 9.2 иллюстрирует указанную процедуру для параметров $v_c = 3, n = 3$: кодер как будто просматривает битовый поток источника через скользящее окно шириной v_c и кодирует все биты, которые в текущий момент видит, в n кодовых

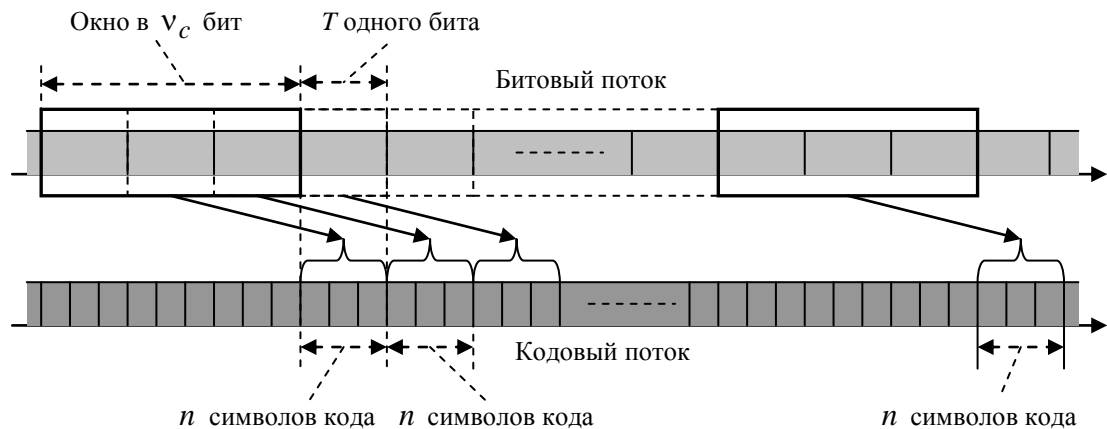


Рис. 9.2. Иллюстрация сверточного кодирования.

символов. После каждого шага процедуры окно сдвигается на длительность одного бита источника и выполняется следующий шаг. Число бит источника, определяющих кодовые символы на одном шаге, называется *длиной кодового ограничения* (*constraint length*). Описанный выше принцип может быть реализован структурой, представленной на рис. 9.3, где регистр сдвига, состоящий из $v_c - 1$ ячейки памяти, хранит $v_c - 1$ предшествующих бит источника. Вместе с поступающим битом они подаются в линейную логическую цепь, содержащую n сумматоров по модулю два. Выход каждой ячейки памяти и вход кодера

может быть, а может и не быть (что объясняет обозначение связей пунктирной линией) соединены с каждым сумматором, причем схема соединений определяет конкретную зависимость выходных кодовых символов от v_c бит источника, т.е. правило кодирования. При поступлении на вход текущего бита источника b_i на выходах сумматоров параллельно

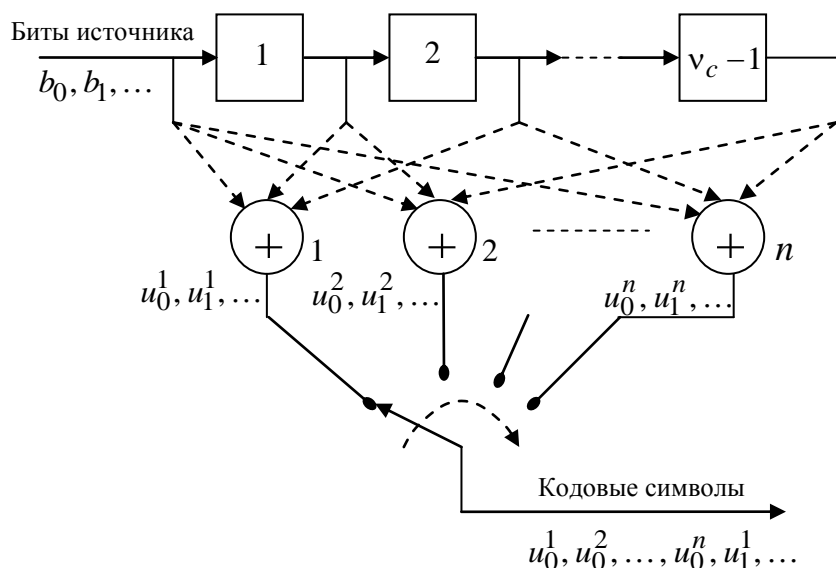


Рис. 9.3. Сверточный кодер.

но появляются n кодовых символов $u_i^1, u_i^2, \dots, u_i^n$. После каждого такта битовый образец в регистре смещается вправо на одну ячейку, подготавливая цепь к формированию следующих n кодовых символов. Выходной ключ в течение одного бита поочередно подключается к выходам n сумматоров и преобразует параллельную форму представления кодовых символов в последовательную, создавая выходной кодовый поток $u_0^1, u_0^2, \dots, u_0^n, u_1^1, u_1^2, \dots, u_1^n, \dots$. Видно, что в устойчивом состоянии кодер на рис.9.3 на каждый новый бит источника отвечает формированием n кодовых символов (см. также рис.9.2), так что скорость кода R_c , измеренная в битах на один кодовый символ, равна $1/n$. Нижеприведенный пример поможет лучше понять принцип сверточного кодирования.

Пример 9.3.1. Рис. 9.4 иллюстрирует реализацию сверточного кодера с длиной кодового ограничения $v_c = 3$ и скоростью $R_c = 1/2$. Входные биты b_0, b_1, \dots порождают два потока кодовых символов u_0^1, u_1^1, \dots и u_0^2, u_1^2, \dots , которые затем мультиплексируются таким образом, что u_i^1 и u_i^2 занимают соответственно четные и нечетные позиции в общем кодовом потоке. Например, битовый поток $\{b_i\} = 10100100\dots$ порождает последовательности $\{u_i^1\} = 11011111\dots$ и $\{u_i^2\} = 10001101\dots$, которые затем мультиплексируются в кодовый поток $\{(u_i^1, u_i^2)\} = 1110001011111011\dots$. □

Естественно возникает следующий вопрос: возможна ли при сверточном кодировании скорость R_c , отличная от $1/n$, т.е. равная k/n , где $1 < k < n$? Существуют два классических способа решения этой задачи. Первый состоит в обобщении вышеприведенного принципа: на каждом шаге kv_c , а не v_c , бит линейно преобразуются в n кодовых символов, после чего k наиболее старых бит (вместо одного) заменяются k новыми и кодер переходит к следующему шагу. Второй способ, называемый *выкалыванием (puncturing)*, ис-

пользует коды со скоростью $1/n$ в качестве подсобного материала и удаляет некоторые кодовые символы в соответствии с заранее назначенным образцом. При соответствующей организации процедуры выкалывания происходит уменьшение числа кодовых символов на бит данных, что обеспечивает достижение скорости k/n . Исходя из реализационных возможностей, процедура выкалывания часто считается предпочтительной, подчеркивая еще раз доминирующий интерес в отношении кодов со скоростями вида $1/n$. Следуя этой тенденции, сконцентрируем наше внимание только на сверточных кодах со скоростью

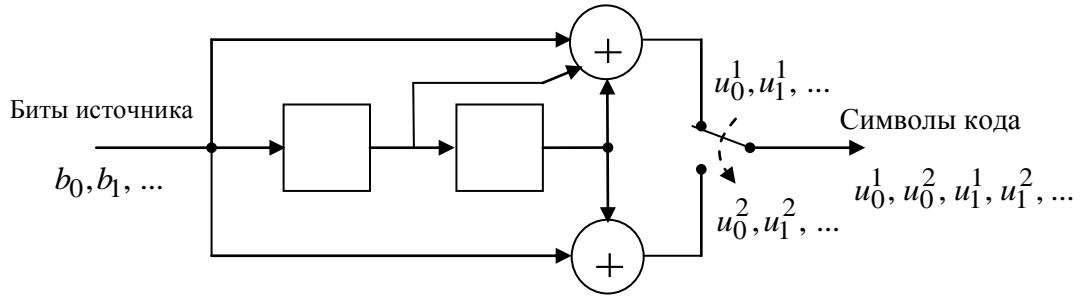


Рис. 9.4. Сверточный кодер со скоростью $1/2$.

$1/n$.

Очевидно, что схема, содержащая регистр сдвига и отдельные сумматоры по модулю два вместе со всеми их соединениями, представляет собой ничто иное, как *фильтр с конечным импульсным откликом* (КИО) (*finite impulse response (FIR) filter*) (см. рис.6.20), сигнал на выходе которого представляет собой свертку входного потока битов и импульсной характеристикой фильтра, что объясняет название рассматриваемых кодов. Данный факт лежит также в основе одного из способов формального описания сверточного кодера. Свертка, связывающая кодовый символ u_i^l (т.е. появляющийся на выходе l -го сумматора при поступлении b_i бита источника) с входным битовым потоком, определяется как

$$u_i^l = \sum_{t=0}^{v_c-1} b_{i-t} g_t^l, \quad i=0,1,\dots; \quad l=1,2,\dots,n, \quad (9.6)$$

где $g_t^l = 1$, если l -й сумматор соединен с t -й ячейкой памяти ($t=0$ соответствует входу кодера) и $g_t^l = 0$ в противном случае; $b_i = 0$ при $i < 0$. В частотной области подходящим инструментом для описания дискретных систем является z -преобразование, уже упоминавшееся в предыдущем параграфе. В z -области свертке соответствует произведение z -преобразований, так что соотношение (9.6) может быть представлено эквивалентной формой

$$u^l(z) = \sum_{i=0}^{\infty} u_i^l z^i = b(z) g_l(z), \quad l=1,2,\dots,n, \quad (9.7)$$

где $b(z) = \sum_{i=0}^{\infty} b_i z^i$ – z -преобразование входного потока бит, а

$$g_l(z) = g_0^l + g_1^l z + \dots + g_{v_c-1}^l z^{v_c-1}, \quad l=1,2,\dots,n \quad (9.8)$$

– передаточная функция l -го КИО-фильтра (т.е. формирующая l -й кодовый символ), называемая также l -м *порождающим полиномом* (*generator polynomial*) сверточного кода. Множество из n порождающих полиномов полностью определяет сверточный код, поскольку их ненулевые коэффициенты конкретизируют соединения сумматоров с регистром сдвига.

Пример 9.3.2. Кодер, изображенный на рис. 9.4, характеризуется порождающими полиномами $g_1(z) = 1 + z + z^2$ и $g_2(z) = 1 + z^2$. Не трудно убедиться, что последовательности кодовых символов, приведенные в примере 9.2.1, могут быть получены из (9.7). \square

Полагая $g_1(z) = 1$, получаем *систематический* сверточный код, в котором биты источника данных занимают вполне определенные позиции. Однако проблема заключается в том, что в рамках фиксированной структуры, изображенной на рис. 9.3, систематические коды, как правило, не оказываются лучшими с точки зрения исправляющей способности (см. задачу 9.15). Последнее обосновывает модификацию регистра сдвига сверточного кодера в структуру с обратной связью, если свойство систематичности играет важную роль. Более детально обсудим данный вопрос при ознакомлении с турбо-кодами в параграфе 9.4.1.

9.3.2. Решетчатая диаграмма, свободное расстояние и асимптотический выигрыш от кодирования.

Регистр сдвига сверточного кодера обладает $2^{v_c - 1}$ возможными состояниями и имеет только два состояния, в которые он может перейти из текущего состояния после очередного такта. Именно значение входного бита источника b_i определяет выбор одного из двух возможных переходов. Если на i -м тактовом интервале состояние регистра определяется, как $(b_{i-1}, b_{i-2}, \dots, b_{i-v_c+1})$, то следующее состояние будет либо $(0, b_{i-1}, \dots, b_{i-v_c+2})$ при поступлении на вход бита источника $b_i = 0$, либо $(1, b_{i-1}, \dots, b_{i-v_c+2})$, если $b_i = 1$. Аналогично, регистр приходит в состояние $(b_i, b_{i-1}, \dots, b_{i-v_c+1})$, если его предыдущее состояние было либо $(b_{i-1}, b_{i-2}, \dots, b_{i-v_c+2}, 0)$, либо $(b_{i-1}, b_{i-2}, \dots, b_{i-v_c+2}, 1)$. Для схематического изображения всех деталей поведения регистра адекватным инструментом служит *решетка (trellis)*. Она включает две колонки из $2^{v_c - 1}$ узлов, причем левая колонка отвечает текущему состоянию регистра, а правая – следующему. Ребра (стрелки) выходят из каждого узла левой колонки и входят в некоторые два узла следующей, причем сплошные и пунктирные ребра показывают пути, соответствующие приходу бита нуля или единицы соответственно. Аналогично, в каждый узел правой колонки входят два ребра, отмеченные или сплошной (нулевой входной бит), или пунктирной (входной бит равен единице) линиями. Каждое ребро помечено n -разрядным блоком, который представляет собой группу из n кодовых символов, выдаваемых кодером в момент, когда входной бит источника переводит его из одного состояния в другое.

Нижеприведенный пример поясняет способ построения решетки для случая сверточного кодера на рис. 9.4.

Пример 9.3.3. Двухразрядный регистр сдвига может находиться в одном из четырех возможных состояниях: (00), (10), (11) и (01), причем запись состояния ведется с левой ячейки памяти. На рис. 9.5 приведена решетка, построенная согласно выше приведенному описанию. Например, ребро из состояния (10) в (01) отмечено сплошной линией, а в состояние (11) – пунктирной. Ребро, идущее из состояния (11) в (01), помечено кодовыми символами 01, поскольку при единицах в обеих ячейках и нулевым битом на входе верхний сумматор на рис. 9.4 вырабатывает нулевой выходной сигнал, а нижний выдает единицу и т.п. \square

В течение интервала каждого такта состояние кодера изменяется согласно движению вдоль какого-то ребра решетки, одновременно выдавая на выход кодовые символы, которыми помечено данное ребро. Отслеживая этот процесс по диаграмме, подобно представленной на рис. 9.5, необходимо производить скачки из правой колонки в тот же самый узел левой колонки на каждом последующем шаге. Для избежания этого повторим решет-

ку столько раз, сколько необходимо, используя правую колонку текущего шага в качестве левой для последующего шага. Тогда кодирование может быть отождествлено с движе-

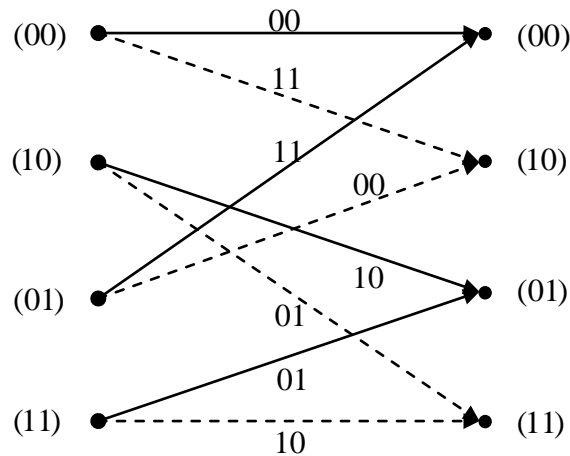


Рис. 9.5. Решетка кодера на рис. 9.4.

ем вдоль полученной *решетчатой диаграммы (trellis diagram)*, регулируемым текущим входным битом, который направляет кодер по сплошному или пунктирному ребру в зависимости от его конкретного значения: нуля или единицы. Для иллюстрации сказанного обратимся к рис. 9.6, на котором представлена решетчатая диаграмма кода из примера 9.3.1. Каждая последовательность-кандидат входных битов данных выбирает определенный путь по решетчатой диаграмме, который можно отследить, например, для последовательности $\{b_i\} = 10100100\dots$. Ее первый бит равен единице, что направляет кодер из узла (00) в узел (10), выдавая выходные кодовые символы 11. Второй бит, равный нулю, переводит кодер из узла (10) в (01), генерируя кодовые символы 10. Третий бит изменяет состояние кодера с (01) на (10), выдавая на выход кодовые символы 00 и т.д. Жирной чертой выделен результирующий путь, отвечающий кодовому слову 1110001011111011.

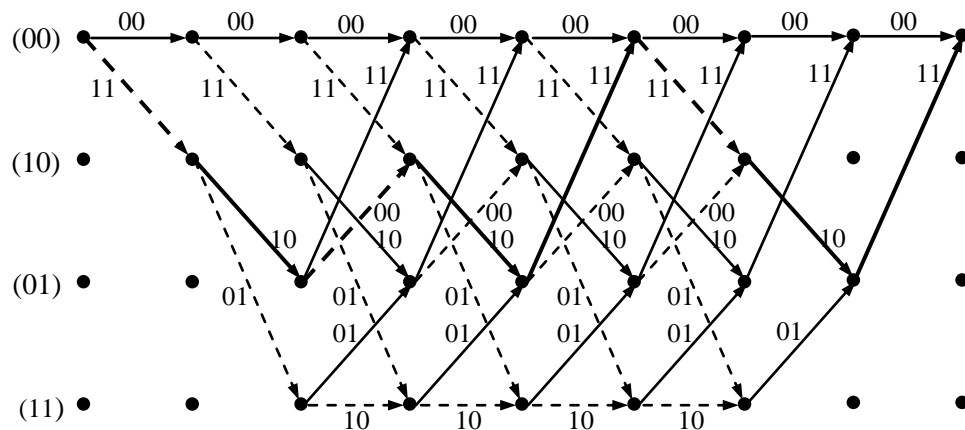


Рис. 9.6. Решетчатая диаграмма кодера на рис. 9.4.

Ранее уже было отмечено, что сверточный код можно трактовать как блочный код соответственной большой длины. Кодовые слова этого блочного кода есть просто различные пути по решетчатой диаграмме, минимальное хэммингово расстояние среди всех их пар дающее кодовое расстояние. В свою очередь, линейность кода упрощает задачу нахождения кодового расстояния: на основании утверждения 9.2.3 минимум расстояния между

путями определяется наименьшим весом Хэмминга среди всех ненулевых слов. Предположим теперь, что закодированный битовый поток заканчивается после некоторого большого (не меньше чем v_c) числа бит и дополняется $v_c - 1$ хвостовыми нулями для установки регистра в нулевое состояние. Реализованное на практике (одним из примеров служит cdmaOne) это дополнение не приведет к значительным затратам, если длина закодированного потока бит во много раз больше длины кодового ограничения. С другой стороны, введенное дополнение приведет к слиянию всех путей в один, который соответствует нулевому состоянию регистра, как показывает рис. 9.6 для кода из примера 9.3.1. Если теперь взять дополненный битовый поток, начинающийся с некоторого числа n_0 нулей, за которыми следует бит, равный «1», и переместить эти n_0 начальных нулей в конец потока, то это будет означать простой сдвиг соответствующего пути на n_0 шагов влево на диаграмме без изменения его веса. После такого сдвига путь отклонится от нулевого пути на самом первом шаге и сольется с ним не позднее, чем за n_0 шагов до последнего хвостового нуля. Вдоль этого пути могут произойти неоднократные возвращения к нулевому пути и последующие отклонения от него (см. рис. 9.7), каждое из которых только увеличивает вес пути, и поскольку целью является отыскание пути минимального веса, то любые отклонения от нулевого пути, кроме первого, должно игнорироваться. Суммируя все вышесказанное, заключаем, что для отыскания расстояния сверточного кода следует исследовать только пути, отклоняющиеся от нулевого пути в начале решетчатой диаграммы и не

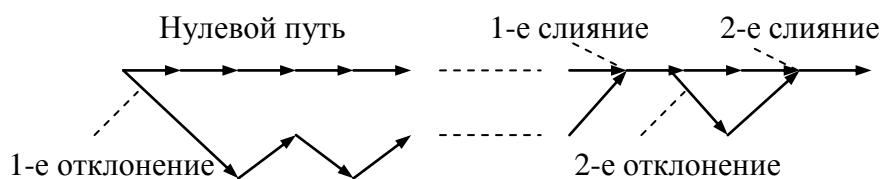


Рис.9.7 Иллюстрация отклонения и слияния путей.

отходящие от него после первого объединения. В теории сверточных кодов эта величина традиционно называется *свободным расстоянием* (*free distance*). Обозначив ее как d_f , можно увидеть, например, что среди всех путей с единственным отклонением от нулевого пути, изображенных на рис. 9.6, кодовое слово 11101100..., кодирующее битовый поток 100..., имеет минимальный вес, так что $d_f = 5$. Очевидно, что свободное расстояние d_f

гарантирует исправление любых $\left\lfloor \frac{d_f - 1}{2} \right\rfloor$ ошибок в символах (см. утверждение 9.2.2), од-

нако, как правило, исправляются также и многие образцы с большим числом ошибок. Существует только несколько примеров эффективных алгебраических правил сверточного кодирования. Большинство же известных сверточных кодов с хорошей корректирующей способностью были найдены в результате компьютерного поиска [31,33,93].

Вследствие специфичности алгоритмов кодирования аналитическое нахождение всех возможных весов слов (векторного спектра) произвольного сверточного кода оказывается не столь трудно, как в случае многих линейных блочных кодов. В частности, непосредственно из решетки (или, что эквивалентно, диаграммы состояний) строится система линейных уравнений, решение которой приводит к точному выражению для векторного спектра [2,7,93].

Выигрыш от кодирования, показывающий во сколько раз может быть уменьшена энергия сигнала на бит или мощность сигнала в результате кодирования при фиксированной вероятности ошибки, является универсальной мерой для оценки эффективности того

или иного кода. Это положение применительно к ортогональной сигнализации обсуждалось в параграфе 2.6 и отмечалось, что асимптотический выигрыш от кодирования в случае АБГШ канала определяется выигрышем в евклидовом расстоянии. При передаче с использованием бинарной ФМ любое различие в символах двух сигналов увеличивает квадрат евклидова расстояния на величину, равную $4E_s$, где E_s – энергия одного символа. Имеется пара слов сверточного кода, имеющая d_f различных символов, и отсутствует пара с меньшим различием (расстоянием Хэмминга). Следовательно, минимум квадрата евклидова расстояния между словами сверточного кода, передаваемыми с помощью бинарной ФМ, составляет $d_{\min, cc}^2 = 4d_f E_s$. В то же время (см. параграф 2.6) аналогичная характеристика при некодированной передаче составляет $d_{\min, u}^2 = 4E_b$, определяя асимптотический выигрыш от кодирования сверточным кодом, как

$$G_a = \frac{d_{\min, cc}^2}{d_{\min, u}^2} = \frac{d_f E_s}{E_b} = d_f R_c,$$

где R_c – скорость кода, измеренная в бит/сим. Для кода из примера 9.3.1 $R_c = 1/2$, $d_f = 5$, так что $G_a = 2.5$ или около 4 дБ. Напомним еще раз, что G_a выводится для АБГШ (не ДСК) канала, другими словами, для случая применения мягкого декодирования. Жесткое декодирование ухудшает эту цифру на 2..3 дБ в зависимости от параметров кода и отношения сигнал-шум на символ [31,33].

9.3.3. Алгоритм декодирования Витерби.

Как уже упоминалось, среди причин широкой популярности сверточных кодов особую роль играет существование эффективного алгоритма декодирования. Рассмотрение его начнем со следующего положения.

Утверждение 9.2.5. Максимально правдоподобное жесткое декодирование с исправление ошибок бинарного кода эквивалентно правилу минимума расстояния Хэмминга:

$$d_H(\hat{\mathbf{u}}, \mathbf{y}) = \min_{\mathbf{u} \in U} d_H(\mathbf{u}, \mathbf{y}) \Rightarrow \hat{\mathbf{u}} \text{ принимается за принятое слово.} \quad (9.9)$$

Как можно видеть, данное правило очень похоже на (2.3) с одним изменением: в случае ДСК расстояние Хэмминга заменяет Евклидово расстояния, адекватное для АБГШ канала. Для доказательства (9.9) достаточно заметить, что соотношение (9.1), переписанное в виде

$$P(\mathbf{y}|\mathbf{u}) = p^{d_H(\mathbf{u}, \mathbf{y})} (1-p)^{n-d_H(\mathbf{u}, \mathbf{y})} = \left(\frac{p}{1-p} \right)^{d_H(\mathbf{u}, \mathbf{y})} (1-p)^n$$

есть просто переходная вероятность ДСК, т.е. вероятность трансформации ДСК посланного кодового вектора \mathbf{u} длины n в двоичное наблюдение \mathbf{y} . Поскольку перекрестная вероятность ДСК $p < 0.5$, то переходная вероятность $P(\mathbf{y}|\mathbf{u})$ есть убывающая функция расстояния Хэмминга между наблюдением \mathbf{y} на выходе ДСК и кодовым вектором \mathbf{u} . Следовательно, максимально правдоподобной оценкой $\hat{\mathbf{u}}$ кодового слова является слово, ближайшее по Хэммингу к \mathbf{y} . \square

Прямая реализация правила (9.9) для произвольного кода заключается в сравнении M расстояний Хэмминга между наблюдением \mathbf{y} и всеми кодовыми словами. Поскольку, как правило, M достаточно велико, то такое решение может оказаться не осуществимым.

С другой стороны, для сверточных кодов вследствие их благоприятной структуры МП декодирование не представляет особой трудности, по крайней мере, для умеренных значений длины кодового ограничения.

Процедура декодирования Витерби реализует максимально правдоподобную стратегию в рекуррентной форме, последовательно осуществляя поиск пути по решетчатой диаграмме, ближайший к двоичному наблюдению \mathbf{y} . Каждый новый шаг декодирования начинается с приема следующей группы из n символов наблюдения. На i -м шаге декодер вначале вычисляет расстояние между n пришедшими символами наблюдения и каждым ребром решетчатой диаграммы, а затем увеличивает расстояния всех путей, рассчитанных за $i-1$ предшествующих шагов. Для произвольного кода расстояния можно вычислять аналогично сразу же по приходу новых символов наблюдения, однако рекуррентная природа сверточных кодов позволяет осуществить эту обычную вычислительную работу более экономно по причине отбрасывания многих путей на каждом шаге.

Рассмотрим все пути, проходящие через фиксированный узел A на i -м шаге, как это показано на рис. 9.8. Продолжение любого пути после i -го шага не зависит от маршрута прихода в A , так что все различные пути, достигающие A , могут в дальнейшем считаться слившимися в один. Однако это означает, что из всех путей, проходящих через A и продолжающихся после i -го шага, один, обладающий минимальным расстоянием от \mathbf{y} вплоть до i -го шага, навсегда останется ближайшим к \mathbf{y} , поскольку общее продолжение внесет равный вклад во все расстояния. Тогда за чем продолжать вычисление расстояний

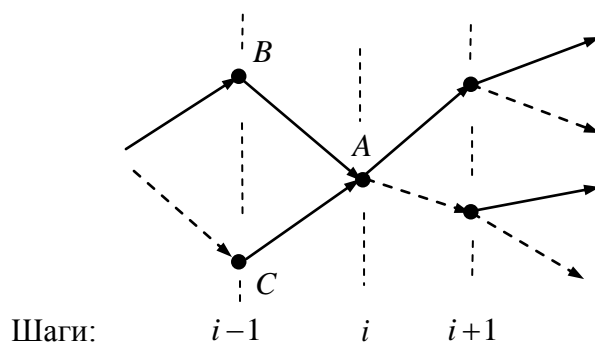


Рис. 9.8. Пути, проходящие через узел A на i -м шаге.

для остальных, если совершенно очевидно, что у них нет шансов оказаться в итоге ближайшими к наблюдению? Вместо этого отбросим их, оставляя только один путь с минимальным расстоянием, достигший узла A . Последний называется *выжившим* (*survivor*) путем и на данный момент будем считать, что имеется только один выживший путь для каждого из узлов решетчатой диаграммы на i -м шаге (см. комментарий ниже). Текущее, т.е. вычисленное по всем символам наблюдения вплоть до i -го шага, расстояние узла A выжившего пути от наблюдения \mathbf{y} называется *метрикой узла* (*node metric*) A .

Теперь вспомним, что имеются только два ребра, входящие в любой узел, что и показывает рис. 9.8 для некоторого узла A . Два ребра, входящие в него, исходят из узлов B , C предшествующего этапа и, следовательно, продолжают выжившие пути B и C . Можно вычислить расстояния двух путей, достигающих A , в результате вычисления только *метрик ребер*, т.е. расстояний ребер от принятой группы из n наблюдаемых символов, и суммирования их с метриками узлов B и C . Путь с меньшим расстоянием считается выжившим для узла A и запоминается вместе с его расстоянием (метрикой узла), тогда как другой отбрасывается. По выполнению этих операций для всех узлов решетчатой диаграммы декодер переходит к следующему этапу.

Резюмируя сказанное, декодер Витерби на каждом шаге вычисляет метрики ребер, добавляет их ко всем метрикам узлов, накопленных ранее, а затем отбрасывает большее расстояние из двух, ведущих к каждому узлу. Поскольку всего имеется 2^{v_c-1} узлов (т.е. состояний регистра), сложность декодера определяется только длиной кодового ограничения v_c и остается фиксированной вне зависимости от теоретически неограниченного числа кодовых слов (путей).

Возвращаясь к сделанному предположению о единственности выживающего пути для каждого узла, отметим, что поскольку расстояние Хэмминга целочисленно, то всегда существует вероятность того, что два пути, ведущие к одному и тому же узлу, обладают одинаковыми текущими расстояниями относительно y . Возможны различные стратегии преодоления неопределенности этого типа. Одна из них использует случайный выбор: присвоение реверса честной монеты одному из путей и провозглашение его выжившим в случае выпадания реверса при бросании монеты. Подобное действие, конечно, нарушает оптимальность декодирования, хотя сопровождающие его энергетические потери, как правило, незначительны. Альтернативой является признание обоих конкурирующих путей выжившими и запоминание их до тех пор, пока дальнейшие шаги не разрешат неопределенность. Последняя стратегия сохраняет оптимальность декодирования ценой увеличения требуемой памяти.

Пример 9.3.4. Рассмотрим декодирование наблюдения $y = 0100110000\dots$ для кода из примера 9.3.1. Рис. 9.9 иллюстрирует данный процесс с указанием метрик узлов, расположенных непосредственно около них, и кадров, содержащих пары принятых символов наблюдения на текущем шаге. Декодер начинает процесс в предположении нулевого (т.е. 00) начального состояния кодирующего регистра. Начальные $v_c - 1 = 2$ шага отвечают переходному состоянию кодирующего регистра, когда только одна ветвь входит в каждое состояние (см. рис. 9.9) и все пути оказываются выжившими. На первом шаге декодер сравнивает первую группу из $n = 2$ символов наблюдения с двумя ребрами, выходящими из состояния (00). Согласно их расстоянию Хэмминга от группы символов 01 как сплошное, так и пунктирное ребра решетки приобретают метрику, равную 1, которая приведена рядом с ними. Следовательно, метрики двух узлов, до которых дошли ребра, становятся равными единице. На следующем шаге расстояние измеряется между второй группой наблюдаемых символов 00 и двумя парами ветвей, исходящих из узлов (00) и (10), приводя к метрикам, которыми помечены ветви. После добавления к метрикам узлов предшествующего этапа они обновляют метрики узлов (00) и (10), а также образуют метрики еще двух узлов (01) и (11). Начиная с третьего шага в любой узел решетчатой диаграммы рис. 9.9 входят по две ветви, означающие, что декодер должен решить какая из них выживает. Начиная с этого шага, откажемся от показа метрик ребер, чтобы не перегружать рисунок. Как видно, на третьем шаге имеются два пути, ведущие к узлу (00). Их расстояния от наблюдаемых символов 010011 составляют величины, равные 3 и 2 соответственно. Первый из них не выживает, и декодер отбрасывает его вместе с соответствующей метрикой, что отражено перечеркиванием пути и отсутствием его на рисунке следующего этапа. Второй путь является выжившим и запоминается со своей метрикой до следующего шага. Аналогичным образом декодер определяет выжившие пути и для остальных узлов. В ходе последующих шагов процедура декодирования продолжается подобно рассмотренному, сохраняя в памяти только $2^{v_c-1} = 4$ выживших путей, и каждая диаграмма на рис. 9.9 изображает только пути, выжившие на предшествующем этапе.

На 7-м шаге декодер впервые сталкивается с проблемой неоднозначности: два пути приходят в узел (01) с одинаковыми расстояниями, как и в узел (11). Выбор выжившего пути, иллюстрируемый рисунком, отражает его реализацию посредством подбрасывания монеты. Аналогичные события происходят на 8-м и 9-м шагах. Читатель может убедиться (задача 9.19), что любое альтернативное решение неопределенности не изменит окончательного результата декодирования за исключением числа шагов, после которых воз-

можно первое считывание декодированных бит (см. ниже).

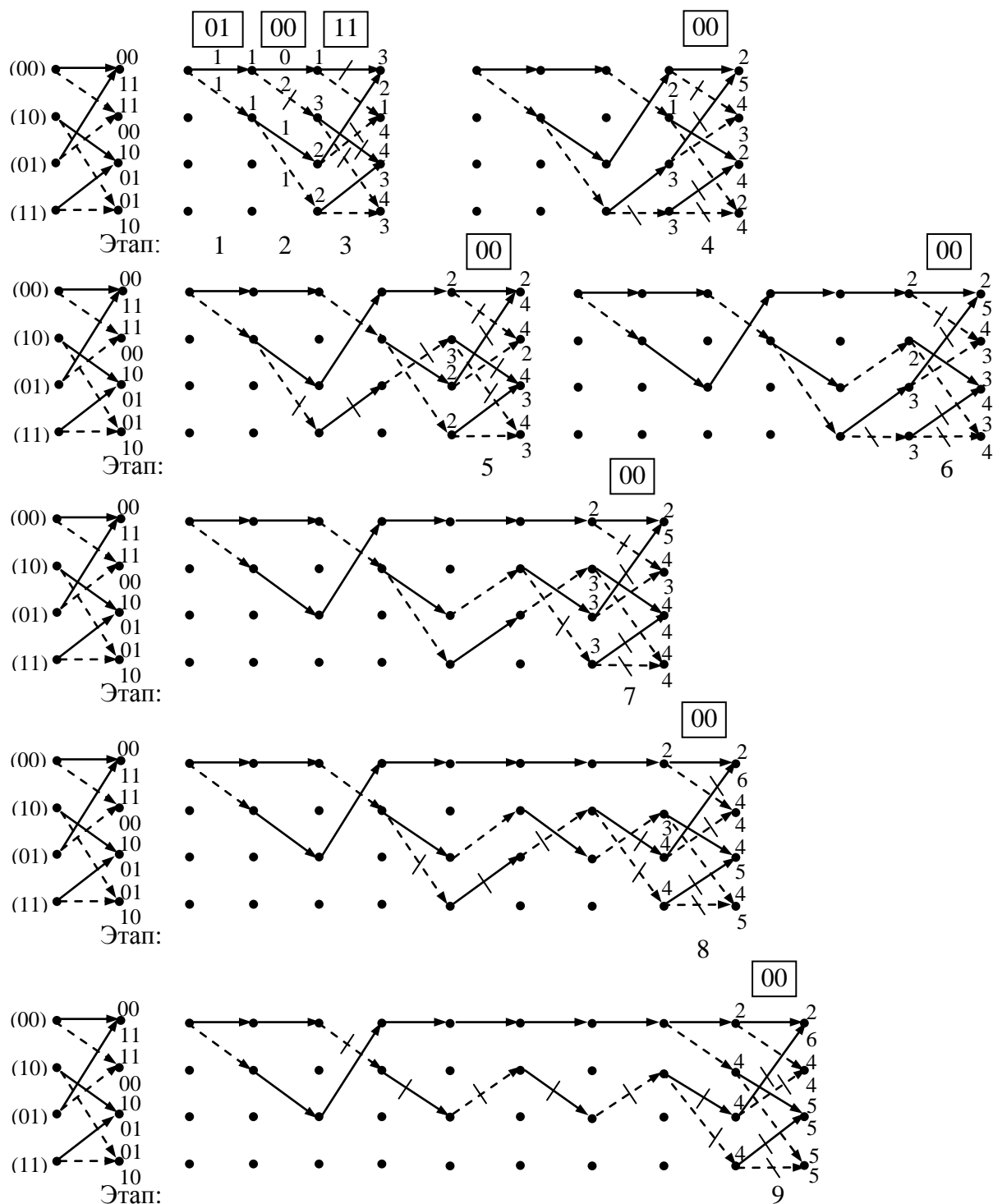


Рис. 9.9 Динамика декодирования сверточного кода из примера 9.3.1.

Ситуация, возникшая после девятого этапа, является чрезвычайно важной: все пути оказываются слившимися вплоть до седьмой группы наблюдаемых символов. Чтобы ни произошло в дальнейшем, эта часть всех слившихся путей навсегда останется общей, означая, что биты данных, соответствующие этому участку, как раз теперь могут быть декодированы. Таким образом, декодер выдает декодированные данные вида 1000000. Сравнимая кодовое слово $\mathbf{u} = 11101100000000\dots$, отвечающее указанному потоку бит данных, с наблюдением $\mathbf{y} = 0100110000$, можно заметить, что расстояние Хэмминга между ними равно 2, и если переданное слово действительно было таким, каким его определил деко-

дер, то были исправлены две ошибки в полном соответствии с величиной свободного расстояния $d_f = 5$. Аналогичные ситуации будут возникать и в дальнейшем, позволяя декодеру выдавать декодированные биты в процесс обработки потока наблюдаемых символов.

Очевидно, что выдача декодированных данных в случайные моменты слияния выживших путей, как в вышеприведенном примере, выглядит не практично и более желательной представляется процедура регулярной выдачи данных. Неоднократно было проверено экспериментально и путем моделирования, что в течении i -го шага декодирования сливающаяся часть всех выживших путей почти никогда не заканчивается после числа бит данных $i - 5v_c$, так что решение о любом бите может регулярно приниматься с задержкой в $5v_c$ [94].

Очень важным свойством сверточных кодов, делающим их еще более привлекательными, является сравнительная простота реализации мягкого декодирования. В самом общем случае блочного кода с M кодовыми словами осуществление мягкого декодирования означает непосредственное вычисление M евклидовых расстояний или корреляций, что представляется достаточно сложным алгебраически по сравнению с синдромным декодированием. При значительных величинах M , типичных для множества приложений, данный факт часто превращает указанную задачу в полностью не осуществимую. В то же время платой за упрощения приемника, использующего жесткие решения, служат энергетические потери, оцениваемые для АБГШ канала в 2 дБ: цифрой, рассматриваемой сегодня, как достаточно значительной. Обратимся вновь к алгоритму Витерби и просто заменим расстояние Хэмминга евклидовым. Очевидно, что данная операция превратит декодирование в его вариант с мягкими решениями, оптимальный для АБГШ канала. Тогда метрики ребер и узлов в точности отвечают соответствующим евклидовым расстояниям (или корреляциям). Подобная модификация ни в малейшей степени не затронет реализационные преимущества алгоритма Витерби. Действительно, метрики узлов вычисляются, как и ранее рекурсивно, путем пошагового увеличения за счет метрик ветвей, а путь, входящий в узел с худшей метрикой, чем альтернативный, снова может быть отброшен на каждом шаге, как не выживший.

Конечно, цифровая реализация декодеров является наиболее предпочтительной, подразумевая квантование входного наблюдения. Общепринятой является тенденция классификации декодеров двоичных кодов, использующих квантование более чем на два уровня, как декодеры с мягким декодированием. Углубленный анализ показывает, что в большинстве случаев 3-битовое (8-ми уровневое) квантование достаточно для достижения почти потенциальных (характерных для непрерывной обработки) параметров [94].

9.3.4. Приложения.

В настоящее время известно множество эффективных сверточных кодов и диапазон их использования в телекоммуникации чрезвычайно широк. В частности, 2G стандарт cdmaOne и 3G стандарт UMTS использует коды с длиной кодового ограничения $v_c = 9$ и скоростями $R_c = 1/2$, $R_c = 1/3$, обеспечивая асимптотический выигрыш от кодирования порядка 7.8 дБ [18,69,92]. 3G стандарт cdma2000 в дополнение к упомянутым выше использует код с параметрами $v_c = 9$, $R_c = 1/4$. Помимо их собственного значения сверточные коды служат основой для построения турбо-кодов, приближающих надежность передачи данных к пределу Шеннона, краткое обсуждение которых представлено ниже.

9.4. Турбо–коды.

Как уже упоминалось ранее в параграфе 9.1, несмотря на огромные усилия, прикладываемые в течение десятилетий с момента начала информационной теории Шеннона, попытки отыскания регулярных правил кодирования, гарантирующих надежность передачи близкую к пропускной способности канала, оставались безуспешными. В свете этого открытие в 1993 г. турбо–кодов явилось впечатляющим прорывом, первоначально встреченным телекоммуникационным сообществом с объяснимым недоверием. Однако в настоящее время турбо–коды широко признаны в качестве эффективного средства обеспечения высококачественной связи, особенно в условиях малого отношения сигнал–шум на бит. Точная и компактная теория этого класса кодов остается скорее разрабатываемой, чем действительно уже существующей. До сих пор интуиция в сочетании с обширными компьютерными поисками играют значительную роль в получении многих результатов.

9.4.1. Кодеры турбо-кодов.

Другим названием турбо-кодов является параллельно сочетающиеся сверточные коды, что отражает основную идею алгоритма кодирования: два параллельных сверточных кодера (компонента) кодируют один и тот же битовый поток источника [95,96]. Кодеры компонентов, как правило, являются идентичными, т.е. обладают одинаковыми длинами кодового ограничения и множествами порождающих полиномов. Первый из них кодирует данные непосредственно, тогда как перед поступлением на второй кодер битовый поток данных подвергается перемежению. Данная операция состоит в перестановке по псевдослучайному закону бит данных в пределах блока фиксированной длины L .

Как было отмечено в параграфе 9.3.1, среди сверточных кодов, основанных на структуре с КИО (см. рис. 9.3), отсутствует эквивалентность между систематическими и несистематическими кодами, причем несистематические, как правило, являются более мощными в смысле дистанционных свойств. В то же время принцип декодирования, связанный с турбо-кодами и придающий им особую привлекательность, в большей степени отвечает кодерам с систематическими компонентами. Применение структуры с *бесконечным импульсным откликом* (БИО) (*infinite impulse response* (ИИР)) вместо КИО открывает путь к формированию систематического сверточного кода, содержащего те же слова и, следовательно, обладающего такими же дистанционными свойствами, как и несистематический. Для объяснения этого факта начнем с рассмотрения кодера, изображенного на рис. 9.3, который описывается множеством порождающих полиномов $g_l(z)$, $l = 1, 2, \dots, n$, и установим однозначное соответствие между двумя битовыми потоками $b(z)$ и $b_1(z)$ как $b(z) = b_1(z)g_1(z)$. Обратимся затем к соотношению (9.7) и увидим, что для входного потока бит $b_1(z)$ оно может быть представлено в виде

$$u_l(z) = b_1(z)g_l(z) = \frac{b(z)}{g_1(z)}g_l(z), \quad l = 1, 2, \dots, n, \quad (9.10)$$

показывающем, что структура, способная осуществить деление битового потока источника $b(z)$ на $g_1(z)$ перед подачей на вход схемы с КИО, изображенной на рис. 9.3, будет кодировать $b(z)$ в те же сверточные слова, как и собственный КИО кодер. Единственное отличие состоит в том, что кодовое слово, присваиваемое ранее $b_1(z)$, теперь переприсваивается $b(z)$, что, учитывая равные права любого битового потока источника, не играет роли, если декодер знает этот новый порядок соответствия. Однако после подобной операции все кодовые слова становятся систематическими, поскольку $u_1(z) = b(z)$. Регистр с обратной связью, представленный на рис. 9.10, реализует деление на полином $g_1(z) = g_r^1 z^r + g_{r-1}^1 z^{r-1} + \dots + 1$, где $r = v_c - 1$. Действительно, согласно общему правилу,

применимому к любой линейной системе (одним из примеров этого служит (8.26)), в z -области передаточная функция $h_f(z)$ замкнутой петли обратной связи выражается (учитывая эквивалентность операций сложения и вычитания в двоичной арифметике) как [2,7]

$$h_f(z) = \frac{h(z)}{1 + \beta(z)h(z)},$$

где $h(z)$ и $\beta(z)$ – передаточные функции разомкнутой петли и схемы обратной связи соответственно. Рассматривая выход крайнего слева сумматора на рис. 9.10 как выход замкнутой петли, имеем $h(z) = 1$, $\beta(z) = g_1^1 z + g_2^1 z^2 + \dots + g_r^1 z^r$, так что $h_f(z) = 1/g_1(z)$. Поскольку поделенный битовый поток $b(z)/g_1(z)$ непосредственно подается на вход регистра, то логическая схема формирования проверочных символов, определяемая полиномами $g_l(z)$, $l = 2, \dots, n$, может быть подсоединена к тому же регистру для дальнейшего обычного сверточного кодирования. Резюмируя, в систематическом кодере с обратной связью би-

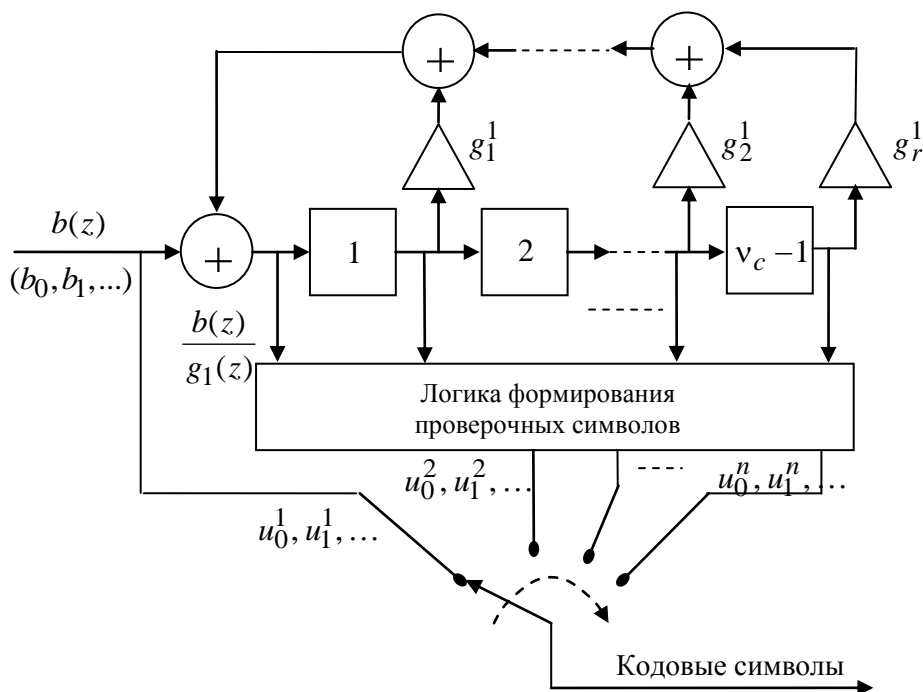


Рис. 9.10. Систематический сверточный кодер.

товый поток источника непосредственно определяет поток бит данных $u_1(z)$, тогда как потоки проверочных символов $u_l(z)$, $l = 2, \dots, n$ формируются той же логической схемой, как и на рис. 9.3, но подсоединенной к регистру с обратной связью. Рис. 9.11 иллюстрирует пример преобразования несистематического кодера, изображенного на рис. 9.4, в систематический.

Полная структура турбо-кодера включает два подобных систематических кодера и перемежитель, как это показано на рис. 9.12 для кодеров компонент со скоростью $R_c = 1/2$. В этой структуре отбрасывается «прямой» поток бит данных со второго компонента кодера, а используются только потоки его проверочных бит. Таким образом, если образующие сверточные коды имеют скорость $R_c = 1/2$, то результирующий турбо-код содержит два проверочных символа u_i^2, u_i^3 на один бит данных b_i , т.е. имеет скорость

$R_c = 1/3$. При желании его скорость может быть увеличена до $1/2$ путем передачи только одного проверочного символа на бит данных, выбираемого поочередно с кодеров компо-

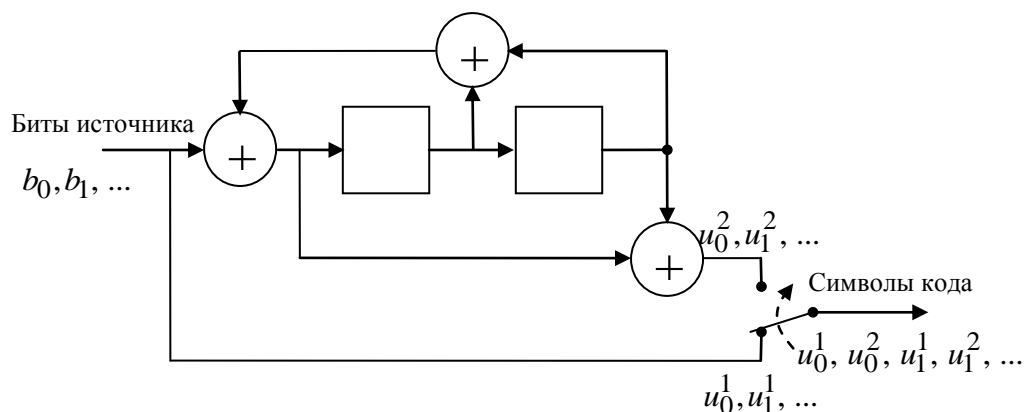


Рис. 9.11. Систематический сверточный кодер для кода из примера 9.3.1.

нент. Перемежитель перемешивает биты данных в пределах блока длины I . Как правило, для установки кодера первого компонента в нулевое состояние используется $v_c - 1$ хво-



Рис. 9.12. Турбо-кодер.

стовых бит, и турбо-код трактуется как блочный код, содержащий I информационных бит.

9.4.2. Процедура итеративного декодирования.

Несмотря на то, что турбо-код состоит из двух сверточных кодов, его оптимальное декодирование не может быть реализовано в виде двух независимых процедур Витерби, поскольку пути на решетчатых диаграммах его компонентов связаны друг с другом вследствие кодирования одного и того же (хотя и с перемежением) потока данных. Открыватели турбо-кодирования [95,96] предложили использовать итеративную версию правила максимума апостериорной вероятности (МАП) (*maximum a posteriori probability* (MAP)). Представленное ниже описание касается только основной ее идеи, отсылая заинтересованного читателя за деталями к [96,97].

Апостериорные вероятности $p(b_i = 0|y)$ и $p(b_i = 1|y)$ характеризуют правдоподобность одного или другого значения i -го бита данных, вычисленную на основе вектора на-

блюдения y . Они содержат в себе полную информацию о b_i , доставляемую y и извлекаемую из него. В случае достоверной связи одна из этих вероятностей близка к единице, тогда как вторая почти равна нулю. Действительно, МАП декодер будет выдавать на выход в качестве оценки \hat{b}_i i -го бита данных такое значение, которому отвечает большая апостериорная вероятность, следуя правилу

$$\Lambda_i = \frac{p(b_i = 0|y)}{p(b_i = 1|y)} \begin{matrix} \hat{b}_i = 0 \\ > 1 \\ \hat{b}_i = 1 \\ < \end{matrix} \quad (9.11)$$

Рекуррентный алгоритм вычисления апостериорной вероятности значения бита разработан в [98]. В рассматриваемой ситуации бит данных b_i фактически присутствует в наблюдениях вследствие применения систематического кодирования. При непосредственной реализации рекуррентный МАП алгоритм вычисляет апостериорную вероятность b_i , используя все ранее полученные наблюдения, включающие данный бит. После принятия всех отсчетов наблюдения, он возвращается назад и пересматривает результаты, содержащие информацию, извлеченную из наблюдений, поступивших после b_i . Таким образом, после прохождения вперед и возвращения назад могут быть найдены апостериорные вероятности для всех бит данных. Для реализации этого алгоритма необходимо знать решетку кода, каналные переходные вероятности и распределение априорной вероятности $q(b_i)$ для каждого бита данных. Очевидно, что не возникнет никаких проблем при реализации алгоритма в отношении каждого из двух компонент сверточных кодов, образующих турбо-код. Однако основная идея турбо-кода состоит в кодировании одних и тех же пе-

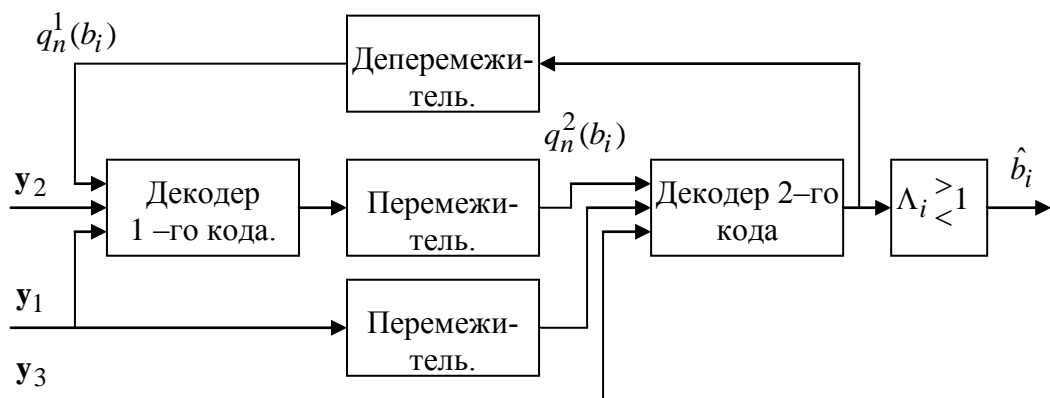


Рис. 9.13. Итеративный турбо-декодер.

ремеженных данных двухкомпонентным кодом, так что информация о b_i может быть восстановлена совместно по обоим из них. Итерационный процесс, организованный, как показано на рис. 9.13, удовлетворяет данному условию.

Полный вектор наблюдения на приемной стороне может быть разложен на три составляющих: $y_l = \mathbf{u}_l + \mathbf{n}_l$, $l = 1, 2, 3$, где $\mathbf{u}_1 = (u_0^1, u_1^1, \dots, u_{M-1}^1)$, $u_i^1 = b_i$ – вектор бит данных b_i непосредственно представлен в любом кодовом слове вследствие систематического характера турбо-кодов, $\mathbf{u}_2, \mathbf{u}_3$ – вектора избыточных символов первого и второго кодеров соответственно, а \mathbf{n}_l – вектор независимых выборок шума.

На первом шаге декодер первого компонента кода вычисляет апостериорные веро-

ятности $p_1(b_i|y_1, y_2)$ всех I бит данных b_i , используя наблюдения y_1, y_2 , связанные с проверочными символами и символами данных этого кода. В качестве исходной информации используется априори равномерное распределение $q_1^1(b_i) = 1/2, b_i = 0,1$, поскольку естественно считать равновероятными все образцы бит данных. После этого декодер второго кода, вычисляющий апостериорные вероятности $p_1(b_i|y_1, y_3)$, может основываться не только на соответствующих наблюдениях y_1, y_3 , но также на информации, доставляемой первым декодером, используя его апостериорное распределение $p_1(b_i|y_1, y_2)$ в качестве априорного: $q_1^2(b_i) = p_1(b_i|y_1, y_2)$. В результате имеем первое приближение $p_1(b_i|y)$ апостериорного распределения $p(b_i|y)$. Поскольку на этом шаге первый декодер не снабжается информацией от второго, то это положение исправляется на второй итерации, когда первый декодер снова декодирует первый компонент кода, но при использовании априорного распределения $q_2^1(b_i) = p_1(b_i|y)$. Продолжая подобным образом, после n -го шага следующая аппроксимация $p_n(b_i|y)$ распределения $p(b_i|y)$ формируется вторым декодером и используется первым декодером в качестве следующего априорного распределения $q_{n+1}^1(b_i)$ для выработки $p_{n+1}(b_i|y_1, y_2)$. Последнее в свою очередь, используется вторым декодером как следующее априорное распределение $q_{n+1}^2(b_i)$ для выработки следующей аппроксимации ожидаемой апостериорной вероятности $p_{n+1}(b_i|y)$ и т.п. Поскольку перемежитель меняет порядок следования бит данных перед подачей их на вход второго кодера, перемежители аналогичным образом воздействуют на наблюдения y_1 и априорное распределение $q_n^2 = p_n(b_i|y_1, y_2)$, подающиеся на второй декодер. Аналогично, деперемежитель восстанавливает первоначальный порядок бит в схеме обратной связи, передавая $p_n(b_i|y)$ с выхода второго декодера на вход первого. С учетом этих преобразований все данные, обрабатываемые обоими декодерами, соединяются должным образом.

Обширное моделирование экспериментально подтвердило сходимость этих итераций, однако теоретическое обоснование до сих пор остается под вопросом.

9.4.3. Характеристики.

Как упоминалось ранее, турбо-коды представляют собой первый пример регулярных кодов, обеспечивающих надежную передачу данных по каналу с ограниченной полосой со скоростью, близкой к пропускной способности, и малой энергии на бит. Для иллюстрации этого на примерах рассмотрим первоначально некоторые фундаментальные ограничения при передаче данных с бинарной ФМ. На основании теоремы отсчетов любой полосный сигнал с полосой W представим вектором размерности $2WT$ (см. параграф 2.3). В случае использования бинарной ФМ каждый компонент такого вектора может принимать только два значения, откуда следует, что в пределах заданного частотно-временного ресурса WT число M бинарных ФМ сигналов подчиняется ограничению $M < 2^{2WT}$, или, что эквивалентно, может быть передано не более чем $2WT$ бит данных. Это ограничивает скорость передачи, достижимую при бинарной ФМ в пределах полосы W , как $R = (\log M)/T \leq 2W$, что в свою очередь определяет границу скорости на один герц: $R/W \leq 2$ бит/сек/Гц. Рассмотрим теперь двоичный код со скоростью $R_c = 1/2$, означающей, что только каждый второй компонент сигнального вектора переносит данные, тогда как остальные предназначены для проверки символов (на один бит приходится два отсчета сигнала), так что отношение между скоростью и полосой будет $R/W = 1$ бит/сек/Гц.

Обратившись к границе Шеннона (1.2), можно заметить, что минимальное значение энергии на бит, нормированной к спектральной плотности шума E_b/N_0 (что составляет половину отношения сигнал-шум по мощности на бит), необходимой для обеспечения безошибочной передачи по АБГШ каналу с такой скоростью, составляет 0 дБ. Однако, имея дело не с произвольным гауссовским каналом, а с каналом, входные символы которого ограничены алфавитом бинарной ФМ (гауссовский канал с бинарным входом). Это ограничение увеличивает минимальное отношение E_b/N_0 , отвечающее $R/W=1$ бит/сек/Гц, до 0.19 дБ [97]. Турбо-код с длиной кодового ограничения 5 и длиной блока 65536, предложенный в [95,96], обеспечивает вероятность ошибки на бит $P_b < 10^{-5}$ (это значение часто используется в качестве практического критерия безошибочной работы беспроводной системы) при $E_b/N_0 \approx 0.7$ дБ, т.е. уступает пределу Шеннона только около 0.5 дБ. В момент опубликования этих работ данные результаты казались фантастическими, поскольку длинная и безуспешная история заставила многих экспертов поверить, что отыскание регулярных правил кодирования, позволяющих работать вблизи границы Шеннона, является безнадежной тратой времени. Вслед за оригинальными работами [95,96] были открыты другие эффективные турбо-коды, как и последовательные коды конкатенации (см. библиографию в [97]).

Следует отметить, что в асимптотическом варианте ($E_b/N_0 \rightarrow \infty$) поведение турбо-кодов не отличается от сверточных при той же скорости и памяти, поскольку они не обладают никаким преимуществом в смысле минимума расстояния. Обратившись к (2.23), можно увидеть, что асимптотически (с ростом отношения сигнал-шум) эффект размножения n_{\min} , т.е. числа сигналов с минимумом евклидова расстояния d_{\min} относительно переданного сигнала, играет второстепенную роль по сравнению с собственно значением d_{\min} вследствие экспоненциального уменьшения Q -функции с ростом отношения сигнал-шум (чтобы убедиться в этом, достаточно взять логарифм от P_e). По этой причине зависимость P_e от E_b/N_0 чуть раньше или чуть позже приобретет характер ограничения, определяемый d_{\min} , которое аналогично для любого кода с одинаковыми значениями минимума расстояния. Данный эффект, однако, происходит при значениях отношения сигнал-шум, обеспечивающих очень малую вероятность ошибки на бит, которые лежат значительно ниже диапазона практических требований. Объяснение тому, почему турбо-коды гарантируют настолько высокое качество работы при низком отношении сигнал-шум, заключается не в их большом значении минимума расстояния, а скорее в относительно малом числе слов, находящихся друг от друга на малом расстоянии, в частности малом увеличении n_{\min} в (2.23). Подобное перераспределение расстояний в направлении большего числа слов, находящихся на большом расстоянии, по сравнению со сверточными кодами объясняется псевдослучайным перемежением бит данных, кодируемых вторым компонентом кодера. Если образец данных оказывается неудачным с точки зрения формирования кодового слова с малым весом первым компонентом кода, то в результате перестановки он может оказаться достаточно отличным от исходного, что позволит второму компоненту сформировать слово значительно большего веса.

9.4.4. Приложения.

Несмотря на короткую историю турбо-коды в настоящее время нашли широкое применение и входят в состав спецификаций многих систем. Наибольший интерес в рассматриваемом контексте представляет их вовлеченность в стандарты мобильных систем третьего поколения. Спецификация UMTS включает турбо-коды со скоростью 1/3, основанные на двух компонентах сверточных кодов с длиной кодового ограничения 4 и перемежении переменной длины в диапазоне от 40 до 5114 [92,97]. В стандарте cdma2000 так-

же используются двухкомпонентные турбо–коды с длиной кодового ограничения 4 и областью перемежения, лежащей в диапазоне от 250 до 4090. Соответствующее выкалывание позволяет получить скорости 1/2, 1/3, 1/4 или 1/5 [69,97].

9.5. Канальное перемежение.

Проведенные выше анализ опирались на модель канала без памяти, в котором искажающие действия канальной помехи, наносимые кодовым символам, были независимыми. В реальных беспроводных каналах с эффектами затенения и замирания (см. параграф 3.5) упомянутые аддитивные эффекты без памяти дополняются мультипликативными: сравнительно медленными и распространяющимися на много кодовых символов спорадическими падениями в уровне принимаемого сигнала. Метод, описанный ниже, является универсальным, не зависящим от мягкого или жесткого метода декодирования, но для большей прозрачности обсуждения положим, что используется процедура жесткого декодирования. Если АБГШ является единственной канальной помехой, то все ошибки в символах являются независимыми и случайно распределенными на интервале кодовых слов. Падение уровня сигнала вследствие замирания вызывает группирование ошибочных символов в пакеты или пачки. Очевидно, что если длина B пакета не превосходит корректирующей (обнаруживающей) способности t_c (t_d), то кодер без труда исправит (обнаружит) B ошибок, поскольку вид конкретного образца ошибок в пределах способности кода к контролю над ошибками не играет роли. Однако природа замирания часто приводит к редким, но достаточно длинным пакетам ошибок, так что выполнение условия $B \leq t_c$ потребует кодов с большим числом избыточных символов, т.е. малой скоростью и расточительным использованием полосы. Эффективной и осуществимой на практике альтернативой является очень популярное т.н. *канальное перемежение* (*channel interleaving*). Суть последней заключается в том, что перед передачей символы кодового потока перемешиваются с целью такого их распределения, при котором близко расположенные друг к другу по возможности разносятся на максимально большое расстояние, а те, которые исходно далеки, становятся близкими. На приемной стороне осуществляется процедура деперемежения, в результате которой все кодовые символы возвращаются на свои исходные позиции. Если в канале возникает пакет ошибок длины B , то искаженные им символы после деперемежения оказываются далеко разнесенными друг относительно друга, как если бы ошибки были независимыми. Если используется блочный код соответствующей длины и расстояния, то с большой вероятностью эти ошибки окажутся в различных словах и будут исправлены декодером. При сверточных кодах вероятность их исправления снова окажется хорошей, поскольку подобные коды исправляют многие образцы ошибок, включая и те, которые превышают свободное расстояние, если они не группируются в слишком плотные пакеты. Простейшей реализацией рассматриваемого метода служит перемежение в рамках блока, заключающееся в записи символов по строкам квадратной матрицы, а считывание их по столбцам. Очевидно, что деперемежение символов осуществляется в обратном порядке.

Перемежение является составной частью большинства современных беспроводных систем связи, включая все стандарты мобильного радио 2–го и 3–го поколений.

Задачи.

9.1. Для передачи $M = 32$ сообщений используется двоичный блочный код длины $n = 9$. Сколько проверочных символов содержится в этом коде? Какова его скорость? Сколько возможно лишних двоичных векторов наблюдения?

9.2. Двоичный блочный код характеризуется минимумом расстояния Хэмминга $d_H = 7$. Каким будет минимальное расстояние Евклида, если двоичные символы передаются с помощью частотной модуляции не перекрывающимися импульсами энергии E_s ?

9.3. Для передачи данных по ДСК используется двоичный блочный код $U = \{10101, 00011, 11000, 01110\}$. Принят вектор наблюдения $Y = (00110)$. В какое кодовое слово будет декодирован вектор при исправлении ошибок? Каков будет ответ, если $Y = (11011)$? Какова исправляющая и обнаруживающая способность этого кода?

9.4. Доказать утверждение 9.2.2.

9.5. Каким должно быть минимальное расстояние двоичного кода, исправляющего вплоть до t_c ошибок и, кроме того, обнаруживающего вплоть до $t_d > t_c$ ошибок?

9.6. Определить число двоичных векторов, обладающих весом не более чем t_d (объем двоичной сферы радиуса t_d).

9.7. Используя результаты задачи 9.6, доказать границу Гильберта: двоичный блочный код, обнаруживающий до t_d ошибок, существует всегда, когда

$$(M - 1) \sum_{i=0}^{t_d} C_n^i < 2^n,$$

где M, n – число кодовых слов и длина кода соответственно.

9.8. Вычислить результат следующих операций над двоичными полиномами:

$$f(z) = (z^3 + z^2 + 1)(z^4 + 1) - (z^2 + 1)^2(z^3 - 1) + z^6 + z^2.$$

9.9. Раскрыть скобки в двоичном полиноме $(z+1)^{2^i}$, где i – целое положительное число.

9.10. Найти остаток от деления $z^5 + z^3 + 1$ на $z^2 + z + 1$ над полем $GF(2)$.

9.11. Пусть $g(z)$ – двоичный полином степени r . Доказать, что двоичный блочный код длины n , использующий его в качестве порождающего полинома (т.е. имеющий кодовые полиномы вида $u(z) = b(z)g(z)$, где $b(z)$ – произвольные двоичные полиномы степени не выше $n - r - 1$) является линейным.

9.12. Двоичный полином $g_1(z) = z^4 + z + 1$ – примитивен. Какова наибольшая длина CRC кода на основе этого полинома, обнаруживающего вплоть до трех ошибок? Является ли полином наблюдения $y(z) = z^8 + z^5 + z + 1$ ошибочным или нет? Что можно сказать о

полиноме $y(z) = z^8 + z^4 + z + 1$?

9.13. Порождающими полиномами сверточного кода являются $g_1(z) = 1$, $g_2(z) = 1 + z$. Какова скорость и длина кодового ограничения порождаемого ими кода? Изобразить схему кодера, нарисовать решетчатую диаграмму и определить свободное расстояние кода.

9.14. Порождающими полиномами сверточного кода являются $g_1(z) = 1 + z + z^2$, $g_2(z) = 1 + z + z^2$, $g_3(z) = 1 + z^2$. Какова скорость и длина кодового ограничения порождаемого ими кода? Изобразить схему кодера, нарисовать решетчатую диаграмму и определить свободное расстояние кода.

9.15. Одним из двух порождающих полиномов сверточного кода является $g_1(z) = 1 + z + z^2$. Какой из полиномов $g_2(z) = 1$, $g_2(z) = 1 + z$ или $g_2(z) = 1 + z^2$ лучше в качестве второго порождающего полинома для получения наибольшего выигрыша от кодирования? Чему равен максимальный выигрыш от кодирования? Какого типа выводы следуют из этой задачи при сравнении систематического и несистематического кода?

9.16. Порождающими полиномами сверточного кода являются $g_1(z) = 1 + z + z^2$ и $g_2(z) = 1 + z^3$. Закодировать битовый поток 110110110110110... и объяснить, почему этот код принадлежит классу так называемых *катастрофических* кодов, которые не рекомендуются к практическому использованию?

9.17. Декодировать максимально возможное число бит при условии приема наблюдения 100101100011000 и порождающих полиномов сверточного кода $g_1(z) = 1$, $g_2(z) = 1 + z$ и $g_3(z) = 1 + z$.

9.18. Декодировать наблюдение 111111100000001111, если известно, что в качестве порождающих полиномов сверточного кода используются полиномы задачи 9.14, а после кодирования четырех бит данных кодер принудительно устанавливается в нулевое состояние хвостовыми битами. Насколько далеко расположено декодированное слово от наблюдения? Если результат декодирования правилен, сколько ошибок исправил декодер?

9.19. Исправьте процедуру декодирования в примере 9.3.4, сохраняя все пути, входящие в узел с одинаковыми метриками, как выжившие. Продолжите процедуру после седьмого шага до первого шага, когда может быть осуществлена выдача бит данных. На каком шаге это произойдет? Сколько бит данных выдается?

9.20. Битовый поток кодируется двоичным сверточным кодом с порождающими полиномами $g_1(z) = 1$, $g_2(z) = 1 + z$. Двоичные символы передаются с помощью бинарной ФМ ($0 \rightarrow +1$, $1 \rightarrow -1$). Канал связи является гауссовским и выходные отсчеты наблюдения заданы вектором $\mathbf{y} = (-0.5, -0.5, -3, -4, -6, 2, -4, 5, 3, -2)$. Декодировать данное наблюдение с использованием жесткого и мягкого (основанного на вычислении корреляций наблюдения с путями по решетчатой диаграмме) алгоритма Витерби. Предположим, что два последних кодовых символа отвечают хвостовым битам, устанавливающим кодер в нуль. Объяснить различие (если оно есть) результатов этих двух процедур. Какая из них заслуживает большего доверия?

9.21. Желательно построить турбо-код, начинающийся сверточным кодом, задаваемым порождающими полиномами $g_1(z) = 1 + z^3 + z^4$ и $g_2(z) = 1 + z + z^2 + z^4$. Изобразить составные части кодера турбо-кода.

10. Некоторые достижения в развитии широкополосных систем.

10.1. Многопользовательский прием и подавление помех множественного доступа.

В параграфе 4.1 были рассмотрены два варианта принятия решения о данных в K -пользовательской CDMA системе. Первый из них реализует оптимальную (максимально правдоподобную) процедуру, осуществляемую т.н. многопользовательским приемником, тогда как второй подразумевает использование однопользовательской или обычной процедуры. Стандартный приемник воспринимает помеху множественного доступа (MAI), как дополнительный случайный шум, полностью игнорируя детерминистскую природу сигнатур и их корреляционные свойства. Напротив, многопользовательский алгоритм использует априорные сведения о кодах сигнатур или, по крайней мере, их ансамблевые корреляционные свойства. В рамках данного параграфа кратко обсудим идеи, лежащие в основе многопользовательского приема, начиная с простейшего случая синхронной CDMA.

10.1.1. Оптимальное (МП) правило многопользовательского приема для синхронного варианта CDMA.

Для того чтобы освободить обсуждение от второстепенных деталей рассмотрим простейшую, однако, и достаточно общую модель K -пользовательской CDMA системы с прямым расширением спектра, включающую реальные сигнатуры и передачу данных с помощью бинарной ФМ. Модель охватывает, среди других, любую систему с бинарными сигнатурами и бинарной модуляцией данных. Как и в параграфе 7.2, в рамках данного пункта рассмотрим полностью синхронизированный случай, когда как чипы, так и границы символов данных (бит) всех пользователей жестко связаны во времени. Последнее, наряду с предположением о независимости последовательных бит данных любого пользователя, позволяет ограничить интервал наблюдения длительностью единственного бита $T = T_b$. Тогда групповой сигнал K пользователей представим в виде

$$s(t; \mathbf{b}) = \sum_{k=1}^K A_k b_k s_k(t), \quad (10.1)$$

где аналогично (4.1) $A_k > 0$ – вещественная амплитуда k -го пользовательского сигнала, $\mathbf{b} = (b_1, b_2, \dots, b_K)$ – вектор бит данных K пользователей (битовый образец), $s_k(t)$ – k -я пользовательская сигнатура.

Как указывалось в параграфе 4.1, глобально оптимальная (МП) процедура диктует искать оценку $\hat{\mathbf{b}} = (\hat{b}_1, \hat{b}_2, \dots, \hat{b}_K)$ образца бит данных K пользователей \mathbf{b} , как такое значение \mathbf{b} , которое минимизирует евклидово расстояние (или его квадрат $d^2(\mathbf{s}, \mathbf{y})$) между наблюдением $y(t)$ и групповым сигналом (10.1). Вычисление $d^2(\mathbf{s}, \mathbf{y})$ аналогично (4.3) приводит к следующему результату

$$d^2(\mathbf{s}, \mathbf{y}) = \int_0^T [y(t) - s(t; \mathbf{b})]^2 dt = \|\mathbf{y}\|^2 - 2 \sum_{k=1}^K A_k b_k z_k + \sum_{k=1}^K \sum_{l=1}^K A_k A_l b_k b_l \rho_{kl}, \quad (10.2)$$

где

$$z_k = \int_0^T y(t) s_k(t) dt \quad (10.3)$$

– как обычно, корреляция наблюдения $y(t)$ с k -й сигнатурой, ρ_{kl} – коэффициент корреляции k -й и l -й сигнатур, а наличие амплитуд A_k позволяет использовать соответствующую нормировку сигнатур:

$$E_k = \|\mathbf{s}_k\|^2 = \int_0^T s_k(t) dt = 1, \quad k = 1, 2, \dots, K.$$

Введем еще две метрики: $\mathbf{G} = \text{diag}(A_1, A_2, \dots, A_K)$ – диагональную $K \times K$ матрицу амплитуд пользователей и $\mathbf{C} = [\rho_{kl}]$, $k, l = 1, 2, \dots, K$ – корреляционную матрицу сигнатур. Введя обозначение $\mathbf{z} = (z_1, z_2, \dots, z_K)$ для вектора корреляций (10.3), выражение для квадрата расстояния (10.2) становится

$$d^2(\mathbf{s}, \mathbf{y}) = \|\mathbf{y}\|^2 - 2\mathbf{bGz}^T + \mathbf{bGCGb}^T, \quad (10.4)$$

где надстрочный символ T символизирует транспонирование вектора–матрицы. Первый член в правой части (10.4) фиксирован для текущего наблюдения $y(t)$ и, следовательно, МП оценка $\hat{\mathbf{b}}$ может быть найдена, как значение \mathbf{b} , максимизирующее разность двух оставшихся членов:

$$2\hat{\mathbf{b}}\mathbf{Gz}^T - \hat{\mathbf{b}}\mathbf{GCG}\hat{\mathbf{b}}^T = \max_{\mathbf{b}} (2\mathbf{bGz}^T - \mathbf{bGCGb}^T). \quad (10.5)$$

Вектор данных \mathbf{b} жестко подчинен ограничению алфавита бинарной ФМ $b_i = \pm 1$, и не существует более эффективной в вычислительном плане процедуры, чем простой перебор всех 2^K возможных битовых образцов и сравнение результатов их подстановки в правую часть (10.5). Следовательно, МП многопользовательское обнаружение, выполняемое в соответствии с (10.5), характеризуется экспоненциальной сложностью от числа пользователей (см. числовой пример в параграфе 4.1). С другой стороны, мотивация к рассмотрению многопользовательского обнаружения может оказаться даже сильнее, когда число пользователей столь значительно, что обычный обнаружитель оказывается неэффективным вследствие высокого уровня помех МАИ. Это объясняет причины интереса, проявляемого к квазиоптимальным многопользовательским алгоритмам, некоторые из которых рассматриваются в оставшейся части параграфа.

10.1.2. Алгоритм декорреляции.

Начнем с обычного (т.е. основанного на вычислении корреляции) приемника данных пользователя номер один. Согласно (10.1) наблюдение представимо как

$$y(t) = s(t; \mathbf{b}) + n(t) = \sum_{k=1}^K A_k b_k s_k(t) + n(t). \quad (10.6)$$

Последнее соотношение после подстановки в (10.3) при $k = 1$ дает

$$z_1 = A_1 b_1 + \sum_{k=2}^K A_k b_k \rho_{k1} + n_1, \quad (10.7)$$

где $n_1 = \int_0^T n(t) s_1(t) dt$ – шумовой отсчет на выходе первого коррелятора. Второй член соотношения (10.7) представляет собой помеху МАИ, и проблема заключается в том, возможно ли ее подавление до нуля с помощью некоторого линейного преобразования входного наблюдения. Какой бы ни была эта линейная операция, в итоге должны иметь свободную от МАИ величину ζ_1 вместо z_1 , т.е. скаляр, обеспечивающий решение о текущем бите первого пользователя как

$$\hat{b} = \text{sign}(\zeta_1). \quad (10.8)$$

Любая линейная операция, преобразующая $y(t)$ в скаляр, может быть описана как корреляция вида

$$\zeta_1 = \int_0^T y(t)u(t)dt, \quad (10.9)$$

отличающаяся от (10.3) только опорным сигналом $u(t)$. Следовательно, подавление помех МАИ будет осуществляться путем отказа от согласованной опоры $s_1(t)$ в пользу рассогласованной $u(t)$, т.е. ценой потерь в отношении сигнал-шум относительно теплового шума. Ранее уже прибегали к подобному методу, обращаясь к фильтрам нулевого подавления, устраняющим боковые лепестки автокорреляции (см. параграф 6.12). Используя (10.6) в (10.9), заменим (10.7) на следующее выражение

$$\zeta_1 = A_1 b_1 \rho_{1u} + \sum_{k=2}^K A_k b_k \rho_{ku} + n'_1, \quad (10.10)$$

где ρ_{ku} – коэффициент корреляции k -й сигнатуры с опорой $u(t)$ с нормировкой $u(t)$ точно такой же, как и для сигнатур, $n'_1 = \int_0^T n(t)u(t)dt$ – шумовой вклад в ζ_1 .

Представим теперь сигнатуры и опорный сигнал $u(t)$ в виде (2.50), типичном для DS CDMA с вещественно значимыми сигнатурами:

$$s_k(t) = \sum_{i=0}^{N-1} a_{k,i} s_0(t-i\Delta), \quad u(t) = \sum_{i=0}^{N-1} u_i s_0(t-i\Delta), \quad (10.11)$$

где $u_i, i=0,1,\dots,N-1$ – вещественная кодовая последовательность опорного колебания $u(t)$. Используя векторное обозначение кодовых последовательностей $\mathbf{a}_k = (a_{k,0}, a_{k,1}, \dots, a_{k,N-1})$, $\mathbf{u} = (u_0, u_1, \dots, u_{N-1})$ (см. параграф 7.2) и полагая без потери общности энергию чипа $E_0=1$, приходим к соотношениям $\rho_{kl} = (\mathbf{a}_k, \mathbf{a}_l) = \mathbf{a}_k \mathbf{a}_l^T$, $\rho_{ku} = (\mathbf{a}_k, \mathbf{u}) = \mathbf{a}_k \mathbf{u}^T$. Для устранения слагаемого помех МАИ в (10.10) независимо от амплитуд и бит мешающих пользователей необходимо выполнить $K-1$ условий: $\rho_{ku} = 0, k > 2$. Другими словами, опорный код \mathbf{u} должен являться решением множества линейных уравнений $\mathbf{a}_k \mathbf{u}^T = \mathbf{u} \mathbf{a}_k^T = 0, k = 2, 3, \dots, K$. Требование сохранения ненулевым полезного эффекта ρ_{1u} означает, что \mathbf{u} есть соответствующим образом масштабированное решение \mathbf{v} уравнения

$$\mathbf{v} \mathbf{A} = \mathbf{e}_1, \quad (10.12)$$

где столбцами $K \times K$ сигнатурной матрицы \mathbf{A} служат кодовые вектора сигнатур: $\mathbf{A} = (\mathbf{a}_1^T, \mathbf{a}_2^T, \dots, \mathbf{a}_K^T)$, а \mathbf{e}_1 – K -мерный вектор вида $\mathbf{e}_1 = (1, 0, 0, \dots, 0)$. При линейной независимости всех векторов сигнатур система (10.12) может иметь множество решений, однако из всех векторов \mathbf{v} , удовлетворяющих (10.12), выберем такой, который представляет собой линейную комбинацию сигнатур, т.е. строк $\mathbf{A}^T : \mathbf{v} = \mathbf{x} \mathbf{A}^T$, где \mathbf{x} – неизвестный K -мерный вектор-строка. Основанием для такого выбора служит то, что включение в \mathbf{v} любого компонента, ортогонального к пространству векторов сигнатур, только увеличит норму \mathbf{v} , т.е. шумовой компонент на выходе коррелятора, без увеличения полезного первого слагаемого в (10.10). С учетом такой подстановки (10.12) становится

$$\mathbf{x} \mathbf{A}^T \mathbf{A} = \mathbf{x} \mathbf{C} = \mathbf{e}_1.$$

Линейная независимость сигнатур (столбцов матрицы \mathbf{A}) определяет ранг K корреляци-

онной $K \times K$ матрицы $\mathbf{C} = \mathbf{A}^T \mathbf{A}$, т.е. ее обратимость и единственность решения вышеприведенного уравнения: $\mathbf{x} = \mathbf{e}_1 \mathbf{C}^{-1}$. Тогда

$$\mathbf{v} = \mathbf{x} \mathbf{A}^T = \mathbf{e}_1 \mathbf{C}^{-1} \mathbf{A}^T = \mathbf{e}_1 (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \quad (10.13)$$

представляет собой искомое решение (10.12), масштабирование которого $\mathbf{u} = \mathbf{v} / \|\mathbf{v}\|$ приводит к нормированной декорреляционной опоре \mathbf{u} , так что $\mathbf{u} \mathbf{a}_1^T = \rho_{1u}$. Практически данная нормировка не является необходимой, поскольку не оказывает никакого влияния на знак ς_1 в решающем правиле (10.8).

Фактически вектор опоры (10.13) представляет собой вектор, ортогональный ко всем сигнатурам, кроме первой, который полностью устраняет помехи МАИ на выходе коррелятора, настроенного на первый пользовательский сигнал. Аналогичным путем можно определить опорный сигнал для k -го пользовательского приемника, заменяя \mathbf{e}_1 в (10.13) на вектор \mathbf{e}_k с единственным k -м ненулевым компонентом.

Основным недостатком приемника с декорреляцией является его работоспособность только с линейно независимыми сигнатурами. Если это условие не выполняется, то любая попытка подавить до нуля помехи МАИ неизбежно приведет также к исчезновению полезного эффекта (первого слагаемого) в (10.10). В то же время, линейная независимость означает, что при $K \leq N$ наиболее адекватным выбором сигнатур служит ортогональное множество (см. параграф 7.2), который влечет за собой оптимальность однопользовательского приемника и автоматическое подавление помех МАИ без потерь в отношении сигнал-шум и, значит, отсутствие необходимости в специальной декорреляционной обработке. Если же рассматривается случай перенасыщения ($K > N$), то линейная независимость сигнатур невозможна и декорреляционный алгоритм не может использоваться.

10.1.3. Обнаружение по минимуму среднеквадратической ошибки.

Снова используем идею рассогласованной обработки в корреляторе, настроенном на сигнал первого пользователя, но на этот раз вместо подавления помех МАИ до нуля постараемся минимизировать общий вредный эффект помех МАИ и шума. Обратившись к (10.7), можно заметить, что единственным полезным компонентом в нем является $A_1 b_1$, тогда как остальные два представляют полную помеху (МАИ плюс шум). В свете этого естественно отыскать линейную операцию (10.9), воспроизводящую полезный вклад с *минимумом среднеквадратической ошибки* (МСКО) (*minimum mean-square error* (ММСЕ)). Для формализации задачи перепишем вначале (10.9) в векторной форме, подставив в него выражение для $u(t)$ из (10.11):

$$\varsigma_1 = \sum_{i=0}^{N-1} u_i y_i = \mathbf{u} \tilde{\mathbf{y}}^T, \quad (10.14)$$

где $\tilde{\mathbf{y}} = (y_0, y_1, \dots, y_{N-1})$ и $y_i = \int_0^T y(t) s_0(t - i\Delta) dt$, $i = 0, 1, \dots, N-1$. На практике y_i могут быть получены в виде отсчетов в соответствующие моменты на выходе фильтра, согласованного с одиночным чипом (см. (2.68)), которые позволяют представить $\tilde{\mathbf{y}}$ в качестве вектора наблюдений после согласованной с чипом фильтрации. Наша задача теперь состоит в соответствующем выборе опорного кодового вектора \mathbf{u} , минимизирующего средний квадрат отклонения $\overline{\varepsilon^2}$ величины ς_1 от $A_1 b_1$:

$$\overline{\varepsilon^2} = \overline{|A_1 b_1 - \varsigma_1|^2} = \overline{|A_1 b_1 - \mathbf{u} \tilde{\mathbf{y}}^T|^2} = \min_{\mathbf{u}}.$$

Отметим, что отсутствуют априорные сведения о подходящей нормировке опоры \mathbf{u} . После возведения в квадрат и почленного усреднения выражение для среднеквадратической ошибки принимает вид

$$\overline{\varepsilon^2} = \overline{(A_1 b_1)^2} - 2A_1 \overline{\mathbf{u} \tilde{\mathbf{y}}^T b_1} + \overline{(\mathbf{u} \tilde{\mathbf{y}}^T)^2} = \overline{(A_1 b_1)^2} - 2A_1 \overline{\mathbf{u} \tilde{\mathbf{y}}^T b_1} + \overline{\mathbf{u} \tilde{\mathbf{y}}^T \tilde{\mathbf{y}} \mathbf{u}^T}, \quad (10.15)$$

где использованы правила элементарной матричной алгебры (коммутативность умножения на скаляр и ассоциативность вектор–матричного умножения, коммутативность скалярного произведения $\mathbf{u} \tilde{\mathbf{y}}^T = \tilde{\mathbf{y}} \mathbf{u}^T$), а также неслучайность \mathbf{u} . Значение i -го компонента $\tilde{\mathbf{y}}$ после подстановки (10.6), а затем (10.11) принимает вид

$$\begin{aligned} y_i &= \int_0^T y(t) s_0(t - i\Delta) dt = \int_0^T \left[\sum_{k=1}^K A_k b_k s_k(t) + n(t) \right] s_0(t - i\Delta) dt = \\ &= \sum_{k=1}^K A_k b_k \sum_{j=0}^{N-1} a_{k,j} \int_0^T s_0(t - j\Delta) s_0(t - i\Delta) dt + v_i = \sum_{k=1}^K A_k b_k a_{k,i} + v_i, \end{aligned} \quad (10.16)$$

где $v_i = \int_0^T n(t) s_0(t - i\Delta) dt$ – i -й отсчет шума на выходе фильтра согласованного с чипом и использовано естественное предположение, что чипы, сдвинутые во времени на ненулевое целое число периодов повторения Δ ортогональны (например, если длительность чипа не больше, чем Δ , то подобные чипы не перекрываются). Теперь видно, что $\overline{y_i b_1} = A_1 a_{1,i}$, а поскольку биты различных пользователей независимы друг от друга ($\overline{b_k b_l} = \delta_{kl}$) и от шума ($\overline{v_i b_k} = \overline{v_i} \cdot \overline{b_k} = 0$), и, следовательно, имеем

$$A_1 \overline{\tilde{\mathbf{y}}^T b_1} = A_1^2 \mathbf{a}_1^T. \quad (10.17)$$

Аналогичным образом вычислим матричное выражение $\overline{\tilde{\mathbf{y}}^T \tilde{\mathbf{y}}}$, элементы которого представляют собой просто корреляционные моменты $\overline{y_i y_j}$ отсчетов y_i . Тогда, согласно (10.16) и, учитывая некоррелированность отсчетов шума на выходе фильтра согласованного с чипом, имеем

$$\overline{y_i y_j} = \sum_{k=1}^K \sum_{l=1}^K A_k A_l \overline{b_k b_l} a_{k,i} a_{l,j} + \overline{v_i v_j} = \sum_{k=1}^K A_k^2 a_{k,i} a_{k,j} + \sigma^2 \delta_{ij},$$

где σ^2 – дисперсия шумовой компоненты y_i . Таким образом, $N \times N$ корреляционная матрица \mathbf{R} вектора наблюдения $\tilde{\mathbf{y}}$ имеет вид

$$\mathbf{R} = \overline{\tilde{\mathbf{y}}^T \tilde{\mathbf{y}}} = \overline{[y_i y_j]} = \mathbf{A} \mathbf{G}^2 \mathbf{A}^T + \sigma^2 \mathbf{I}_N, \quad (10.18)$$

где \mathbf{I}_N – единичная матрица N -го порядка. Подстановка (10.17) и (10.18) в (10.15) после отбрасывания первого слагаемого, не зависящего от \mathbf{u} , дает следующую скалярную функцию для минимизации по \mathbf{u}

$$f(\mathbf{u}) = \mathbf{u} \mathbf{R} \mathbf{u}^T - 2A_1^2 \mathbf{u} \mathbf{a}_1^T. \quad (10.19)$$

В точке \mathbf{u} , доставляющей экстремум функции $f(\mathbf{u})$, градиент $f(\mathbf{u})$, т.е. вектор, компоненты которого являются производными $f(\mathbf{u})$ по каждому компоненту вектора \mathbf{u} , должен быть равен нулю. Градиент $f(\mathbf{u})$ легко находится (см. задачу 10.3), как $2(\mathbf{R} \mathbf{u}^T - A_1^2 \mathbf{a}_1^T)$. Таким образом, учитывая обратимость матрицы \mathbf{R} , вектор \mathbf{u} , доставляющий экстремум

функции $f(\mathbf{u})$, определяется уравнением $\mathbf{u}^T = A_1^2 \mathbf{R}^{-1} \mathbf{a}_1^T$ или после транспонирования

$$\mathbf{u} = A_1^2 \mathbf{a}_1 \mathbf{R}^{-1}, \quad (10.20)$$

где \mathbf{R} определяется соотношением (10.18). Читателя, желающего проверить, что только что найденный экстремум является минимумом (10.19), отсылаем к задаче 10.3.

Ясно, что данный алгоритм не базируется на обратимости корреляционной матрицы сигнатур $\mathbf{C} = \mathbf{A}^T \mathbf{A}$; обратимой должна быть только корреляционная матрица (10.18) наблюдений, что практически всегда выполняется. Следовательно, в отличие от (10.13) решение (10.20) является универсальным, не зависящим от соотношения между K и N . В тоже время, по крайней мере в одном важном частном случае решение (10.20) вырождается в однопользовательский алгоритм. Пусть множество сигнатур удовлетворяет границе Велча, означающее, что строки матрицы сигнатур \mathbf{A} ортогональны (см. пункт 7.2.2), т.е. $\mathbf{A} \mathbf{A}^T = \mathbf{I}_N$. Если все сигналы имеют одинаковую интенсивность A , то $\mathbf{G}^2 = A^2 \mathbf{I}_K$, и корреляционная матрица наблюдений (10.18) становится простейшей: $\mathbf{R} = (A^2 + \sigma^2) \mathbf{I}_N$, приводя к результату $\mathbf{u} = [A^2 / (A^2 + \sigma^2)] \mathbf{a}_1$, который представляет собой взвешенную первую сигнатуру, т.е. опорный сигнал в обычном приемнике. Таким образом, для сигнатур равной мощности, удовлетворяющих границе Велча, не существует особенного алгоритма обработки по критерию МСКО. Этот факт является достаточно тривиальным при $K \leq N$, поскольку в этом случае подобные сигнатуры ортогональны и стандартный приемник полностью устраняет помехи МАИ при наилучшей фильтрации шума, однако, для варианта с перенасыщением ($K > N$) данное утверждение не предсказуемо заранее.

В литературе результат (10.20), как правило, приводится в несколько иной форме, в явном виде включающем корреляционную матрицу сигнатур $\mathbf{C} = \mathbf{A}^T \mathbf{A}$ [99–101]. Получение ее возможно, например, с помощью обратной матричной леммы, приведенной здесь в форме, согласованной с контекстом:

$$\mathbf{R}^{-1} = (\mathbf{A} \mathbf{G}^2 \mathbf{A}^T + \sigma^2 \mathbf{I}_N)^{-1} = \frac{1}{\sigma^2} \mathbf{I}_N - \frac{1}{\sigma^2} \mathbf{A} (\mathbf{A}^T \mathbf{A} + \sigma^2 \mathbf{G}^{-2})^{-1} \mathbf{A}^T. \quad (10.21)$$

Доказательство этого результата состоит в прямой проверке (Задача 10.4). Отметим, что оно справедливо всегда, когда матрица \mathbf{G} обратима, что автоматически выполняется, если все амплитуды пользователей ненулевые. Используя (10.21) и соотношение $A_1^2 \mathbf{a}_1 = \mathbf{e}_1 \mathbf{G}^2 \mathbf{A}^T$ в (10.20), получаем

$$\begin{aligned} \mathbf{u} &= \mathbf{e}_1 \mathbf{G}^2 \mathbf{A}^T \mathbf{R}^{-1} = \frac{1}{\sigma^2} \mathbf{e}_1 \mathbf{G}^2 \left[\mathbf{A}^T - \mathbf{A}^T \mathbf{A} (\mathbf{A}^T \mathbf{A} + \sigma^2 \mathbf{G}^{-2})^{-1} \mathbf{A}^T \right] = \\ &= \frac{1}{\sigma^2} \mathbf{e}_1 \mathbf{G}^2 \left[\mathbf{A}^T - (\mathbf{A}^T \mathbf{A} + \sigma^2 \mathbf{G}^{-2}) (\mathbf{A}^T \mathbf{A} + \sigma^2 \mathbf{G}^{-2})^{-1} \mathbf{A}^T \right] + \mathbf{e}_1 (\mathbf{A}^T \mathbf{A} + \sigma^2 \mathbf{G}^{-2})^{-1} \mathbf{A}^T, \end{aligned}$$

и окончательно

$$\mathbf{u} = \mathbf{e}_1 (\mathbf{C} + \sigma^2 \mathbf{G}^{-2})^{-1} \mathbf{A}^T. \quad (10.22)$$

Возвращаясь назад к (10.14), запишем окончательно правило решения о бите первого пользователя в виде

$$b_1 = \text{sign}(\zeta_1) = \text{sign}[\mathbf{e}_1 (\mathbf{C} + \sigma^2 \mathbf{G}^{-2})^{-1} \mathbf{A}^T \tilde{\mathbf{y}}^T]. \quad (10.23)$$

Распространение этого правила на приемник данных k -го пользователя следует очевидным образом: необходимо заменить \mathbf{e}_1 на \mathbf{e}_k .

Подчеркивая еще раз, что рассматриваемое правило является универсальным неза-

висимо от обратимости корреляционной матрицы сигнатур, тем не менее, следует учитывать, что если \mathbf{C} является не сингулярной (необходимым условием этого является $K \leq N$), а тепловой шум незначителен, то МСКО обнаружитель асимптотически сходится к декорреляционному:

$$\mathbf{u} = \mathbf{e}_1(\mathbf{C} + \sigma^2 \mathbf{G}^{-2}) \mathbf{A}^T \xrightarrow{\sigma^2 \rightarrow 0} \mathbf{e}_1 \mathbf{C}^{-1} \mathbf{A}^T.$$

Для демонстрации эффективности МСКО алгоритма наиболее естественным является сравнение отношения сигнал-(помеха плюс шум) SINR для результата приема ζ_1 в случае опорного сигнала (10.22) и при стандартном приеме $\mathbf{u} = \mathbf{a}_1$. Как следует из (10.14), вклад k -го сигнала в ζ_1 составляет $A_k \mathbf{u} \mathbf{a}_k^T$, так что полезная мощность (создаваемая сигналом первого пользователя) будет $A_1^2 (\mathbf{u} \mathbf{a}_1^T)^2$, а мощность помех MAI – $\sum_{k=2}^K A_k^2 (\mathbf{u} \mathbf{a}_k^T)^2$. В свою очередь, некоррелированность отсчетов шума v_i на выходе фильтра, согласованного с чипом, означает, что их мощности суммируются после взвешивания u_i в (10.14), так что полная мощность шумовой компоненты ζ_1 составляет величину $\sigma^2 \sum_{i=0}^{N-1} u_i^2 = \sigma^2 \|\mathbf{u}\|^2$. Тогда с учетом полученных результатов искомое SINR будет

$$q_I^2 = \frac{A_1^2 (\mathbf{u} \mathbf{a}_1^T)^2}{\sum_{k=2}^K A_k^2 (\mathbf{u} \mathbf{a}_k^T)^2 + \sigma^2 \|\mathbf{u}\|^2} = \frac{q_{1b}^2 (\mathbf{u} \mathbf{a}_1^T)^2}{\sum_{k=2}^K q_{kb}^2 (\mathbf{u} \mathbf{a}_k^T)^2 + \|\mathbf{u}\|^2}, \quad (10.24)$$

где $q_{kb}^2 = A_k^2 / \sigma^2$ – отношение сигнал-шум по мощности на бит данных k -го пользователя.

Пример 10.1.1. Рассмотрим перенасыщенную синхронную CDMA систему с сигнатурами, удовлетворяющими границе Велча. Двоичные ансамбли, удовлетворяющие границе Велча, существуют для любого $K > N$, допускающего существование $K \times K$ матрицы Адамара. Тогда сигнатурами являются K столбцов этой матрицы после отбрасывания из нее $K - N$ любых строк. С учетом упоминавшегося ранее случай равных мощностей не демонстрирует никакого преимущества МСКО алгоритма относительно стандартного обнаружения. Для значений $K = 64$, $N = 48$ был выбран случайный образец амплитуд пользователей, полученный путем взятия 64 отсчетов релейской случайной величины для имитации релейского канала. На рис. 10.1 представлены зависимости отношения SINR, рассчитываемые согласно (10.24), для МСКО и стандартного обнаружителей от отношения SNR на бит при некотором «щадящем» образце амплитуд. Приведенные кривые свидетельствуют, что выигрыш МСКО алгоритма иногда достигает значительных величин (на рис. 10.1 вплоть до 10 дБ). Стоит еще напомнить, что такой выигрыш является делом случая: для некоторых образцов амплитуд он может оказаться даже большим, однако чем более равномерно распределены амплитуды в образце, тем меньше будет разность в отношениях SINR между МСКО и стандартным приемником. Следует упомянуть еще об одном достоинстве, заключающемся в том, что МСКО детектор в большей степени приспособлен к разбросу в интенсивностях пользовательских сигналов (конечно, если сигнатуры не являются ортогональными) по сравнению со стандартным приемником, что делает его особенно привлекательным при отсутствии жестких требований к управлению мощностью. \square

10.1.4. Слепой МСКО обнаружитель.

Несмотря на то, что вычислительная сложность МСКО алгоритма (точно также как и декорреляционного) не является совсем затруднительной для практической реализации,

существует одна реализационная проблема, мотивирующая дальнейшие исследования. Как следует из (10.20), ключевой операцией МСКО алгоритма является процедура обращения корреляционной матрицы наблюдений \mathbf{R} , определяемой соотношением (10.18). Для ее построения приемник k -го пользователя должен знать наряду со своей собственной сигнатурой также и сигнатуры всех остальных пользователей. В некоторых ситуациях подобное требование выглядит чрезмерным. Одним из наиболее типичных применений синхронного варианта CDMA является прямой канал связи мобильного радио, и необходимость знания каждым мобильным абонентом всех сигнатур, используемых в текущий момент остальными пользователями, значительно усложнит работу системы. К счастью, корреляционная матрица наблюдаемого процесса может быть оценена экспериментально из самого наблюдения в случае, если период наблюдения достаточно большой. Данное положение является основной идеей *слепых (blind)* многопользовательских алгоритмов.

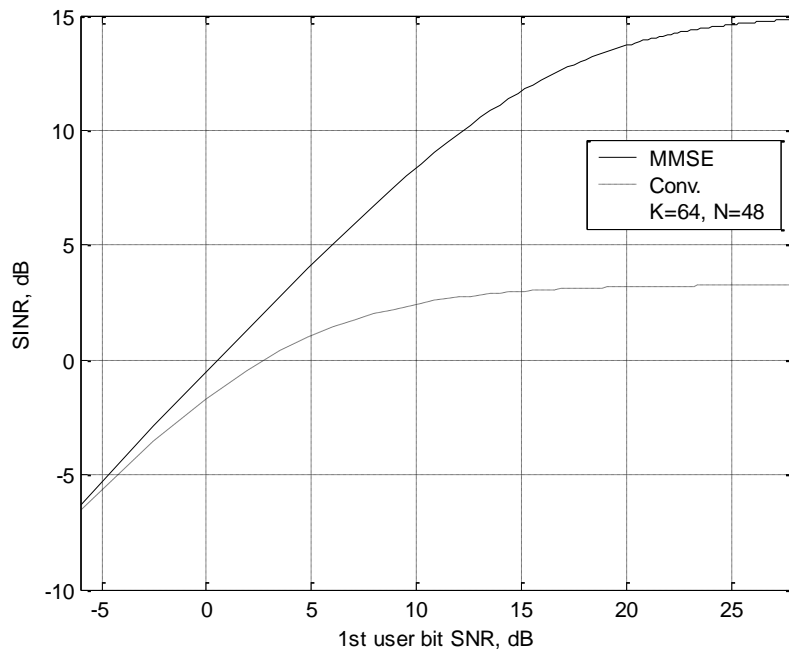


Рис.10.1. Примерные кривые зависимости отношения SINR для МСКО и однопользовательского приемников.

Предположим, что $\tilde{\mathbf{y}}_i$ представляет собой N -мерный вектор–строку отсчетов на выходе фильтра согласованного с чипом, соответствующий битам данных с номерами $i = 0, 1, \dots$. Тогда оценка $\hat{\mathbf{R}}_i$ корреляционной матрицы наблюдений может быть найдена как

$$\hat{\mathbf{R}}_i = \frac{1}{i+1} \sum_{t=0}^i \tilde{\mathbf{y}}_t^T \tilde{\mathbf{y}}_t = \frac{i}{i+1} \cdot \frac{1}{i} \sum_{t=0}^{i-1} \tilde{\mathbf{y}}_t^T \tilde{\mathbf{y}}_t + \frac{1}{i+1} \tilde{\mathbf{y}}_i^T \tilde{\mathbf{y}}_i = \frac{i}{i+1} \hat{\mathbf{R}}_{i-1} + \frac{1}{i+1} \tilde{\mathbf{y}}_i^T \tilde{\mathbf{y}}_i.$$

Следовательно, вычисление $\hat{\mathbf{R}}_i$ возможно осуществить рекуррентным образом: как только поступает часть отсчетов наблюдения, охватывающих новый бит, она используется для обновления оценки $\hat{\mathbf{R}}_i$. После этого полученный результат подставляется в (10.20) для нахождения текущей опоры в виде $\mathbf{u}_i = A_1^2 \mathbf{a}_1 \hat{\mathbf{R}}_i$, вычисления с помощью (10.14) $\zeta_1 = \mathbf{u}_i \tilde{\mathbf{y}}_i^T$ и принятия решения об i -м бите согласно (10.23). Вариации и улучшения слепого многопользовательского алгоритма многочисленны и могут быть найдены в литературе (см. [19,100,101] и приведенной в них библиографии).

10.1.5. Устранение помех.

Можно трактовать низкую сложность как декорреляционного, так и МСКО обнаружителей на основании того факта, что они используют однопользовательскую философию, т.е. линейную операцию умножения вектора наблюдений $\tilde{\mathbf{y}}$ на вектор рассогласованной опоры \mathbf{u} . Стратегия устранения помех также базируется на стандартном приемнике, дополненного петлей вычитания слагаемого помех МАИ из выходного эффекта, определяемого (10.7). Предположим, приемник первого пользователя знает сигнатуры и амплитуды всех пользователей и тем или иным путем им получены оценки $\hat{b}_2, \hat{b}_3, \dots, \hat{b}_K$ бит данных сторонних пользователей. Тогда он способен восстановить сигналы всех посторонних пользователей, вычесть их сумму из наблюдения $y(t)$ и использовать результат в качестве входного сигнала (предположительно свободного от помех МАИ) для стандартного приемника. Очевидно, что эффективность подобного обнаружителя в сильной степени зависит от достоверности знаний амплитуд сторонних сигналов и точности оценок бит сторонних пользователей. Среди прочих в литературе широко обсуждается многоэтапная процедура [19,102]. Ее первый этап заключается в последовательном оценивании бит пользователей от более сильных к более слабым сигналам и использовании уже полученных оценок бит $k-1$ пользователей для устранения соответствующих им помех МАИ при оценке бита k -го пользователя. Осуществив оценивание всех K бит согласно этому алгоритму, процедура переходит к следующему этапу, на котором повторяются все те же операции. К этому моменту знания приемника о помехах значительно более полные по сравнению с предшествующим этапом и вычитание воссозданных помех МАИ начинается с самого начала этапа, т.е. с оценивания бита самого сильного пользователя. Этапы, подобные рассмотренному, повторяются столько раз, сколько желательно, каждый из которых начинается с обновления воссоздаваемых помех МАИ и последовательного их уточнения во время оценивания образца бит пользователей. После удовлетворения того или иного критерия завершения, процедура выдает окончательную оценку бита интересующего пользователя.

10.1.6. Асинхронные многопользовательские обнаружители.

Хотя кратко рассмотренные ранее многопользовательские алгоритмы обнаружения ориентированы на синхронный вариант CDMA, они, после соответствующей модификации, применимы также и для асинхронного случая. Обратимся к соотношению (10.6) и преобразуем его к асинхронной ситуации, иллюстрацией которой служит рис. 7.15, полагая согласованными границы чипов всех пользователей и оперируя только с вещественными видеосигналами сигнатур (т.е. игнорируя взаимный фазовый сдвиг полосных сигналов). Предположим также, что все сигнатуры периодичны с периодом, равным длительности бита $T = T_b$, и рассмотрим, как и ранее, приемник первого пользователя ($\tau_1 = 0$) на интервале наблюдения бита данных под нулевым номером. Тогда

$$y(t) = \begin{cases} A_1 b_{1,0} s_1(t) + \sum_{k=2}^K A_k b_{k,-1} s_k(t - \tau_k) + n(t), & 0 \leq t \leq \tau_1, \\ A_1 b_{1,0} s_1(t) + \sum_{k=2}^K A_k b_{k,0} s_k(t - \tau_k) + n(t), & \tau_1 \leq t \leq T. \end{cases} \quad (10.25)$$

Как следует из (10.25), характер помех МАИ, создаваемых k -м пользователем, зависит теперь от изменения или нет значения бита данных k -го пользователя в течение интервала наблюдения, поскольку во втором случае k -я сигнатура меняет свою полярность в момент $t = \tau_k$. Предположим, что приемнику первого пользователя известны временные задержки τ_k всех пользователей. В этом случае число потенциально опасных векторов помех МАИ,

создаваемых $K-1$ сторонними пользователями, будет $2(K-1)$ в отличие от синхронного варианта, при котором аналогичная величина равна только $K-1$. С учетом первой сигнатуры всего имеется $2K-1$ векторов. Если все они являются линейно независимыми, то для полного исключения влияния помех МАИ вне зависимости от значений бит, передаваемых пользователями, может быть использован модифицированный вариант декорреляционного алгоритма (10.13). Ясно, что в пределах размерности (показателя расширения) N максимальное число пользователей K , допускающее подобную обработку, ограничено соотношением $2K-1 \leq N$, т.е. не может быть больше $\lfloor N/2 \rfloor$. Необходимая модификация (10.13) заключается в замене «чистой» $N \times K$ сигнатурной матрицы A на расширенную, размерности $N \times (2K-1)$, столбцами которой являются первая сигнатура, $K-1$ сторонних сигнатур с соответствующими временными сдвигами и их инвертированные по полярности копии, начинающиеся с нулевого номера чипа (см. рис. 7.15).

МСКО обнаружитель воспринимает помехи МАИ статистически, т.е. как случайный процесс, описываемый своей корреляционной матрицей. Очевидно, аналогичная интерпретация применима и к асинхронному случаю, и снова необходима некоторая ревизия, учитывающая изменение корреляционной матрицы наблюдений [19].

Поскольку в асинхронной системе каждый бит стороннего пользователя охватывает два последовательных бита интересующегося пользователя, корреляция помехи МАИ простирается за длительность одного бита. Последнее объясняет тот факт, что, как правило, характеристики асинхронного многопользовательского обнаружителя улучшаются, если интервал наблюдения охватывает несколько бит данных.

Ограниченность предоставленного объема не дает возможности более глубоко обсудить проблему многопользовательского обнаружения. Эта задача в настоящее время является одной из наиболее интересных и быстро развивающихся. Читателю, стремящемуся к более глубокому проникновению в ее настоящее состояние, а также к ознакомлению с перспективами ее применения в 3G стандартах и за этими рамками рекомендуется обратиться к книгам [19,100,102–104] и обширной библиографии, в них представленной.

10.2. Модуляция со многими несущими и OFDM.

В последнее время метод передачи, называемый модуляцией со *многими несущими* (*multicarrier* (МС)) приобрел большую популярность в телекоммуникации. В самом общем смысле МС модуляция означает ничто иное, как привлечение множества параллельных *поднесущих* (*subcarriers*) для передачи символов одного и того же потока данных. Выбор конкретного варианта использования поднесущих может изменяться в зависимости от выдвигаемой цели. В рассматриваемом контексте наиболее интересной задачей является связь между МС модуляцией и концепцией расширения спектра. Как следует из ранее рассмотренного материала, все достоинства, присущие широкополосности, объясняются большой величиной частотно–временного произведения (выигрышем от обработки, коэффициентом расширения) WT . Существуют различные способы расширения спектра сигнала при заданной его длительности. Расширение спектра прямой последовательностью решает данную задачу путем замены обычных «медленных» импульсов символов «быстрыми» сигналами некоторой подходящей формы, т.е. управлением полосы сигнала во временной области. МС модуляция с этой точки зрения может рассматриваться как альтернативный метод расширения спектра путем его прямого формирования в частотной области.

10.2.1. Метод CDMA с прямым расширением спектра и многими несущими.

Простейшим вариантом реализации МС метода в CDMA является обычное частотное мультиплексирование данных одного и того же пользователя. Оно осуществляется путем передачи демупльтиплексированных данных с прямым расширением спектра на M_c поднесущих с не перекрывающимися спектрами сигналов. Рис. 10.2 поясняет этот способ передачи, называемый CDMA методом с прямым расширением спектра и многими несущими – multicarrier DS CDMA (МС-DS-CDMA). Источник «быстрого» потока данных со скоростью $R = 1/T_b$ демупльтиплексирован (разделен на части) в M_c параллельных «медленных» битовых потоках со скоростью R/M_c (или длительностью бита $T'_b = M_c T_b$) каждый. Медленный i -й поток бит передается на некоторой своей поднесущей частоте $f_i, i = 1, 2, \dots, M_c$ с использованием прямого расширения спектра и некоторого вида модуляции данных (бинарной ФМ, квадратурной или др.). При длительности чипа расширяющей последовательности Δ' и длительности символа данных, передаваемого на поднесущей, T'_p полоса и коэффициент расширения спектра сигнала на поднесущей составляют $W' \approx 1/\Delta'$ и $N' = T'_p / \Delta' \approx WT'_p$ соответственно. Разнос F соседних поднесущих частот должен исключать перекрытие спектров, т.е. удовлетворять ограничению $F = W' + F_g$, где F_g – защитный интервал, отводимый для внеполосных боковых лепестков спектра. Таким образом, полная полоса, занимаемая системой, составит

$$W = (M_c - 1)F + W' = M_c W' + (M_c - 1)F_g. \quad (10.26)$$

Приемник МС-DS-CDMA системы состоит из M_c параллельных идентичных приемных каналов, каждый из которых настроен на свою собственную поднесущую и осуществляет обработку сигнала независимо от других. Затем демодулированные медленные битовые потоки с их выходов мультиплексируются для восстановления переданного источником быстрого битового потока.

Сравнение данной системы с обычной DS CDMA, занимающей такую же общую полосу W и имеющую такую же скорость передачи данных R , показывает, что вследствие необходимости введения защитных интервалов МС-DS-CDMA система уступает DS CDMA системе в коэффициенте расширения. Если обе системы используют один и тот же

вид модуляции, то длительность символа T_p в DS CDMA системе должна быть в M_c раз короче, чем аналогичная характеристика T'_p в MC-DS-CDMA. Однако, длительность чипа в DS CDMA может быть взята равной $\Delta \approx 1/W$, что обеспечивает коэффициент расширения $N = T_p / \Delta \approx WT'_p / M_c$. Тогда, согласно (10.26), отношение данной характеристики к

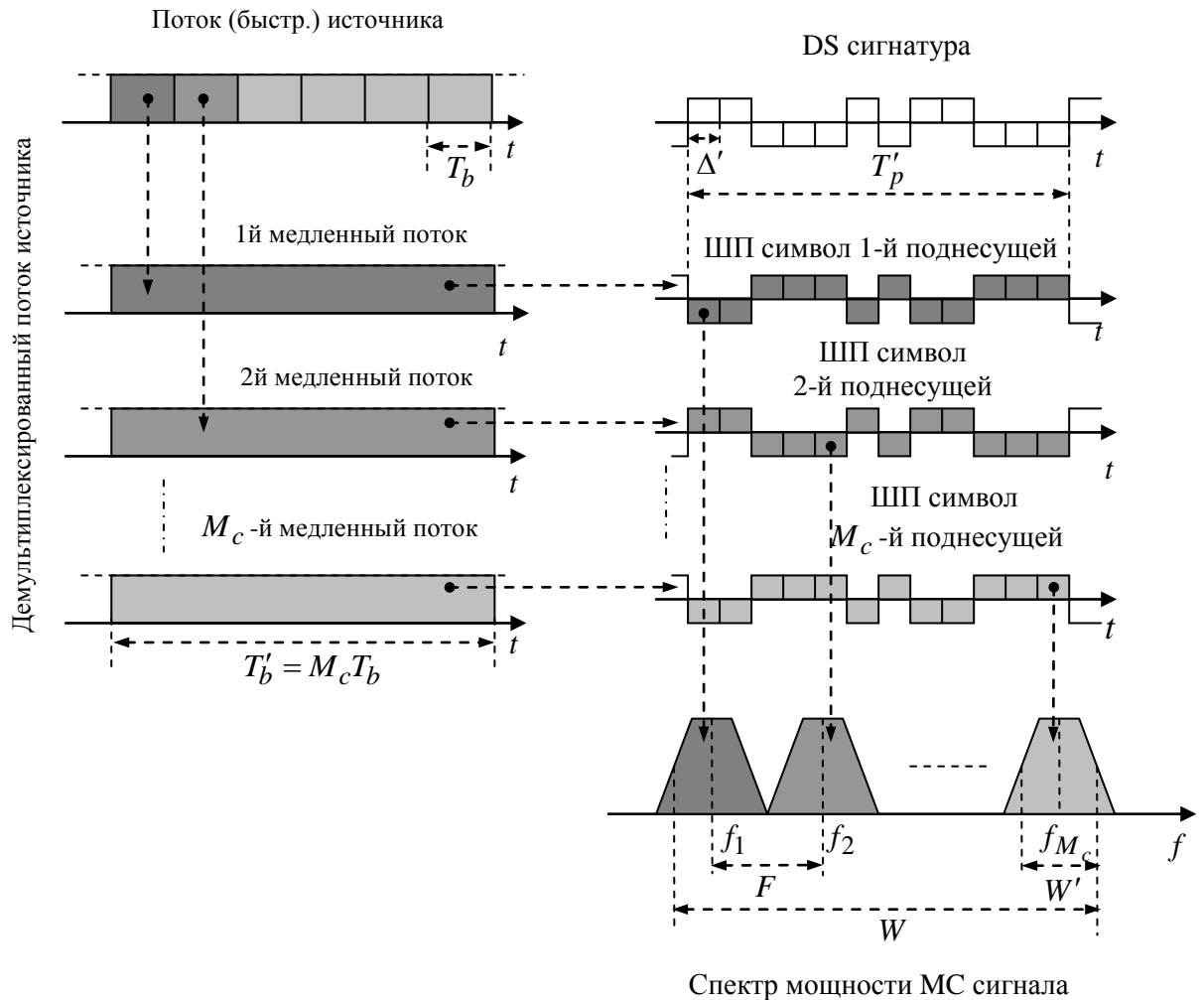


Рис. 10.2. Иллюстрация MC-DS-CDMA.

коэффициенту расширения MC-DS-CDMA составит

$$\frac{N}{N'} = 1 + \frac{M_c - 1}{M_c} \cdot \frac{F_g}{W'} = 1 + \frac{(M_c - 1)F_g}{W - (M_c - 1)F_g}, \quad (10.27)$$

что значительно превосходит единицу. Поскольку коэффициент расширения спектра является наиболее важным параметром CDMA, определяющим абонентскую емкость, помехоустойчивость и пр., MC-DS-CDMA оказывается не лучшим потенциальным способом утилизации доступного спектрального ресурса. Однако иногда могут перевесить и другие факторы, как это имеет место в спецификации cdma2000, рекомендующей MC-DS-CDMA для организации прямого канала, на основании взаимной совместимости с cdmaOne. В этом случае формат с единственной несущей cdmaOne (слегка модифицированный) просто повторяется на трех специально выбранных несущих.

10.2.2. Стандартная МС передача и OFDM.

Отступим на время от линии расширения спектра и CDMA с целью лучшего понимания причин, лежащих в основе значительного интереса к МС методам в современной беспроводной телекоммуникации. Предположим, что желательно передавать битовый поток источника с использованием обычного (не широкополосного) метода модуляции (BPSK, QPSK или др.). При M -ичной модуляции и необходимой скорости передачи данных R длительность импульса символа данных составит $T_p = (\log_2 M) / R$. Предположим, что канальная полоса когерентности B_c (см. параграф 3.6) значительно уже полосы символов данных ($B_c \ll W \approx 1/T_p = R / \log_2 M$) или, другими словами, задержка распространения τ_{\max} значительно превосходит длительность символа. Тогда при «прямой» передаче (см. рис. 10.3, а) будет наблюдаться глубокая межсимвольная интерференция (МСИ), искажающая многие символы данных следующие за текущим. Для противодействия этому эффекту приемник должен содержать достаточно сложный эквалайзер с большой памятью, как правило, реализуемый в виде адаптивного КИО фильтра, т.е. линию задержки с отводами и перестраиваемыми весами отводов.

Передача на многих несущих предлагает альтернативное решение (рис. 10.3, б), избавляющее от необходимости использования сложного эквалайзера. Снова осуществим демультиплексирование «быстрого» потока данных источника со скоростью R , образовав $M_c \geq W / B_c$ параллельных «медленных битовых потоков» со скоростью R / M_c каждый. Очевидно, что общая скорость, обеспечиваемая всеми медленными битовыми потоками, будет равна изначальной скорости, т.е. R . Возьмем теперь M_c поднесущих f_1, f_2, \dots, f_{M_c} , расположенных равномерно с интервалом $F = W' = W / M_c$, и используем каждую из них для передачи одного из M_c медленных битовых потоков с тем же типом модуляции, что и ранее. Каждая индивидуальная поднесущая образует отдельный подканал, функционирующий независимо от других и передающий медленный битовый поток более длинными импульсами (символами) длительности $T'_p = M_c T_p$, т.е. занимающий в M_c раз меньшую полосу $W' = W / M_c$, чем ранее. Последнее означает, что в пределах подканала замирания более не являются частотно-селективными, поскольку $W' = W / M_c \leq B_c$. При медленном замирании задержка распространения не превышает длительности одного импульса, МСИ не столь значительна, как это было первоначально, и может быть нейтрализована сравнительно простыми эквалайзерами. Общая полоса, занимаемая МС системой, примерно составляет $W \approx M_c / T'_p = 1/T_p = R / \log_2 M$, т.е. совпадает с полосой системы с одной несущей. Фактически спектральная эффективность МС системы может оказаться даже лучше, поскольку форма ее реального спектра близка к прямоугольной.

В рассматриваемом варианте не использованы никакие частотные защитные интервалы, более того, допускается даже перекрытие спектров подканалов, поскольку традиционная оценка полосы подканала $W' \approx 1/T'_p$ обычно оставляет значительные внеполосные боковые лепестки спектра. Однако взаимная помеха между подканалами может быть полностью подавлена. Предположим, что форма символьного импульса является прямоугольной и, по соглашению, его полоса определяется, как $W' = 1/T'_p$. Тогда частотный разнос между соседними поднесущими, равный $F = W' = 1/T'_p$, гарантирует ортогональность подканальных сигналов, т.е. полное исключение взаимных помех между МС подканалами. По этой причине данная версия метода МС модуляции получила название *мультиплексирования с ортогональным частотным разделением* – *orthogonal frequency division multiplexing* (OFDM).

Пусть модуляционные символы (комплексные амплитуды) M_c поднесущих OFDM определены как $b_i = A_i \exp(j\phi_i)$, $i=1,2,\dots,M_c$, где A_i и ϕ_i – соответственно вещественные амплитуда и фаза, а символы передаются прямоугольными импульсами. Тогда результи-

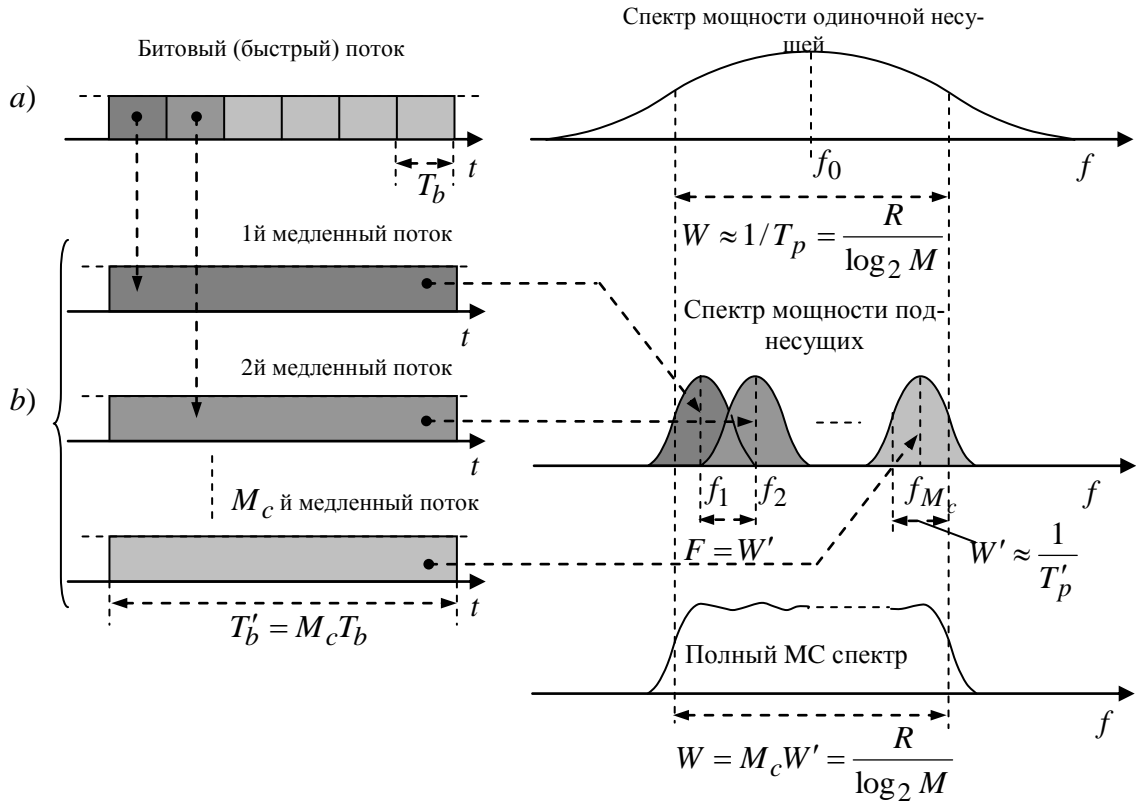


Рис.10.3. Передача данных на одной несущей (а) и на МС (b).

рующий сигнал в комплексном представлении, которое фактически является комплексной огибающей, перенесенной затем на некоторую центральную частоту f_0 , будет иметь вид

$$\dot{S}(t) = \sum_{i=1}^{M_c} b_i \exp(j2\pi f_i t) = \sum_{i=1}^{M_c} b_i \exp\left[\frac{j2\pi(i-1)t}{T'_p}\right], \quad (10.28)$$

где f_1 считается равной нулю и, следовательно, $f_i = (i-1)F = (i-1)/T'_p$. Последнее предположение не ведет к потере общности, поскольку окончательное значение центральной частоты устанавливается позднее по соглашению. Сигнал вида (10.28) представляет собой OFDM символ, причем число различных OFDM символов (объем OFDM алфавита M_{OFDM}) определяется объемом алфавита модуляционных символов M и числом частот: $M_{OFDM} = M^{M_c}$. Например, для бинарной OFDM $M_{OFDM} = 2^{M_c}$.

Взятие отсчетов из (10.28) с интервалом $T_s = T'_p / M_c$ дает последовательность

$$\dot{S}_l = \dot{S}(lT'_p / M_c) = \sum_{i=1}^{M_c} b_i \exp\left(\frac{j2\pi(i-1)l}{M_c}\right), \quad l=0,1,\dots,M_c-1, \quad (10.29)$$

повторяющую (за исключением несущественного постоянного коэффициента) обратное дискретное преобразование Фурье (ОДПФ) последовательности модуляционных символов $\{b_i\}$. Последнее объясняет основную причину популярности OFDM: реализация этого МС

метода не требует модулирования параллельными генераторами M_c несущих и суммирования результатов. Аналогичный выходной эффект получается с помощью ОДПФ символов модуляции. Таким образом, типичная структура OFDM передатчика (рис. 10.4) включает демультиплексор, блок ОДПФ, выдающий вектор ОДПФ (10.29), который затем преобразуется из параллельной в последовательную форму следующих друг за другом отсчетов и интерполируется для получения непрерывного OFDM символа (10.28). Последний (после введения префикса, см. ниже) преобразуется и передается на желаемой центральной

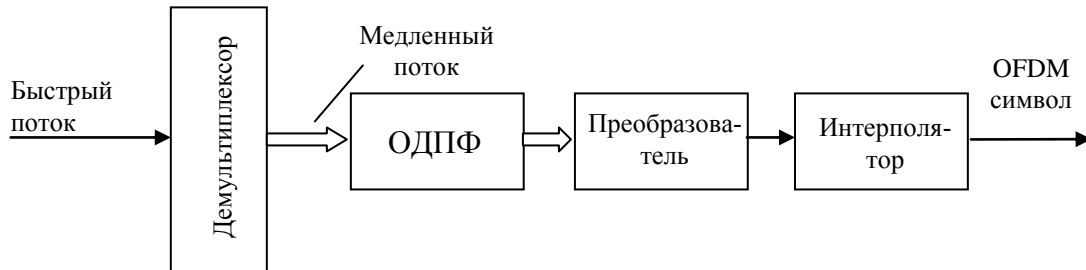


Рис.10.4. Формирование OFDM символа.

ной частоте f_0 .

На приемной стороне также отсутствует необходимость в использовании M_c параллельных приемников, каждый из которых настроен на свою собственную несущую, поскольку можно извлечь последовательность $\{b_i\}$ из (10.29) с помощью прямого дискретного преобразования Фурье (ПДПФ):

$$\sum_{l=0}^{M_c-1} \dot{s}_l \exp\left[-\frac{j2\pi(i-1)l}{M_c}\right] = M_c b_i, \quad i = 1, 2, \dots, M_c. \quad (10.30)$$

Последнее соотношение показывает, что блок ДПФ является подходящим устройством для раскладывания принятого OFDM символа на M_c подканальных результатов, необходимых для восстановления переданных данных. На выходе реального канала приемник еще не имеет в распоряжении «чистый» OFDM символ. Действительно, он наблюдает комплексную огибающую $\dot{Y}(t)$, содержащую OFDM символ, который искажен шумом и МСИ. Как следует из общего принципа, МС метод, увеличивая длительность символа, ограничивает глубину распространения МСИ на символ, следующий за текущим. Для исключения и этого остаточного влияния можно ввести защитный интервал длительности $T_g \geq \tau_{\max}$ между соседними МС символами. Этот интервал не обязательно должен быть пустым. Более того, заполнение его циклическим префиксом OFDM символа значительно облегчает нейтрализацию эффектов канального многолучевого распространения. Введение циклического префикса позволяет превратить свертку переданного сигнала и импульсного отклика канала в циклическую, соответствующую произведению их ДПФ отображений. Обозначим через $\nu = \lceil \tau_{\max} / T_s \rceil$ целое число интервалов взятия выборок при максимальной канальной задержки и добавим ν последних отсчетов (10.29) как префикс в передаваемый OFDM символ. Приемник будет пропускать первые ν отсчетов $\dot{S}_l, l = -\nu, -\nu+1, \dots, M_c-1$, так что многолучевые копии предшествующего OFDM символа не затронут M_c оставшихся отсчетов. Задержанные копии текущего символа не повлияют на более поздние отсчеты вследствие канального многолучевого распространения. Если отсчетами импульсного отклика канала являются $\dot{H}_0, \dot{H}_1, \dots, \dot{H}_\nu$, то любой отсчет наблюдения $\dot{Y}_l, l = 0, 1, \dots, M_c-1$ (при незначительном уровне шума) находится как сверт-

ка вида

$$\dot{Y}_l = \sum_{m=0}^v \dot{S}_{l-m} \dot{H}_m, \quad l=0,1,\dots,M_c-1. \quad (10.31)$$

Благодаря циклическому префиксу последовательность \dot{S}_{l-m} , $l=0,1,\dots,M_c-1$ представляет собой циклический сдвиг последовательности \dot{S}_l при любом $m=0,1,\dots,v$, следовательно (10.31) является циклической сверткой, и ее ДПФ–спектр есть произведение ДПФ \tilde{S}_i и \tilde{H}_i последовательностей \dot{S}_l и \dot{H}_l . ДПФ первой есть просто масштабированная последовательность модуляционных символов $\{b_i\}$ (см. (10.30)), тогда как

$$\tilde{H} = \sum_{l=0}^v \dot{H}_l \exp\left(\frac{j2\pi(i-1)l}{M_c}\right)$$

есть передаточная функция канала на частотах $f_i = f_1, f_2, \dots, f_{M_c}$, так что

$$\tilde{Y}_i = M_c b_i \tilde{H}_i, \quad i=1,2,\dots,M_c. \quad (10.32)$$

Данный результат свидетельствует, что для исключения влияния канала на OFDM сигнал, т.е. осуществления эквалайзинга, достаточно просто поделить каждый отсчет на выходе блока ДПФ приемника на значение передаточной функции канала \tilde{H}_i на соответствующей частоте.¹ Для определения текущего состояния канала, т.е. \tilde{H}_i , как правило, используется вполне определенная обучающая процедура. Отметим также, что поскольку защитный интервал создает нечто подобное заголовку, уменьшая скорость передачи данных, часто бывает рационально увеличить значение M_c , превращая защитный интервал в малую часть длительности OFDM символа.

При передаче данных с использованием фазовой манипуляции и отдельной обработке подканальных сигналов, т.е. независимо от других использований i -го компонента ДПФ (10.32) при демодуляции модуляционного символа b_i , вышеупомянутое выравнивание может быть упрощено до простой компенсации канального фазового сдвига, поскольку знание амплитуды поднесущей избыточно для принятия решения. В отличие от рассмотренного, если необходима совместная обработка подканалов, например, в случае MC, основанного на CDMA, что рассматривается ниже, амплитуды \tilde{H}_i играют особую важность, и последнее выравнивание может оказаться предпочтительным.

Суммируя вышесказанное, можно представить структуру OFDM приемника в виде, изображенном на рис. 10.5. Строблирующее устройство (временной селектор) обеспечивает взятие отсчетов \dot{Y}_l , префикс которых затем отбрасывается. Далее последовательность отсчетов трансформируется в параллельную форму. Блок прямого ДПФ выдает компоненты ДПФ спектра \tilde{Y}_i , которые представляют собой искаженные шумом и канальными эффектами символы данных b_i . Поэтому после выравнивания (которое состоит в простом умножении на соответствующий весовой коэффициент w_i^*) они могут послужить основой для выработки оценок символов данных \hat{b}_i , осуществляемой обычным образом для BPSK, QPSK, QAM или любого другого вида модуляции.

¹ Подобное выравнивание передаточной функции канала известно как нуль–форсирующее выравнивание, ранее упомянутое в параграфе 6.1.2.

10.2.3. Метод CDMA со многими несущими.

Модуляционная схема с МС легко адаптируема к многопользовательской среде для обеспечения мультиплексирования с кодовым разделением. В отличие от DS CDMA, где соответствующий выбор формы сигнатур во временной области обеспечивает разделение пользовательских сигналов, в CDMA со многими несущими (МС-CDMA) сигнатуры формируются в частотной области, контролируя амплитуды и фазы поднесущих специфическим для каждого потребителя образом. Одним из способов пояснения МС-CDMA служит связь DS-CDMA с МС передачей. Возвратимся к рис. 10.3 и представим, что вместо быстрого потока бит источника имеем поток символов данных k -го пользователя, расширенный с помощью k -й пользовательской DS сигнатуры длины (коэффициента расширения) N . В результате имеем N DS чипов длительности Δ на один символ данных k -го пользователя длительности T_p . Демультимплексируем этот быстрый DS широкополосный поток в $M_c = N$ медленных потоков, развертывая каждый символ данных в $M_c = N$ параллельных чипов большой длительности $\Delta' = T_p = N\Delta$. Каждый из этих параллельных медленных потоков чипов далее передается МС (OFDM) способом, так что каждый пользователь обладает своим специфическим законом модуляции поднесущих.

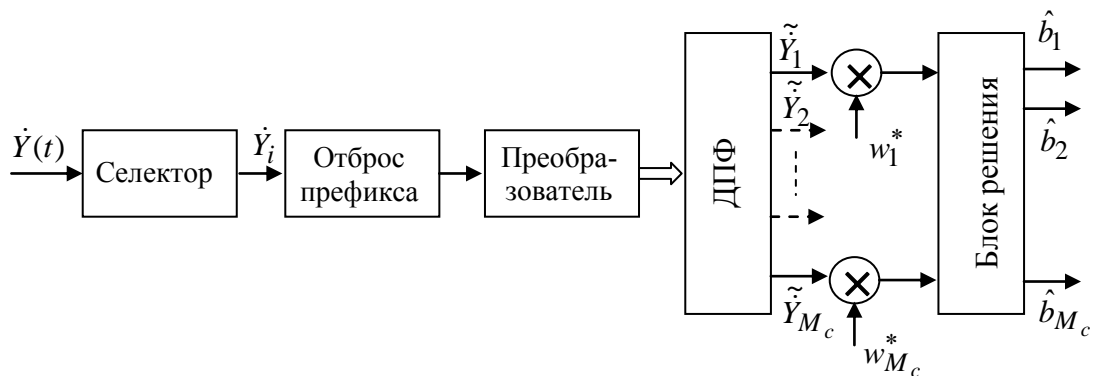


Рис.10.5 Структура OFDM приемника.

Опишем теперь вышесказанное более подробно. Пусть b_k – текущий символ данных k -го пользователя. Пусть $\mathbf{a}_k = (a_{k,0}, a_{k,1}, \dots, a_{k,N-1})$ – вектор сигнатуры k -го пользователя, используемый теперь в частотной области. Для формирования МС-CDMA сигнала N компонент вектора $b_k \mathbf{a}_k$ параллельно манипулируют по амплитуде и фазе $M_c = N$ поднесущие f_1, f_2, \dots, f_N в течение длительности импульса T_p . Суммирование всех манипулированных поднесущих дает i -й МС-CDMA символ длительности T_p , передаваемый k -м пользователем. Очевидно, что при $F = 1/T_p$ реализация МС-CDMA более осуществима в типичной для OFDM форме на основе ДПФ, однако прямой способ формирования МС-CDMA сигнала, показанный на рис. 10.6 для случая вещественного (например, BPSK) алфавита символов данных и сигнатур, является более прозрачным для иллюстрации идеи. Ее обобщение на комплексный алфавит осуществляется без особых проблем. При реализации OFDM варианта блок ОДПФ заменяет многоканальную структуру, представленную на рис. 10.6.

В синхронных не перенасыщенных системах, подобных каналу вниз мобильного радио, любое множество $K \leq N$ ортогональных векторов сигнатур (функции Уолша, например) могло бы обеспечить свободное от воздействия помех MAI разделение OFDM МС-CDMA пользовательских сигналов, поскольку ортогональность ДПФ спектра гаран-

тирует ортогональность OFDM символов. Выбор сигнатур в асинхронных системах (например, каналах «вверх» мобильного радио) не столь прямолинеен, хотя характеристики некоторых минимаксных ансамблей сигнатур асинхронных DS CDMA (см. параграф 7.5) тоже могут вызвать интерес и для MC-CDMA [105–107]. Существует еще одно осложнение, связанное с конструированием MC-CDMA сигнатур и особенно важное для каналов «вверх» мобильного радио: вещественная огибающая MC-CDMA сигнала в отличие от аналогичного параметра DS-CDMA сигнала подвержена значительным изменениям, приводя к пик-фактору, значительно большему единице. Необходимо учитывать эту особенность и, при прочих равных характеристиках, из множества кандидатов на роль множества сигнатур следует отдавать предпочтение такому ансамблю, который обеспечивает наименьшее значение пик-фактора.

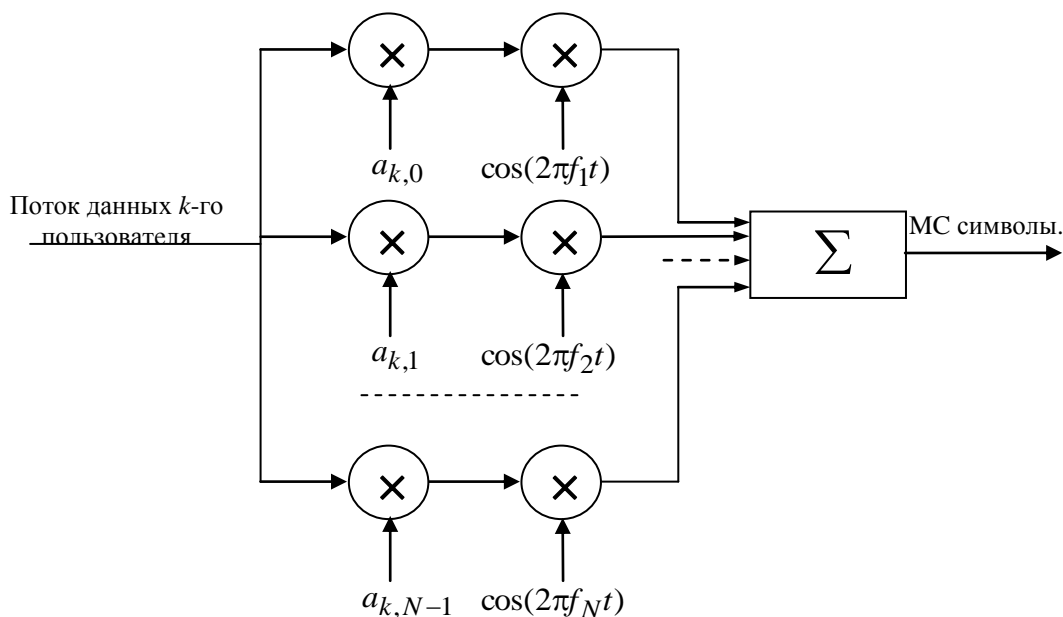


Рис.10.6 Формирование MC-CDMA сигнала k -го пользователя.

На рис. 10.7 представлена общая (не на основе ДПФ) структура MC-CDMA приемника k -го пользователя. Она содержит $N = M_c$ каналов, каждый из которых настроен на свою собственную частоту и реализован в виде комплексного коррелятора, обрабатывающего комплексную огибающую наблюдения $\dot{Y}(t)$. Для борьбы с медленными замираниями в рамках каждого частотного подканала комплексная величина $\dot{z}_{k,i}$ с выхода i -го коррелятора взвешивается комплексным коэффициентом w_i^* и умножается на комплексно сопряженный символ сигнатуры $a_{k,i}^*$. Последняя операция есть ничто иное, как снятие расширения в частотной области. Суммирование подобных произведений по всем подканалам дает статистику \dot{z}_k , которая используется при оценке текущего символа данных b_k k -го пользователя. Снова отметим, что при реализации с ДПФ блок прямого ДПФ заменяет множество корреляторов.

Кратко коснемся теперь проблемы выбора весовых коэффициентов w_i^* , $i = 1, 2, \dots, N$. В случае MC-CDMA варианта эта задача является отчасти более сложной, чем при обычной MC передаче вследствие необходимости контроля уровня помех MAI. Предположим, что сигналы всех пользователей распространяются по одному и тому же каналу, как это имеет место, например в прямом канале мобильной системы радиосвязи. Поскольку разнос поднесущих F не менее чем полоса когерентности канала, то значе-

ния канальной передаточной функции $\tilde{H}_i, i = 1, 2, \dots, N$ на поднесущих частотах могут независимо распределяться в большом диапазоне. Нормирование канальной передаточной функции, как $N^{-1} \sum_{i=1}^N |\tilde{H}_i|^2 = 1$, позволит характеризовать k -й сигнал независимо от интенсивности амплитуды A_k . Тогда мощности P_k, P_{nk}, P_l , создаваемые на выходе k -го

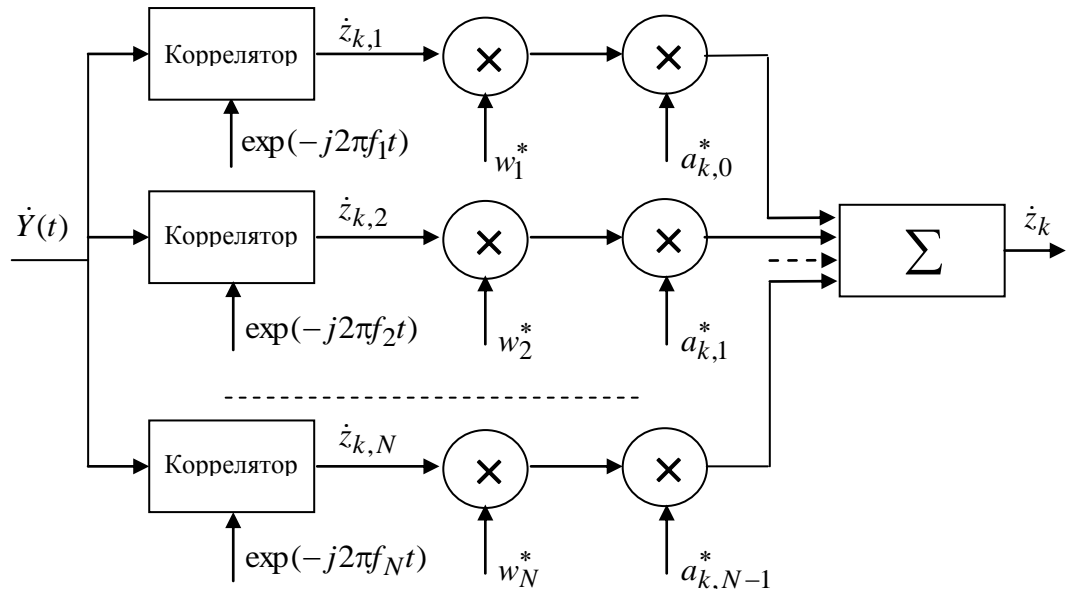


Рис.10.7. Общая схема MC-CDMA приемника.

приемника полезным сигналом, шумом и l -й помехой MAI соответственно, вычисляются следующим образом

$$P_k = A_k^2 \left| \sum_{i=1}^N |a_{k,i-1}|^2 \tilde{H}_i w_i^* \right|^2, \quad P_{nk} = \sigma^2 \sum_{i=1}^N |a_{k,i-1}|^2 |w_i|^2, \quad P_l = A_l^2 \left| \sum_{i=1}^N a_{l,i-1} a_{k,i-1}^* \tilde{H}_i w_i^* \right|^2,$$

где σ^2 – мощность подканального шума. Теперь видно, что даже если сигнатуры изначально ортогональны:

$$\sum_{i=0}^{N-1} a_{l,i} a_{k,i}^* = \delta_{kl},$$

частотная селективность по каналам может нарушить ортогональность, усиливая одни и подавляя другие поднесущие. В результате возникают помехи MAI, так что не все $P_l, l \neq k$ будут нулевыми. Для сохранения ортогональности сигнатур на приемной стороне независимо от текущего состояния канала следует выбирать $w_i^* = 1/\tilde{H}_i, i = 1, 2, \dots, N$, т.е. реализовать нуль-форсирующее выравнивание, полностью компенсирующее каналные эффекты. Однако это приводит к рассогласованности обработки, если канальная амплитудная передаточная функция неравномерна, а, значит, платой за полное подавление помех MAI является потери в отношении сигнал-шум $q_{k,zf}^2$, соответствующие нуль-форсирующей комбинации:

$$q_{k,zf}^2 = \frac{P_k}{P_{nk}} = \frac{A_k^2}{\sigma^2} \frac{\left| \sum_{i=1}^N |a_{k,i-1}|^2 \right|^2}{\sum_{i=1}^N |a_{k,i-1}|^2 |\tilde{H}_i|^{-2}} \leq q_{k,mf}^2 = \frac{A_k^2}{\sigma^2} \sum_{i=1}^N |a_{k,i-1}|^2 |\tilde{H}_i|^2,$$

где $q_{k,mf}^2$ – отношение сигнал-шум по мощности, достижимое при согласованной обработке, т.е. комбинации с максимальным отношением (параграф 3.6.1) $w_i^* = \tilde{H}_i^*$. Для случая первоначально равных интенсивностей всех компонентов поднесущих $|a_{k,i}| = 1/\sqrt{N}, i = 1, 2, \dots, N$ потери в отношении сигнал-шум $\gamma = q_{k,mf}^2 / q_{k,zf}^2$ комбинации с нуль-форсированием относительно комбинации с максимальным отношением (см. (6.42)) составят

$$\gamma = \frac{1}{N} \sum_{i=1}^N |\tilde{H}_i|^{-2}. \quad (10.33)$$

Из (10.33) видно, что при заметной неравномерности канальных амплитудно-частотных искажений (некоторые $|\tilde{H}_i|$ близки к нулю), потери в отношении сигнал-шум, т.е. плата за радикальное устранение помех МАИ, могут оказаться не допустимыми, и более рациональным является отыскание компромисса между уровнями подавленных помех МАИ и шумом. Один из подобных подходов ведет к МСКО выравниванию, основная идея которого аналогична рассмотренной в предыдущем разделе применительно к многопользовательскому обнаружению. Детали данного подхода, также как и дальнейшее ознакомление с широкополосной МС философией заинтересованный читатель может найти в [105-106] и приведенных там многочисленных ссылках.

В заключение вновь подчеркнем, что не существует жестких барьеров между DS и МС CDMA. Они представляют собой лишь параллельные технические пути достижения одного и того же результата: широкополосной сигнатуры. Последняя всегда может быть синтезирована либо как суперпозиция гармоник в частотной области (МС), либо непосредственным заданием формы во временной области (DS).

10.2.4. Применение.

В настоящее время проникновение МС технологии в цифровые телекоммуникации очень широко. Среди примеров ее практического использования можно упомянуть стандарты цифрового аудио- и видеовещания DAB, DVB-T и др. Накопленный к настоящему времени позитивный опыт обещает большую привлекательность CDMA версиям на основе МС. В частности, метод МС-CDMA рассматривается сейчас в качестве одного из наиболее вероятных вариантов эфирного интерфейса мобильной связи 4-го поколения.

10.3. Разнесение на передаче и пространственно-временное кодирование в CDMA системах.

10.3.1. Разнесение на передаче и задача пространственно-временного кодирования.

Из краткого обсуждения, проведенного в параграфе 3.7, видно, что привлечение нескольких приемных и передающих антенн является продуктивным путем организации ветвей разнесения, необходимых для противодействия деструктивным эффектам замирания. На практике для реализации указанной цели используются антенные решетки, состоящие из элементов, расположенных на расстояниях в несколько длин волн, что обеспечивает независимость их многолучевых образцов. Термин *множественный вход – множественный выход* (*multiple input – multiple output* (MIMO)) служит для определения системы, совместно обрабатывающей сигналы, принятые несколькими приемными антеннами от нескольких передающих антенн. Рис. 10.8 иллюстрирует общую концепцию MIMO канала, содержащего n_T передающих и n_R приемных антенн соответственно. Подканал

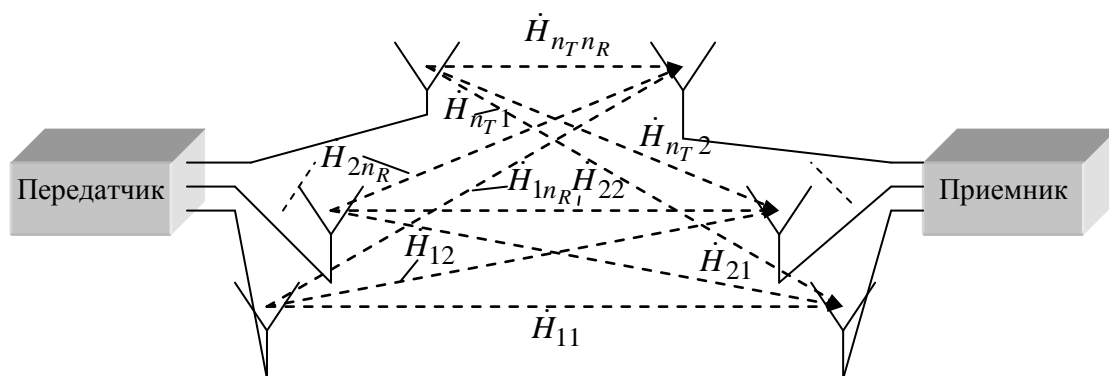


Рис.10.8. Обобщенная модель MIMO системы.

данной системы образуют i -я передающая и j -я приемная антенны, текущее состояние которого характеризуется комплексным коэффициентом замирания \dot{H}_{ij} , $i = 1, 2, \dots, n_T$, $j = 1, 2, \dots, n_R$, который в общем случае может зависеть как от времени, так и от частоты. Как правило, благодаря наличию специального пилот-сигнала, приемнику известно состояние канала, и он может использовать коэффициенты \dot{H}_{ij} для эффективной совместной обработки (т.е. комбинирования) подканальных сигналов. Что касается передатчика, то он может быть уведомлен о состоянии канала и способен адаптировать сигнал к текущим условиям распространения только при условии организации надежной информационной обратной связи «приемник-передатчик» (замкнутой петли разнесения по передаче).

Предположим, что имеется только одна передающая и n_R приемных антенн и, следовательно, n_R подканалов с коэффициентами замирания $\dot{H}_1, \dot{H}_2, \dots, \dot{H}_{n_R}$. Тогда использование потенциала разнесения на приеме в самом общем виде заключается только в соответствующем комбинировании сигналов, принятых параллельными антеннами. Данная задача, по крайней мере в принципе, не вызывает особых проблем, поскольку приемник наблюдает сигналы различных антенн на отдельных выходах и (зная состояния всех ветвей разнесения \dot{H}_j) может обработать их наилучшим из возможных способов. Ситуация кардинально изменяется в случае привлечения также и разнесения на передаче. При одновременной работе параллельных передающих антенн приемник сталкивается с проблемой

разделения перекрывающихся друг с другом на входе каждой приемной антенны сигналов, что необходимо для использования в дальнейшем знания о состоянии подканала \dot{H}_{ij} и осуществления комбинирования подканальных сигналов соответствующим образом. Для обеспечения возможности подобного разделения передача данных по n_T параллельно передающим антеннам должна быть тщательно организована и нахождение путей достижения этого составляют сущность задачи, называемой *пространственно-временным кодированием (space-time coding)*. Название отражает тот факт, что группа передаваемых бит данных однозначно отображается в двумерное $n_T \times n$ кодовое слово $[u_t^i]$. Элементы массива u_t^i с индексами i, t являются кодовыми символами, передаваемыми i -й антенной в t -й момент времени, а n – длина кода. Отметим, что в некоторых случаях разнесение на приеме может оказаться неосуществимым, например, в прямом канале мобильного радио, где малые размеры аппарата не дают достаточного пространства для нескольких приемных антенн. В подобных ситуациях разнесение на передаче и, следовательно, адекватное пространственно-временное кодирование становится особенно значимым. В материале, рассматриваемом ниже, будем полагать, что используется только одна приемная антенна с тем, чтобы сконцентрироваться только на исследовании эффективности разнесения на передаче. Данное предположение позволяет упростить обозначения коэффициентов канального замирания, оставив только один индекс, указывающий номер передающей антенны: $\dot{H}_{i1} = \dot{H}_i$.

10.3.2. Эффективность разнесения на передаче.

Многочисленные исследования были предприняты по оцениванию пропускной способности Шеннона, т.е. потенциальной скорости безошибочной передачи данных, ММО каналов и была доказана продуктивная роль множественности антенн для основных моделей замираний [108-110]. Нет ничего удивительного в полезности разнесения на приеме, поскольку дополнительные приемные антенны утилизируют энергию сигнала в тех точках пространства, которая была бы безвозвратно потеряна в случае единственной антенны. При комбинировании с максимальным отношением n_d идентичных приемных ветвей разнесения среднее отношение сигнал-шум по мощности возрастает в n_d раз (см. пункт 3.6.1), и хотя данный фактор не является определяющим в снижении вероятности ошибки и канальной емкости, однако делает данное улучшение легко предсказуемым. В отличие от предшествующего природа выигрыша в пропускной способности или уменьшении вероятности ошибки за счет разнесения на передаче не столь очевидна, учитывая распределение ограниченного мощностного ресурса между несколькими передающими антеннами. Действительно, при фиксированной полной мощности выигрыш в среднем отношении сигнал-шум при комбинировании с максимальным отношением идентичных ветвей разнесения отсутствует за исключением случая, когда передатчику известно состояние канала, и он способен координировать передачу по различным ветвям так, чтобы на приемной антенне сигналы подканалов суммировались бы когерентно. В самом деле, пусть информация о состоянии канала недоступна передатчику, а полная мощность P равномерно распределяется между n_d идентичными ветвями разнесения (антеннами, частотными каналами и т.п.). Тогда среднее мощностное отношение сигнал-шум на одну ветвь составляет q^2/n_d , где q^2 – среднее отношение сигнал-шум по мощности, которое существовало бы на входе приемника в отсутствии разнесения. Очевидно, что комбинирование с максимальным отношением способно только увеличить среднее мощностное отношение сигнал-шум на ветвь в n_d раз, т.е. только сделать его равным отношению сигнал-шум в отсутствии разнесения.

Таким образом, в случае разнесения на передаче (также как и при частотном разнесении) имеют место две противоречивые тенденции. С одной стороны, при заданном уровне полной мощности увеличение числа ветвей приводит к большему числу независимых подканалов, которые, поддерживая друг друга, обеспечивают большую вероятность того, что, по крайней мере, некоторые из них будут энергетически состоятельными. С другой стороны, условия (отношение сигнал-шум в ветви) в каждом из подканалов разнесения становятся хуже с ростом n_d . Тщательный теоретический анализ показывает, что первый из этих факторов перевешивает второй. Оставим в стороне математические выкладки, касающиеся пропускной способности канала, с которыми можно ознакомиться в литературных источниках (например, [108-110]), однако в задачах 10.11 и 10.12 содержатся простейшие примеры, иллюстрирующие данный вопрос. Что же касается положительного воздействия разнесения при передаче на вероятность ошибки, то оно становится очевидным из следующего рассмотрения передачи данных с помощью бинарной ФМ по релеевскому каналу с замираниями.

Пусть $A_i = |\dot{H}_i|$ – амплитуда коэффициента замирания i -й ветви разнесения со средним квадратом, нормированным как $\overline{A_i^2} = 1$. Тогда при комбинировании с максимальным отношением текущее отношение сигнал-шум на входе приемника (см. (3.15)) составит $q_r^2 = \sum_{i=1}^{n_d} A_i^2 (q^2 / n_d)$ и, согласно (2.19), условная вероятность ошибки на бит $P_e(A_1, A_2, \dots, A_{n_d})$ при фиксированных состояниях подканалов $A_i, i = 1, 2, \dots, n_d$ будет

$$P_e(A_1, A_2, \dots, A_{n_d}) = Q(q_r). \quad (10.34)$$

Для получения безусловной вероятности ошибки на бит P_e необходимо усреднить (10.34) по всем подканальным амплитудам $A_i, i = 1, 2, \dots, n_d$, используя их совместную плотность распределения $W(A_1, A_2, \dots, A_{n_d})$. Вследствие независимости ветвей разнесения данная плотность распределения есть просто произведение n_d одномерных плотностей распределения всех амплитуд, так что

$$P_e = \int_0^\infty \int_0^\infty \dots \int_0^\infty Q(q_r) \left[\prod_{i=1}^{n_d} W(A_i) \right] dA_1 dA_2 \dots dA_{n_d}. \quad (10.35)$$

Для разделения переменных интегрирования в (10.35) аппроксимируем дополнительную функцию ошибок $Q(x)$ ее верхней границей (см. задачу 10.13) $Q(x) \leq (1/2) \exp(-x^2/2)$, $x \geq 0$, приходя к соотношению вида

$$P_e \leq \frac{1}{2} \prod_{i=1}^{n_d} \int_0^\infty \exp\left(-\frac{A_i^2 q^2}{2n_d}\right) W(A_i) dA_i.$$

Для канала с релеевскими флюктуациями плотность распределения $W(A_i)$ подчиняется закону (3.12), подстановка которого приводит к соотношению

$$P_e \leq \frac{1}{2} \prod_{i=1}^{n_d} \int_0^\infty 2A_i \exp\left[-A_i^2 \left(1 + \frac{q^2}{2n_d}\right)\right] dA_i = \frac{1}{2} \left(\frac{2n_d}{q^2 + 2n_d}\right)^{n_d}. \quad (10.36)$$

Рис. 10.9 иллюстрирует поведение вероятности ошибки на бит в зависимости от общего среднего отношения сигнал-шум q для 1, 2, 3, 4, 5 и 6 ветвей разнесения. Из приведенных графиков следует чрезвычайно важный факт, свидетельствующий, что при фиксированной вероятности ошибки две ветви разнесения обеспечивают значительный энер-

гетический выигрыш. Например, если допустимая ошибка на бит составляет величину, не большую 10^{-4} , то организация двух ветвей разнесения снижает необходимую энергию излучения более чем на 15 дБ. В то же время при дальнейшем увеличении числа ветвей разнесения энергетический выигрыш растет с меньшей скоростью и, например, переход от 5 к 6 ветвям обещает сокращение излучаемой энергии лишь на 1 дБ. Последнее служит основанием применения во многих реальных системах (например, в прямом канале мобильного радио) только двух передающих антенн, что рассматривается как удовлетворительный компромисс между выигрышем от разнесения и сложностью оборудования.

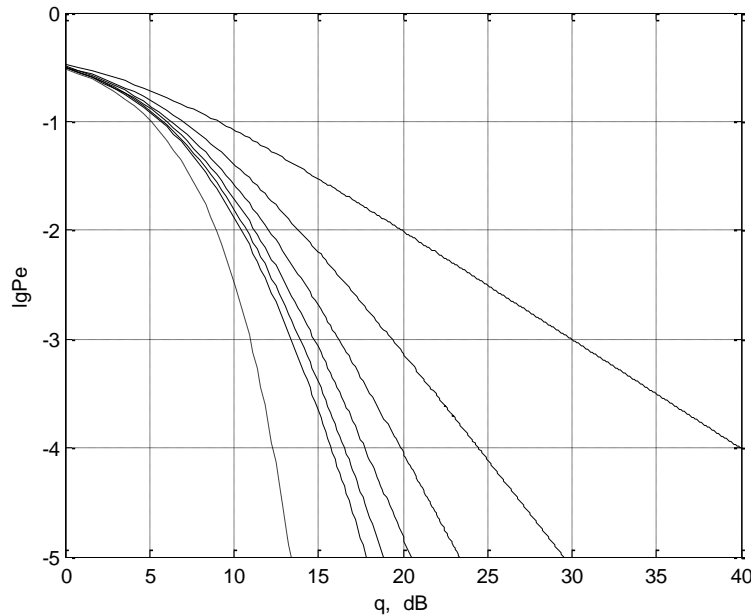


Рис.10.9. Зависимость вероятности ошибки на бит от общего отношения сигнал-шум в схеме разнесения на передаче для 1,2,3,4,5 и 6 ветвей разнесения.

Отметим, что при неограниченном увеличении числа ветвей разнесения правая часть соотношения (10.36) обращается в $(1/2)\exp(-q^2/2)$ (см. задачу 10.14), т.е. верхнюю границу (пунктирная линия на рис. 10.9) вероятности ошибки на бит для случая гауссовского канала без замираний. Другими словами, увеличение числа ветвей разнесения при передаче может в пределе (по крайней мере теоретически) полностью устранить деструктивный эффект множественности путей распространения.

10.3.3. Пространственно–временной код с переключением во времени.

В зависимости от модели замирания возможны различные подходы к конструированию пространственно–временных кодов. Быстрые замирания (см. параграф 3.5) предполагают настолько быстрые изменения образца многолучевого распространения, что значения коэффициента замирания на двух соседних интервалах символов одного и того же подканала оказываются независимыми. При дальнейшем обсуждении будем считать, что имеет место противоположный случай медленных замираний, полагая, что коэффициенты замирания подканалов остаются постоянными на всей длительности кодового слова.

Для подчеркивания не тривиального характера задачи конструирования кодов, обеспечивающих разделимость на приемной стороне сигналов различных передающих антенн, начнем с рассмотрения простейшего примера.

Пример 10.1.1. Пусть фиксированный мощностной ресурс P в равных частях де-

лится между $n_T = 2$ излучающими антеннами, одновременно передающими один и тот же символ данных без принятия мер, позволяющих разделить сигналы подканалов в приемнике (пространственно–временной код с повторением). Пусть используется единственная приемная антенна, а коэффициенты замирания \dot{H}_1, \dot{H}_2 двух подканалов – независимые гауссовские комплексные величины с нулевым средним и одинаковыми дисперсиями. Данная ситуация в точности отвечает случаю релейских замираний, поскольку амплитуды $A_i = |\dot{H}_i|$, $i=1,2$ подчиняются релейской плотности вероятности. Если $\dot{S}(t)$ – комплексная огибающая передаваемого сигнала, то комплексная огибающая принятого имеет вид $(\dot{H}_1 + \dot{H}_2)\dot{S}(t)/\sqrt{2}$, где корень квадратный из двух отвечает разделению мощности. Теперь становится очевидным, что два подканала образуют результирующий канал с общим коэффициентом замираний $\dot{H} = (\dot{H}_1 + \dot{H}_2)/\sqrt{2}$, который снова является гауссовской величиной с нулевым средним и такой же дисперсией, как и у любого из H_i . Следовательно, амплитуда $A = |\dot{H}|$ распределена по релейскому закону, а результирующий канал опять является каналом с релейскими замираниями. Если средние квадраты A_i нормированы к единице, то и $\overline{A^2} = 1$ тоже. Таким образом, результирующий канал абсолютно идентичен каждому из подканалов и использование в рассматриваемой ситуации двух антенн не может дать какого-либо преимущества по сравнению со случаем единственной антенны. Другими словами, код с повторением является вырожденным, не обеспечивающим реального разнесения. \square

Как показывает рассмотренный пример, число эффективных ветвей разнесения может оказаться меньше числа передающих антенн. Одним из наиболее важных параметров любого пространственно–временного кода служит *коэффициент разнесения (diversity gain)*, т.е. реально существующее число ветвей разнесения, обеспечиваемое схемой кодирования (см. задачу 10.15).

Очевидным путем обеспечения разделения сигналов различных передающих антенн в приемнике является передача одного и того же символа данных n_T антеннами по очереди, т.е. без перекрытия во времени. Другими словами, одна и только одна из передающих антенн излучает сигнал в каждый момент времени, используя весь мощностной ресурс. Данная ситуация абсолютно подобна схеме многопользовательских TDMA систем связи, т.е. сигнал каждой антенны идентифицируется своей временной позицией и ортогонален другим вследствие отсутствия перекрытия во временной области. Следовательно, приемник наблюдает все сигналы подканалов не в смеси, а просматривает их последовательно во времени без взаимной интерференции. Тогда, зная текущие коэффициенты замираний \dot{H}_i , приемник полностью в курсе того, какой из сигналов более или менее достоверен и способен осуществить их комбинирование любым подходящим образом, например, максимизировать отношение сигнал-шум, применяя взвешивание с максимальным отношением. Подобная простейшая схема кодирования отвечает *пространственно–временному коду с переключением во времени (time-switched space-time code)*.

Пример 10.3.2. Снова рассмотрим ситуацию с двумя антеннами ($n_T = 2$), повторно передающими один и тот же текущий символ данных, но работающими на этот раз в перемежающемся режиме: когда первая излучает энергию, вторая находится в пассивном состоянии и наоборот. Приемник последовательно один за другим наблюдает сигналы, прошедшие подканалы с коэффициентами замираний \dot{H}_1 и \dot{H}_2 и искаженные аддитивным шумом мощности σ^2 . Для реализации процедуры комбинирования с максимальным отношением эти наблюдения суммируются с весами \dot{H}_i^* , $i=1,2$ соответственно, приводя в итоге к отношению SNR по мощности (см. (3.15)) $q_r^2 = (A_1^2 + A_2^2)q_s^2/2$, где q_s^2 – полное

«без замираний» отношение SNR по мощности, приходящееся на один символ данных, которое уполовинено в результате деления фиксированной энергии символа между двумя антеннами. Для идентичных подканалов с амплитудами, нормированными к единице ($A_i^2 = 1$), среднее отношение SNR по мощности, конечно, вновь будет таким, как и в случае передачи всей энергии символа только по единственному подканалу: $\overline{q_r^2} = 2q_s^2/2 = q_s^2$. Однако если выход устройства комбинирования рассматривать как выход результирующего канала, то его сигнал более не является релейским и имеет вероятность ошибки меньшую, чем в случае распределения Релея при том же значении SNR (см. рис. 10.9). Таким образом, реально существуют две ветви разнесения, и они обеспечивают предсказанный выигрыш, который достигается в обмен за два раза меньшую скорость передачи на единицу полосы. Действительно, при фиксированной скорости передачи в бит/сек каждая антенна передает теперь любой символ данных за длительность $T_p/2$, т.е. занимает удвоенную полосу. Еще одним показательным примером является сравнение с системой, использующей две ветви разнесения на приеме ($n_T = 1, n_R = 2$). Легко увидеть, что если единственная передающая антенна использует полный интервал T_p для передачи текущего символа, то приемное устройство, реализующее комбинирование с максимальным отношением, обеспечивает среднее SNR по мощности $\overline{q_r^2} = 2q_s^2$, т.е. на 3 дБ выше, чем при аналогичном разнесении на передаче при одинаковых энергиях на символ. Причина энергетических потерь разнесения на передаче по сравнению с разнесением на приеме уже неоднократно указывалась и заключается в разделении общей фиксированной энергии между антеннами в схеме разнесения на передаче. \square

Пространственно–временные коды с переключением во времени чрезвычайно просты, однако они реализуют максимальный выигрыш n_T от разнесения в обмен на расширение полосы (при фиксированной скорости) и прерывистость передачи, т.е. увеличение пик–фактора сигнала. Указанные причины в значительной степени мотивируют поиск пространственно–временных кодов, допускающих разделение сигналов от различных антенн вне зависимости от их перекрытия во времени. Простейший, но и очень важный пример подобных кодов представлен в следующем пункте.

10.3.4. Пространственно–временные коды Аламоути.

Схема кодирования, предложенная в [111], использует две передающие антенны, функционирует без дополнительного расширения полосы и реализует максимально возможный для двух антенн выигрыш от разнесения $n_d = n_T = 2$. Пусть b_0 и b_1 – два последовательных символа данных, стоящих на четной и нечетной временной позиции соответственно и принадлежащих некоторому фиксированному модуляционному алфавиту (PSK, QAM и т.п.). Кодовые слова *пространственно–временного кода Alamouti* представляют собой 2×2 массивы вида

$$\mathbf{u} = \begin{bmatrix} b_0 & -b_1^* \\ b_1 & b_0^* \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix}, \quad (10.37)$$

означающего, что длина кода $n = 2$. Как видно, на четном символьном интервале две антенны одновременно передают кодовые символы $u_0^1 = b_0$ (первая антенна) и $u_0^2 = b_1$ (вторая антенна), тогда как на нечетном интервале излучаемыми символами являются $u_1^1 = -b_1^*$ (первая антенна) и $u_1^2 = b_0^*$ (вторая антенна). В альтернативном представлении антенны одновременно передают последовательности длины 2 $\mathbf{u}_1 = (u_0^1, u_1^1) = (b_0, -b_1^*)$

(первая антенна) и $\mathbf{u}_2 = (u_0^2, u_1^2) = (b_1, b_0^*)$ (вторая антенна). Подобное построение превращает последовательности, излучаемые двумя антеннами, т.е. вектора \mathbf{u}_1 и \mathbf{u}_2 , в ортогональные $(\mathbf{u}_1, \mathbf{u}_2) = b_0 b_1^* - b_1^* b_0 = 0$, гарантируя разделимость в приемнике перекрывающихся сигналов различных подканалов. На самом деле, однако, отсутствует необходимость осуществлять разделение подканалов в виде специальной процедуры, поскольку оптимальное (МП) обнаружение символов данных b_0 и b_1 автоматически включает ее, также как и комбинирование с максимальным отношением. По очевидной причине положим, что одиночный кодовый символ, излучаемый в текущий момент одной антенной, в среднем использует половину общей средней энергии символа E_s . Пусть $\dot{\mathbf{Y}} = (\dot{Y}_0, \dot{Y}_1)$ – вектор наблюдения, компонентами которого \dot{Y}_t , $t=0,1$ являются отсчеты комплексной огибающей на выходе фильтра, согласованного с одиночным символом, для четной и нечетной позиций соответственно, нормированные по соглашению делителем $E_s / \sqrt{2}$. Тогда

$$\dot{\mathbf{Y}} = \dot{H}_1 \mathbf{u}_1 + \dot{H}_2 \mathbf{u}_2 + \mathbf{n}, \quad (10.38)$$

где \mathbf{n} – двумерный вектор независимых комплексных отсчетов гауссовского шума с нулевым средним и одинаковой дисперсией. Тогда МП правило (см. главу 2) выдаст \hat{b}_0 и \hat{b}_1 в качестве оценок символов данных b_0 и b_1 , если они минимизируют евклидово (в квадрате) расстояние между наблюдением $\dot{\mathbf{Y}}$ и полезной компонентой $\dot{H}_1 \mathbf{u}_1 + \dot{H}_2 \mathbf{u}_2$:

$$d^2(\dot{H}_1 \mathbf{u}_1 + \dot{H}_2 \mathbf{u}_2, \dot{\mathbf{Y}}) = \|\dot{\mathbf{Y}} - \dot{H}_1 \mathbf{u}_1 - \dot{H}_2 \mathbf{u}_2\|^2 = (\dot{\mathbf{Y}} - \dot{H}_1 \mathbf{u}_1 - \dot{H}_2 \mathbf{u}_2, \dot{\mathbf{Y}} - \dot{H}_1 \mathbf{u}_1 - \dot{H}_2 \mathbf{u}_2).$$

На основании аксиом дистрибутивности и симметрии скалярного произведения $((\mathbf{u}, \mathbf{v}) = (\mathbf{v}, \mathbf{u})^*)$ и с учетом ортогональности $\mathbf{u}_1, \mathbf{u}_2$ получаем

$$d^2 = \|\dot{\mathbf{Y}}\|^2 - 2\operatorname{Re}[\dot{H}_1^* (\dot{\mathbf{Y}}, \mathbf{u}_1)] - 2\operatorname{Re}[\dot{H}_2^* (\dot{\mathbf{Y}}, \mathbf{u}_2)] + |\dot{H}_1|^2 \|\mathbf{u}_1\|^2 + |\dot{H}_2|^2 \|\mathbf{u}_2\|^2,$$

где d^2 – сокращенное обозначение рассматриваемого квадрата расстояния, или, после подстановки $\mathbf{u}_1, \mathbf{u}_2$ из (10.37)

$$d^2 = \|\dot{\mathbf{Y}}\|^2 - 2\operatorname{Re}[b_0^* (\dot{H}_1^* \dot{Y}_0 + \dot{H}_2^* \dot{Y}_1^*)] - 2\operatorname{Re}[b_1^* (\dot{H}_2^* \dot{Y}_0 - \dot{H}_1^* \dot{Y}_1^*)] + (|\dot{H}_1|^2 + |\dot{H}_2|^2) (|b_0|^2 + |b_1|^2).$$

Преобразованные отсчеты наблюдения

$$\dot{z}_0 = \dot{H}_1^* \dot{Y}_0 + \dot{H}_2^* \dot{Y}_1^*, \quad \dot{z}_1 = \dot{H}_2^* \dot{Y}_0 - \dot{H}_1^* \dot{Y}_1^*, \quad (10.39)$$

также как и норма вектора наблюдения не зависят от переменных b_0, b_1 , относительно которых и минимизируется d^2 . Следовательно, в вышеприведенном соотношении допустима замена $\|\dot{\mathbf{Y}}\|^2$ на $|\dot{z}_0|^2 + |\dot{z}_1|^2$, приводящая к виду

$$d^2 = |\dot{z}_0|^2 - 2\operatorname{Re}(b_0^* \dot{z}_0) + |b_0|^2 + (|\dot{H}_1|^2 + |\dot{H}_2|^2 - 1) |b_0|^2 + |\dot{z}_1|^2 - 2\operatorname{Re}(b_1^* \dot{z}_1) + |b_1|^2 + (|\dot{H}_1|^2 + |\dot{H}_2|^2 - 1) |b_1|^2.$$

Теперь становится очевидным, что минимизация d^2 относительно b_0, b_1 заключается в

раздельной минимизации двух функций одной переменной $d^2(b_0) = |\dot{z}_0 - b_0|^2 + H^2|b_0|^2$ и $d^2(b_1) = |\dot{z}_1 - b_1|^2 + H^2|b_1|^2$, где $H^2 = |\dot{H}_1|^2 + |\dot{H}_2|^2 - 1$, относительно b_0 и b_1 . Таким образом, оценки символов данных b_0 , b_1 находятся как

$$\hat{b}_l = \arg \max_{b_l} \left(|\dot{z}_l - b_l|^2 + H^2|b_l|^2 \right), \quad l = 0, 1, \quad (10.40)$$

где \dot{z}_l определено (10.39), а минимизация осуществляется по всем значениям b_l , определяемым заданным алфавитом символа данных. В частном случае PSK модуляции данных $|b_l|^2 = 1$ и единственным компонентом первого члена соотношения (10.40), зависящим от b_l , является $\text{Re}(b_l^* \dot{z}_l)$, которое превращает решающее правило в обычную PSK демодуляцию (см. пункт 7.1.2), однако основанную на модифицированных статистиках согласованного фильтра \dot{z}_l :

$$\hat{b}_l = \arg \max_{b_l} [\text{Re}(b_l^* \dot{z}_l)], \quad l = 0, 1.$$

Зафиксируем передаваемые символы $b_l, l = 0, 1$ и коэффициенты замираний $\dot{H}_i, i = 1, 2$. Тогда полезные компоненты решающих статистик $\dot{z}_l, l = 0, 1$ могут быть оценены усреднением \dot{Y}_0 и \dot{Y}_1 в (10.39) относительно аддитивного шума. Обозначив эту операцию через $E_n\{\cdot\}$, получаем (см. (10.38)) $E_n\{\dot{Y}_0\} = \dot{H}_1 b_0 + \dot{H}_2 b_1$, $E_n\{\dot{Y}_1\} = -\dot{H}_1 b_1^* + \dot{H}_2 b_0^*$, так что

$$E_n\{\dot{z}_0\} = (A_1^2 + A_2^2)b_0, \quad E_n\{\dot{z}_1\} = (A_1^2 + A_2^2)b_1,$$

где, как и прежде, $A_i = |\dot{H}_i|, i = 1, 2$. Аналогичным образом, дисперсия σ_z^2 вещественной или мнимой частей аддитивного шума, входящего в \dot{z}_l , составит величину $\sigma_z^2 = (A_1^2 + A_2^2)\sigma^2$, где σ^2 – дисперсия реальной части шумового отсчета в (10.38). Тогда SNR по мощности $q_{z_l}^2$ для каждой из статистик $\dot{z}_l, l = 0, 1$ будет

$$q_{z_l}^2 = \frac{|E_n\{\dot{z}_l\}|^2}{\sigma_z^2} = \frac{(A_1^2 + A_2^2)|b_l|^2}{\sigma^2}.$$

Теперь учтем случайность значений $A_i, i = 1, 2, b_l, l = 0, 1$ и усредним $q_{z_l}^2$ по всем случайным факторам с целью получения среднего значения SNR $\overline{q_{z_l}^2}$ для каждой из статистик $\dot{z}_l, l = 0, 1$. При естественной нормировке коэффициентов замираний и модуляционного алфавита $\overline{A_i^2} = 1, i = 1, 2, \overline{|b_l|^2} = 1, l = 0, 1$ величина $\overline{q_{z_l}^2} = 2/\sigma^2$, а поскольку \mathbf{n} в (10.38) нормирован к $E_s/\sqrt{2}$ и дисперсия шума на выходе фильтра, согласованного с символом, составляет $N_0 E_s/2$ (см. (2.15)), то $\sigma^2 = N_0/E_s$. Тогда окончательное соотношение примет вид $\overline{q_{z_l}^2} = 2E_s/N_0 = q_s^2$, показывающее, что схема Аламоути, точно также как и код с переключением во времени, сохраняет тоже среднее значение SNR на символ, как и схема без разнесения, обеспечивая выигрыш от разнесения $n_d = n_T = 2$. Еще раз подчеркнем,

что достоинство кода Аламоути относительно конструкции с переключением во времени заключается в отсутствии пауз в излучении, что влечет улучшение в величине пик-фактора и более высокую спектральную эффективность.

Код Аламоути является полноскоростным кодом, означающим, что два независимых модуляционных символа передаются в течение двухсимвольного интервала. В общем случае пространственно-временной код длины n , который позволяет передачу k независимых модуляционных символов данных, обладает скоростью $R = k/n$. Полноскоростные коды являются предпочтительными, поскольку они не требуют дополнительной полосы при передаче единственной антенной. Существование большинства полноскоростных кодов, обеспечивающих максимальный выигрыш при разнесении $n_d = n_T$, жестко зависит от модуляционного алфавита. Для вещественного модуляционного алфавита (например, BPSK) полноскоростные коды существуют для нескольких значений числа n_T передающих антенн, тогда как для комплексных модуляционных символов (QPSK, QAM и др.) код Аламоути является единственным¹ [112]. В тоже время несколько интересных конструкций пространственно-временных блочных кодов становятся достижимыми как для комплексного, так и вещественного алфавитов в случае снятия полноскоростного ограничения [110, 112] (см. также задачи 10.16, 10.17). Для получения большего количества информации по данной проблеме и ознакомления с другими аспектами пространственно-временного кодирования отсылаем заинтересованного читателя к работам [110–113] и статьям, упомянутым в [110].

10.3.5. Разнесение на передаче в широкополосных приложениях.

Широкополосная концепция, по-видимому, предоставляет непосредственный и простой путь реализации разнесения при передаче. Действительно, основным узким местом процедуры разнесения при передаче является проблема разделения на приемной стороне сигналов, излученных одновременно различными передающими антеннами. Можно обойти этот камень преткновения посредством использования широкополосных сигналов различных антенн, полученных с помощью различных (ортогональных) широкополосных кодов. На первый взгляд это очевидное средство организации разнесения при передаче не оказывается, однако, столь универсальным. На самом деле ортогональные последовательности в CDMA системах представляют собой дефицитный ресурс, поскольку их число определяет потенциальное число абонентов. Таким образом, в насыщенных ($K = N$), а тем более в перенасыщенных ($K > N$) прямых CDMA каналах отсутствуют свободные ортогональные последовательности для организации разнесения на передаче, что делает пространственно-временные коды с экономным использованием полосы настолько же ценными и в CDMA приложениях.

Способ включения кода Аламоути в прямой DS CDMA канал не представляет особых сложностей и не требует дополнительного привлечения сигнатурного ресурса. Пусть $\dot{S}_k(t)$ – комплексная огибающая k -й пользовательской сигнатуры (трактуемой как сигнал такой же длительности T_p , что и символ данных), а $b_{k,0}, b_{k,1}$ – четный и нечетный символы данных модуляции, передаваемые k -м пользователем. Тогда достаточно только применить в массиве (10.37) DS расширенные символы $b_{k,0}\dot{S}_k(t)$ и $b_{k,1}\dot{S}_k(t)$ вместо b_0, b_1 соответственно для организации разнесения на приеме на основе фиксированной сигнатуры $\dot{S}_k(t)$. Тогда первая и вторая антенны передают сигналы $b_{k,0}\dot{S}_k(t) - b_{k,1}^*\dot{S}_k^*(t - T_p)$ и

¹ Не рассматриваем в качестве различных вариантов тривиальные модификации (10.37), сохраняющие ортогональность строк и нормировку строк, такие как общее сопряжение или/и умножение строк на -1, а также на любое фиксированное комплексное число с единичной амплитудой.

$b_{k,1}\dot{S}_k(t) + b_{k,0}^*\dot{S}_k^*(t - T_p)$ соответственно в течение двух последовательных символьных интервалов. Сужение этих сигналов с помощью опорного сигнала $\dot{S}_k^*(t) + \dot{S}_k(t - T_p)$ удаляет сигнатуру и превращает данную задачу в одну из рассмотренных в предыдущем параграфе. Этот метод с незначительными изменениями используется в прямом канале UMTS для организации не замкнутой петли (без обратной связи абонент–БС) разнесения при передаче [114,115]. Для справедливости следует отметить, что обработка в пользовательском терминале сигналов синхронизации (измерение временной задержки), кодированных кодом Аламоути, окажется значительно сложнее в сравнении с демодуляцией данных. По этой причине метод разнесения при передаче, используемый в канале синхронизации, основывается на кодировании с переключением во времени, что затрагивалось в пункте 10.3.3 [115]. Еще одним интересным примером служит использование разнесения с замкнутой петлей в выделенных (т.е. назначенных конкретному пользователю) физических каналах. Основываясь на данных обратной связи МС–БС, базовая станция знает текущее состояние канала, связывающего БС с конкретным пользователем, и управляет фазами сигналов двух передающих антенн таким образом, чтобы они складывались когерентно на входе мобильного терминала. Возможна также и регулировка амплитуд передаваемых сигналов для правильной реализации комбинирования с максимальным отношением в приемной антенне и приведения эффективности разнесения при передаче как можно ближе к той, что имеет место при одиночном приеме. Спецификация прямого канала cdma2000 включает некоторые аналогичные решения, касающиеся разнесения при передаче.

Задачи.

10.1. Пусть первая сигнатура \mathbf{a}_1 является линейной комбинацией остальных сигнатур. Показать, что линейная система (10.12) не имеет решения, т.е. подавление помех MAI автоматически устраняет также и полезный эффект.

10.2. Рассматривается DS CDMA система для трех пользователей. Бинарные сигнатуры, обладающие коэффициентом расширения $N=3$, имеют вид: $\mathbf{a}_1 = (+ - -)$, $\mathbf{a}_2 = (- + -)$, $\mathbf{a}_3 = (- - +)$. Найти опорный сигнал первого пользовательского декорреляционного приемника, продемонстрировать подавление помех MAI и оценить потери в отношении SNR декорреляционного алгоритма по отношению к согласованной фильтрации.

10.3. Определить производную функции (10.19) и показать, что вектор \mathbf{u} , определяемый (10.20), является точкой минимума этой функции.

10.4. Доказать лемму об обратимости матриц в форме (10.21).

10.5. Найти опорный вектор МСКО обнаружителя для условий задачи 10.2, полагая все амплитуды сигнатур и дисперсию шума равными единице. Вычислить отношение SINR на выходе МСКО обнаружителя, сравнить этот показатель с аналогичным для декорреляционного обнаружителя и согласованного фильтра (задача 10.2) и объяснить полученные результаты.

10.6. Синхронная CDMA система обслуживает 128 пользователей с коэффициентом расширения в пределах $N=96$. Сигнатурами являются столбцы матрицы Адамара порядка 128, в которой вычеркнуты 32 строки. Найти опорные вектора всех пользователей, обеспечивающие МСКО, если сигналы пользователей имеют одинаковую интенсивность.

10.7. Прямой канал MC-DS-CDMA системы реализован с использованием трех поднесущих. Данные передаются на каждой поднесущей бинарной ФМ со скоростью 32 кбит/сек и с коэффициентом расширения 64. Защитный частотный интервал составляет $0.5/\Delta'$, где Δ' – длительность чипа сигнатуры. Как изменится потенциальное число пользователей при замене MC-DS-CDMA на DS-CDMA?

10.8. При использовании QPSK данные передаются со скоростью 2.88 Мбит/сек по каналу с полосой когерентности $B_c = 50$ кГц. Найти минимальное число поднесущих, необходимое для организации MC передачи. Какова минимальная длина ДПФ в OFDM схеме, если заголовок, обусловленный защитными интервалами не должен превосходить десяти процентов?

10.9. По прямому каналу MC-CDMA системы в версии OFDM данные с использованием QPSK передаются со скоростью 40 кбит/сек по каналу с максимальной задержкой распространения $\tau_{\max} = 10$ мксек. Сколько пользователей он может обслужить, если все не искажаемые сигнатуры являются ортогональными, а общая полоса составляет 5 МГц?

10.10. Рационально ли использовать комбинирование с нуль-форсированием в MC-CDMA системе, работающей в релеевских подканалах?

10.11. Предположим, что n_R антенн параллельно принимают сигнал, переданный

единственной антенной, причем интенсивности всех принятых сигналов одинаковы, также как и независимого гауссовского шума, искажающего сигналы. Насколько шенноновская пропускная способность такого канала отличается от пропускной способности канала, отвечающего варианту без разнесения, если приемнику известны различия длин путей всех сигналов?

10.12. Предположим, что передатчик способен передавать данные с привлечением n_d независимых идентичных ветвей разнесения при фиксированной полной мощности излучения. Предположим, что интенсивности всех принятых сигналов одинаковы, также как и интенсивность независимого гауссовского шума, искажающего сигналы, и что приемнику (не передатчику) известны различия длин путей всех сигналов. С точки зрения пропускной способности Шеннона: лучше передавать один и тот же или различные потоки данных по n_d ветвям?

10.13. Доказать верхнюю границу дополнительной функции ошибок: $Q(x) \leq (1/2)\exp(-x^2/2)$ для любого $x \geq 0$.

10.14. Доказать сходимость правой части (10.36) к верхней границе вероятности ошибки для единственной ветви без замираний: $(1/2)\exp(-q^2/2)$, когда число ветвей разнесения неограниченно возрастает.

10.15. Согласно строгому определению [110,112,113], выигрыш от разнесения равен минимальному рангу среди всех разностей пар различных пространственно-временных кодовых слов, т.е. $n_T \times n$ матриц $[u_i^t]$. Доказать, что для кода Аламоути выигрыш от разнесения равен двум.

10.16. Определить скорость (в числе символов данных на один кодовый символ) и выигрыш от разнесения пространственно-временного кода с вещественными символами кодовых слов [112]

$$\mathbf{u} = \begin{bmatrix} b_0 & -b_1 & -b_2 & -b_3 \\ b_1 & b_0 & b_3 & -b_2 \\ b_2 & -b_3 & b_0 & b_1 \end{bmatrix}.$$

10.17. Определить скорость (в числе символов данных на один кодовый символ) и выигрыш от разнесения пространственно-временного кода с комплексными символами кодовых слов [116]

$$\mathbf{u} = \begin{bmatrix} b_0 & b_1^* & b_2^* & 0 \\ -b_1 & b_0^* & 0 & -b_2^* \\ -b_2 & 0 & b_0^* & b_1^* \end{bmatrix}$$

11. Примеры действующих беспроводных широкополосных систем.

11.1. Предварительные замечания.

Как уже неоднократно отмечалось в предыдущих главах, широкополосность является основной философией многочисленных современных и проектируемых беспроводных систем, охватывающих такие области как локацию и навигацию, так и сети мобильного радио и локальной области. Коммерческие достоинства широкополосной технологии отчетливо проявились в рыночном успехе системы мобильного телефона второго поколения cdmaOne (IS-95), а также консолидацией международного телекоммуникационного сообщества в вопросе принятия технологии CDMA в качестве основы систем третьего и последующих поколений. Из всего множества примеров, иллюстрирующих практическую применимость принципа широкополосности и CDMA, автор выбрал только три из них, которые подробно рассматриваются в данной главе, учитывая их выдающееся значение, как для сегодняшнего дня, так и на будущее, и высокую образовательную поучительность. Рассматриваемые системы уже неоднократно упоминались в контексте примеров практической реализации тех или иных частных принципов и идей.

11.2. Система глобального позиционирования.

Совершенно естественным является стремление движущегося по земной поверхности или в пространстве объекта знать свое текущее положение. Задача позиционирования является основной целью навигации. Современное навигационное оборудование необходимо для информирования потребителя о его мгновенных координатах, скорости передвижения относительно всех координатных осей, точном текущем времени, предполагаемом положении в тот или иной момент времени и т.п. Наиболее передовой и универсальной системой, обеспечивающей решение указанных задач, является Система Глобального Позиционирования (Global Positioning System (GPS)), работоспособное состояние которой официально провозглашено США в 1993 г.

11.2.1. Основные принципы системы и архитектура.

Тремя координатами можно описать положение любого объекта на или над земной поверхностью, например широтой, долготой и высотой. Для их определения объект может измерить расстояние (дальность) до трех фиксированных точек (маяков), координаты которых известны заранее с высокой точностью. Эти измерения приводят к трем уравнениям, тремя неизвестными которых являются координаты объекта. Разрешив эту систему, объект определяет свое собственное положение.

В радионавигации измерение дальности осуществляется посредством измерения задержки на распространение сигналов, излученных маяками: расстояние есть результат произведения задержки на распространение и скорости света. Как правило, все передатчики маяков жестко синхронизированы друг с другом, работая в едином *системном времени*. В отличие от этого генератор любого индивидуального пользователя обычно имеет смещение относительно системного времени и вследствие частотного дрейфа и нестабильности временной сдвиг опорного генератора пользователя относительно системного времени следует учитывать в качестве четвертого неизвестного. Фиксация времени прихода сигнала маяка относительно своей местной временной шкалы не позволяет потребителю определить истинное значение задержки на распространение, поскольку там содержится ранее упомянутое временное расхождение. Простейшим путем исключения этого неизвестного вклада может служить измерение расстояния с помощью запросного метода, при котором маяк излучает

свой сигнал в ответ на пользовательский дальномерный сигнал. Последнее, однако, предполагает активизацию пользователей, т.е. излучение некоторой энергии. Необходимость маяку различать сигналы отдельных пользователей и отвечать каждому из них индивидуально жестко лимитирует число обслуживаемых абонентов. Возможность обслуживания неограниченного числа пользователей существует лишь при однопроходном измерении дальности, т.е. при пассивном режиме работы абонентов, работающих только на прием, что очень часто является желательным, учитывая требования безопасности. В этом случае приемник пользователя сможет преодолеть проблему временного смещения системной и пользовательской шкал, измеряя время прихода сигналов от четырех, а не трех маяков. Каждое измеренное время прихода содержит «истинную» задержку на распространение и одно и то же неизвестное временное временное расхождение шкал. Умноженные на скорость распространения света эти измерения дадут *псевдодальности*: истинное расстояние и результат произведения смещения шкал и скорости света. Таким образом, в распоряжении пользователя имеются четыре уравнения с четырьмя неизвестными, решение которых обеспечивает пользователю определение своего местоположения, а также оценку временного смещения шкал. Такого рода обработка позволяет пользователю осуществить одновременные измерения времени и положения.

Система GPS является системой космического базирования, означающей, что спутники, являющиеся носителями передатчиков, используются в качестве маяков. Расположение передатчиков на борту спутников открывает возможность к использованию СВЧ колебаний для передачи навигационных сигналов. По сравнению с длинноволновыми характеристиками более ранних систем наземного базирования (Лоран, Омега и др.) СВЧ волны распространяются только по прямой, без дифракции, так что попытки использовать их в наземных передатчиках приведут к возможности позиционирования только в пределах зоны прямой видимости маяка. В то же время СВЧ диапазон является более благоприятным в сравнении с длинноволновым в том, что касается массо-габаритных характеристик передающих и приемных антенн, а также многих других частей оборудования. Теперь становится очевидным способ достижения согласия между желанием использовать СВЧ диапазон и добиться большой зоны покрытия навигационным передатчиком: расположить СВЧ-передатчик на борту космического аппарата. Если возвышение спутника над земной поверхностью будет достаточно большим, передатчик охватит большой участок земной поверхности, т.е. удовлетворит требованиям к размеру навигационной зоне покрытия.

Структура системы GPS включает три основных компонента. *Космический сегмент* состоит из 24 основных спутников (плюс несколько резервных), расположенных на 6 около круговых орбитах по 4 на каждой, с реальным периодом обращения около 12 часов. Орбиты сдвинуты друг относительно друга на 60° по долготе и имеют наклон в 55° относительно плоскости экватора. Подобное пространственное расположение позволяет одновременно наблюдать не менее 4-х спутников в любой момент времени и любой точке земной поверхности при угле возвышения 10° . Однако в большинстве случаев число видимых спутников больше, достигая в некоторые моменты времени 10. Чем больше число одновременно наблюдаемых космических носителей, тем выше точность местоопределения и времени. На борту каждого спутника имеется атомный (цезиевый или рубидиевый) стандарт частоты, характеризующийся чрезвычайно высокой стабильностью: за сутки уход частоты от номинала не превышает 10^{-13} . Учитывая незначительное влияние погодных условий на распространение волн в диапазоне системы GPS, ее космический сегмент обеспечивает точное местоопределение в любой точке земной поверхности вне зависимости от времени и погодных условий.

Как уже указывалось ранее, потребитель способен вычислить свое положение на основании измеренного расстояния при условии знания координат маяков. Поскольку при передвижении по своим орбитам спутники непрерывно вращаются, то их координаты постоянно изменяются, и их мгновенные значения необходимы потребителю в любой момент времени, в которой осуществляется его местоопределение. Мгновенное положение спутника на

орбите не является полностью детерминистическим вследствие случайных воздействий той или иной природы. Аналогично, спутниковые стандарты частоты, хотя и очень стабильные, раньше или позже накапливают относительную временную расстройку, ухудшающую точность местоопределения. Таким образом, необходимо отслеживать положение спутников и поведение спутниковых генераторов частоты и сообщать об этом потребителю. Для реализации указанных потребностей в состав системы GPS входит *сегмент управления (control segment)*, состоящий из основной станции управления в Колорадо Спрингс (штат Колорадо) и пяти станций слежения, расположенных в Колорадо Спрингс, Гавайях, Кваджалейне, Диего Гарсия и острове Вознесения. Станции слежения осуществляют непрерывный контроль за всеми спутниками, измеряя их частотные параметры, тестируя состояние спутника и т.п. и посредством специальной линии связи передают данные на главную станцию управления. Последняя вычисляет текущее и предсказанное положение спутников, сдвиг шкал времени и другие важные параметры и передает обновленные данные одной из трех наземных станций управления, расположенных в тех же местах, что и станции слежения (Кваджалейн, Диего Гарсия и остров Вознесения). Наземные станции управления загружают данные на спутники GPS системы, используя выделенный S-канал связи.

Пользовательский сегмент охватывает всех пользователей, снабженных GPS-приемниками. В настоящее время существует множество (не менее сотен) различных моделей приемников, от самых простых и дешевых наручных приборов (используемых в спортивных целях и для отдыха) до очень сложных и дорогих, предназначенных для военных целей, топографии и т.п.

11.2.2. Дальномерные сигналы системы GPS.

Основной атомный стандарт частоты каждого GPS спутника вырабатывает частоту 10.23 МГц. Когерентное частотное умножение на 154 и 120 образование двух частот в L-диапазоне с номиналами $L1=1575.42$ МГц и $L2=1227.60$ МГц, которые в дальнейшем модулируются GPS дальномерными кодами. Используются два дальномерных кода: C/A-код (*код грубого поиска (coarse acquisition)* или *легкого доступа (clear access)*) и P-код (*точный (precision)* или *защищенный (protected)* код). Каждому спутнику назначается своя специфическая пара C/A- и P-кодовых последовательностей. Более высокая несущая частота L1 является основной и ее сигнал доступен любому GPS приемнику, тогда как более низкая частота L2 предназначена к использованию только авторизированным пользователями для высокоточных измерений.

Объяснение тому, почему наличие двух несущих способствует точному местоопределению лежит в характере распространения волн в ионосфере: верхнем атмосферном слое, лежащем в 50 км от поверхности земли. Обладая высокой концентрацией свободных ионов и электронов, обусловленной ионизирующим эффектом солнечного ультрафиолетового и рентгеновского излучения, ионосфера представляет собой дисперсную среду, в которой дальномерный сигнал распространяется со скоростью, отличающейся от скорости распространения в свободном пространстве, приобретая некоторую дополнительную задержку. К сожалению, эта дополнительная задержка не может быть достаточно точно предсказана вследствие быстрых случайных флуктуаций параметров среды. С другой стороны, общий характер зависимости ионосферной задержки от частоты хорошо аппроксимируется обратной квадратичной зависимостью с неизвестным коэффициентом пропорциональности. Последний может быть определен только как решение уравнения, полученного в результате измерения задержек идентичных сигналов на двух особым образом разнесенных частотах и вычитания результатов. Завершая рассмотрение, отметим, что ионосферная поправка может быть определена и, значит, удалена из измеряемой псевдодальности.

Последовательность C/A-кода служит для организации шкалы позиционирования умеренной точности. Она представляет собой последовательность Голда (см. параграф 7.5.2) длины $N = 1023$ с длительностью чипа чуть меньше 1 мкс, что приводит к периоду сигнала в

точности равному 1 мс. Очевидно, что число (1025) существующих последовательностей Голда длины $N=1023$ во много раз превосходит необходимое для спутников GPS их количество, так что используются только некоторые из них. Периодически повторяемые последовательности Голда применяются только для модуляции частоты L1 и постоянно передаются спутниками. Несмотря на жесткую синхронизацию всех спутников расстояние между ними и потребителем изменяется в ходе эволюций спутников, так что сигналы различных спутников достигают приемной антенны потребителя со взаимными временными сдвигами, изменяющимися в значительном диапазоне. Поскольку каждый спутник идентифицируется своей специфической последовательностью Голда, и все данные сигналы обладают низким уровнем асинхронной взаимной корреляцией, то пользовательский приемник способен выделить любой индивидуальный сигнал спутника от других. Другими словами, шкала на основе C/A-кода космического сегмента GPS использует ансамбль последовательностей Голда для организации асинхронного варианта CDMA. Передача спутниковых координат (эфемерид) наряду с другими важными данными (смещением шкалы времени системы GPS, поправками на эффекты распространения, состоянием спутника и др.) организована методом DS расширения (см. параграф 7.1): поток бит данных со скоростью 50 бит/сек непосредственно бинарно манипулирует последовательность Голда перед модулированием несущей L1. Таким образом, любой бит данных занимает интервал, равный 10 периодам последовательности Голда, что теоретически нейтрализует ухудшение их корреляционных свойств вследствие манипуляции данными (см. параграф 7.3). Минимальная посылка данных (кадр), необходимая для местоопределения, составляет 1500 бит (30 сек), структурированная в 5 подкадров, содержащих по 300 бит каждый. Первые три подкадра повторяются без изменения (за исключением случая загрузки новых данных) в каждом кадре, тогда как содержимое остальных меняется в течение суперкадра, охватывающего 25 кадров. Для более надежной передачи поток данных кодируется (32,26) расширенным кодом Хэмминга. Важной разновидностью данных, передаваемых в 4-м и 5-м подкадрах, служат альманах: грубые координаты остальных спутников системы. Приняв сигнал любого спутника, потребитель получает доступ к грубому положению других и может использовать эту информацию для ускорения процедуры поиска сигналов других спутников (см. параграф 8.3).

C/A-код обеспечивает стандартную точность местоопределения и бесплатно доступен любому пользователю, обладающему GPS приемником. P-код предназначен для позиционирования с высокой точностью и поэтому имеет длительность чипа в 10 раз меньшую, чем C/A-код (меньше 100 нсек), или в 10 раз более широкую полосу (см. пункт 2.12.2 для напоминания связи между шириной полосы и точностью измерения дальности). Для реализации описанного выше двухчастотного метода компенсации ионосферной ошибки P-код (DS модулированный потоком бит данных аналогично C/A-коду) передается на обеих частотах L1 и L2, причем на частоте L1 C/A- и P-сигналы находятся в квадратуре с превышением на 3 дБ интенсивности первого из них. В свою очередь интенсивность сигнала на частоте L2 на 3 дБ ниже, чем на L1. Структура P-кода описана в открытых документах по системе GPS. Он формируется путем посимвольного сложения по модулю 2 двух очень длинных бинарных последовательностей, отличающихся по длине на 37 чипов. Результирующий период последовательности, образованной подобным способом, составляет порядка 266 дней. Неперекрывающиеся 7-дневные ($6.187104 \cdot 10^{12}$ чипов) сегменты этой последовательности используются в качестве P-кодов различных спутников. Министерство обороны США рекомендовало проектировщикам GPS позаботиться о жестком ограничении доступа к P-коду, считая, что несанкционированное его использование может угрожать национальной безопасности. Шифрование P-кода реализовано его суммированием по модулю 2 с маской или ключевым W-кодом, структура которого закрыта. Результирующий Y-код обладает совершенной секретностью (см. пример 3.3.1).

11.2.3. Обработка сигналов.

Основные операции, осуществляемые одночастотным (L1) приемником GPS, присущи любой обычной широкополосной системе с прямым расширением спектра. После грубого поиска C/A–сигнала (см. параграф 8.2-8.3), облегченного возможным априорным знанием положения спутника, петля слежения по задержке (см. параграф 8.4) замыкается, начиная выдавать последовательность оценок псевдодальности до спутника. Как правило, современные GPS приемники содержат ряд каналов, параллельно обрабатывающих C/A сигналы всех видимых спутников. По завершению поиска последнего отслеживаемого спутника приемник готов к выдаче координат потребителя, что является установившимся процессом, длящимся неограниченно долго в соответствии с пожеланием пользователя.

Санкционированный пользователь совершает еще аналогичные операции в отношении P–кода на обеих несущих, затрачивая незначительное время на поиск этих сигналов, поскольку кадр данных, поступивший от сигнала на частоте L1, содержит специальное передаточное слово, которое способствует установке местного генератора P–кода в соответствующее начальное состояние.

Во многих современных GPS приемниках эти основные операции дополняются или заменяются рядом других, направленных на достижение большей точности, ускорение начальной фиксации времени, повышение удобств пользования и пр. Например, дополнительная точность может быть достигнута за счет измерения псевдодальностей путем интегрирования несущей частоты принятого сигнала. Мгновенное значение частоты пропорционально радиальной скорости спутника относительно пользователя. Следовательно, интеграл от частоты за некоторый интервал времени пропорционален изменению расстояния между спутником и потребителем за тот же период времени. Начиная с точки с точно известными координатами, приемник может определять свое дальнейшее местоположение путем интегрирования мгновенной частоты сигналов видимых спутников, т.е. путем приращения расстояния. Более того, существуют методы разрешения неопределенности, позволяющие производить местоопределение путем интегрирования частоты даже без точного знания начальной точки [117-118].

Другим очень популярным методом служит т.н. *дифференциальное* или относительное местоопределение, основная идея которого заключается в следующем. Пусть один из GPS приемников установлен в опорном месте с точным знанием его координат. Тогда сравнение вычисленных расстояний до спутников с измеренными позволяет приемнику в опорной точке определить систематические ошибки (смещения), вызванные системными несовершенствами. Пусть теперь имеется другой приемник, расположенный в точке с неизвестными координатами, удаленной от опорной. Если базовая линия, т.е. расстояние между реперным и удаленным приемником, не очень велика (порядка десятков км), систематические ошибки в опорном и удаленном местах сильно коррелированы, так что удаленный приемник может вычесть смещение, оцененное опорным приемником, из измеренной псевдодальности, повышая тем самым точность своего местоопределения. Очевидно, что подобная модификация системы, должна содержать линию связи, обеспечивающую доставку данных опорного приемника удаленным пользователям. В настоящее время по всему миру организовано значительное число опорных точек с использованием передачи дифференциальных поправок ФМ станциями, ретрансляционными спутниками, радиомаяками, сотовыми системами, Интернетом и пр. [117-118].

11.2.4. Точность местоопределения.

Первоначально планируемая точность местоопределения по C/A–коду системы GPS составляла около 100 м в горизонтальном и 156 м в вертикальном направлениях с вероятностью удержания ошибки местоопределения в этих пределах, равной 95%. Аналогичные цифры для P–кода составляли 16 и 23 м. Однако, многочисленные усовершенствованные струк-

туры приемников, разработанные производителями, показали значительно лучшую точность, даже без привлечения P-кода. Данное положение стало предметом беспокойства институтов США, ответственных за национальную безопасность, и в 1990 году был введен режим избирательной доступности, загромождающий данные эфемерид и времени, передаваемые со спутника, и значительно ухудшающий точность местоопределения. Однако в течение следующей декады широкую популярность приобрела дифференциальная навигация, которая почти полностью позволила исключить подобного рода ошибки, так что режим избирательной доступности оказался практически бессмысленным и был отменен в 2000 году. В настоящее время для рынка GPS оборудования характерен широкий спектр предложений с гарантированной точностью от десятков метров до нескольких миллиметров и лучше.

11.2.5. Система ГЛОНАСС и ГНСС.

Российская навигационная система спутникового базирования ГЛОНАСС имеет много общих черт с системой GPS. Ее космический сегмент состоит из 24 спутников, расположенных на 3-х околокруговых орбитах с номинальным периодом обращения 11 часов 15 минут и наклоном в 64.8° к плоскости экватора. Для обеспечения ионосферной коррекции снова используются две частоты L1 и L2 (соответственно в полосе 1.5 и 1.2 ГГц), причем C/A-код передается на несущей L1, а P-код – на обеих несущих. Текущие эфемериды и другие важные данные кодируются (32,26) кодом Хэмминга, после соответствующего структурирования в кадры и подкадры накладываются на дальномерные коды методом прямого расширения спектра и передаются спутниками со скоростью 50 бит/сек. Сегмент управления обеспечивает непрерывное слежение за спутниками, вычисление/предсказание параметров их орбит и загрузку их в память, находящуюся на спутниках.

Существенное отличие системы ГЛОНАСС от GPS состоит в том, что все спутники передают одинаковый C/A-код, являющийся бинарной m -последовательностью длиной $N=511$ с реальным периодом во времени в 1 мсек. Разделение индивидуальных сигналов спутников возможно благодаря малому взаимному сдвигу несущих частот, превращающему общий C/A-код в ансамбль частотно-сдвинутых копий m -последовательности, описанный в параграфе 7.5.1. Для экономии полосы противоположные спутники на одной и той же орбите (никогда не наблюдаемые одновременно потребителем) используют одинаковое смещение частоты.

Точность в системе ГЛОНАСС имеет тот же уровень, что и в системе GPS. Обе эти системы рассматриваются в настоящее время как содействующие друг другу, образующие интегрированную Глобальную Навигационную Спутниковую Систему (ГНСС). Как уже подчеркивалось, увеличение числа обрабатываемых спутниковых сигналов увеличивает точность местоопределения, так что очевидна выгода от совместного использования обоих спутниковых созвездий. Кроме того, не редки следующие сценарии, когда некоторые спутники над горизонтом заслонены (например, крылом самолета), так что общее число доступных дальномерных сигналов только системы GPS или только ГЛОНАСС не достаточно для местоопределения. Тогда опять совместная обработка сигналов GPS и ГЛОНАСС предоставит значительный выигрыш в целостности местоопределения. Большое число моделей приемников, представленных на рынке или планируемых к разработке, пригодны к совместной обработке сигналов обеих систем.

11.2.6. Применение.

Роль глобальных систем навигации спутникового базирования в современном мире и будущем трудно переоценить. Только простое перечисление областей их применения образует достаточно длинный список, включающий традиционную навигацию кораблей, летательных аппаратов и наземных мобильных объектов (легковых автомобилей, грузовиков и др.), транспортные системы, услуги проектирования (например, путепроводов), мониторинг

лесных и природных ресурсов, сельское хозяйство, гражданское строительство, геодезическая съемка, предсказывание землетрясений, аэрокосмическая съемка, морские исследования и многое другое. Будучи не в состоянии углубиться в рассмотрении данного интересного вопроса, адресуем заинтересованного читателя к источникам [117-119] и сопутствующей им библиографии.

11.3. Эфирный интерфейс систем мобильной радиосвязи стандартов cdmaOne (IS-95) и cdma2000.

11.3.1. Вводные замечания.

Первые промежуточные спецификации CDMA сотового телефона стандарта IS-95 (называемого сейчас как cdmaOne) были опубликованы в 1993-95 годах, а рабочая фаза сети IS-95 стартовала в 1996 г. В настоящее время сети этого стандарта охватывают обширные территории, обслуживая десятки миллионов потребителей. Ее впечатляющий коммерческий успех, широко признанное высокое качество обслуживания и открытость к дальнейшей модернизации явились теми решающими факторами, которые обусловили признание CDMA философии в качестве основы следующих поколений систем мобильного радио (3G и др.).

Первоначально стандарт IS-95 был предназначен для постепенной замены (с сохранением совместимости) американского аналогового стандарта AMPS, работающего в диапазоне 800 МГц. Документы по стандарту IS-95 определяют только частотный разнос прямого (869-894 МГц) и обратного (824-849) каналов¹ и не накладывают никаких ограничений на повторное использование частот в соседних сотах или секторах. Стандартная полоса сигнала IS-95 составляет 1.25 МГц, так что в пределах выделенной полосы в 25 МГц оператор обладает значительной свободой в выборе несущих и частотном планировании сети. Все БС, входящие в сеть, жестко синхронизированы с помощью сигналов GPS для работы по единой шкале времени, позволяющей МС легко переключаться с одной БС на другую (процедура эстафетной передачи). Стандарт IS-95 и его эволюционный наследник cdma2000 являются типичными широкополосными системами с прямым расширением спектра, которые явным образом проявляют все преимущества этой технологии. Они также имеют очень высокое познавательное значение, поскольку в прозрачной форме демонстрируют практические пути материализации многих ранее рассмотренных идей. В последующем изложении проанализируем только наиболее общие идеи расширения спектра, формирования каналов, кодирования и модуляции сигналов в эфирных интерфейсах IS-95 и cdma2000. Читателя, заинтересованного в более глубоком ознакомлении с данной проблемой, советуем обратиться к [18,69,83,120,121] и другим источникам.

11.3.2. Коды расширения в стандарте IS-95.

Расширяющие последовательности, используемые в стандарте IS-95, частично упоминались в ранее рассмотренных примерах. Они конструировались для обеспечения кодового разделения физических каналов, различения сигналов различных БС, поступающих на вход приемника МС, и секретности передачи данных. Синхронное кодовое мультиплексирование физических каналов прямой линии связи, обслуживаемое данной БС, реализовано на основе последовательностей Уолша длины $N=64$ (см. параграф 2.7.3). Ортогональность последовательностей Уолша теоретически позволяет разделить соответствующие 64 физических канала без взаимных помех. Длительность чипа последовательностей Уолша приблизительно равна 0.81 мксек, а скорость изменения чипов – 1.2288 Мчип/сек, что приводит к упоминавшейся ранее полосе 1.25 МГц. Очевидно, что число образованных таким способом прямых физических каналов составляет 64, и в соответствии с CDMA принципом они занимают одну и ту же общую полосу без какого-либо частотного или временного смещения. Все базовые станции используют одно и то же множество из 64 функций Уолша, а расширение спектра осуществляется т.н. *короткими кодами*, которые позволяют разделить сигналы различных базовых станций друг от друга

¹ Термин «прямой» и «обратный» каналы являются синонимами канала вниз (от БС к МС) и вверх (от МС к БС), принятыми в спецификациях по стандартам cdmaOne и cdma2000.

на входе приемника МС. Используются две различные бинарные псевдослучайные последовательности короткого кода PN-I и PN-Q, применяемые соответственно в синфазной и квадратурной ветвях модулятора БС. Первоначально они формируются как m -последовательности генераторами на основе линейного регистра сдвига с обратной связью, содержащего 15 ячеек памяти (см. параграф 6.6), и задаются примитивными полиномами $f_I(x) = x^{15} + x^{13} + x^9 + x^8 + x^7 + x^5 + 1$ для PN-I и $f_Q(x) = x^{15} + x^{12} + x^{11} + x^{10} + x^6 + x^5 + x^4 + x^3 + 1$ для PN-Q. Полученные последовательности обладают периодом $L = 2^{15} - 1$, но для получения коротких кодов PN-I и PN-Q они расширяются еще на один нулевой символ, добавляемый после 14 последовательных нулей. В итоге получаем последовательности длиной $N = L + 1 = 2^{15} = 32768$ чипов, и при той же чиповой скорости, что и кодов Уолша, на интервале в 2 секунду укладывается 37.5 периодов коротких кодов или 75 периодов на интервале в 2 секунды. Для распознавания различных базовых станций каждой из них присваиваются специфические копии основных последовательностей коротких кодов, получаемые за счет сдвига во времени. Всего имеется 512 подобных пар копий, каждая последующая сдвинута относительно предыдущей пары на 64 чипа или примерно 52 мксек. При сетевом планировании присвоение пар коротких кодов базовым станциям осуществляется таким образом, чтобы исключить вероятность приема любой МС сигнала случайной БС, имеющего задержку на распространение, как и сигнал полезной БС, и интенсивность, достаточную для их перепутывания. Следует также отметить, что относительные временные сдвиги между базовыми станциями, входящими в сеть, устанавливаются одномоментно и остаются постоянными навсегда, поскольку все БС используют GPS приемники для синхронизации работы своих опорных генераторов.

Еще одним расширяющим кодом является *длинный код*, первоначально генерируемый как бинарная m -последовательность памяти 42. В соответствие со спецификацией примитивный полином 42-разрядного LFSR генератора длинного кода предписывает образование отводов обратной связи от ячеек с номерами 7, 9, 11, 15, 16, 17, 20, 21, 23, 24, 25, 26, 32, 35, 36, 37, 39, 40, 41 и 42 при нумерации их слева направо. Полученная таким образом m -последовательность снова расширяется вставкой одного дополнительного нуля после серии из 41 последовательного нуля, образуя период равный $N = 2^{42}$. При прежней чиповой скорости длинный код обладает периодом в 42 суток. Различные, специфические для каждого пользователя временные сдвиги длинного кода используются в прямом канале для скремблирования потока данных, а в обратном канале – для организации CDMA разделения сигналов МС и одновременно для защиты передаваемых данных (см. ниже).

11.3.3. Каналы прямой линии связи стандарта IS-95.

Существуют 4 типа логических, т.е. отражающих содержание передаваемых данных, прямых каналов:

- пилотный канал (*pilot channel*),
- канал синхронизации (*synchronization channel*),
- каналы вызова (*paging channels*),
- каналы трафика (*traffic channels*).

11.3.3.1. Пилотный канал.

Для конкретной БС пилотный канал является единственным, и по нему в «чистом» виде передается короткий код без модуляции данными. Он может трактоваться как канал, в котором последовательность, состоящая из одних нулей, расширяется с помощью короткого кода. Короткий код, принятый МС, обеспечивает синхронизацию местного генератора с системным (т.е. БС) с точностью, достаточной для формирования приемного ко-

герентного образца, необходимого для снятия расширения и демодуляции принятого сигнала. Поскольку период короткого кода содержит целое число (512) периодов последовательностей Уолша, то последние, сформированные в приемном устройстве, автоматически оказываются согласованными с расширяющими кодами Уолша принятого сигнала после завершения поиска приемником и захвата пилот сигнала схемой слежения за задержкой.

Функция Уолша, являющаяся первой строкой матрицы Адамара размера 64 и состоящая только из единиц, фактически выделяет (канализирует) пилотный канал от остальных. Наиболее удобно генерировать функции Уолша в цифровом виде, т.е. используя двоичную $\{0, 1\}$ логику, а затем заменить нули и единицы на плюс и минус единицы в соответствии с (6.15). На рис. 11.1, *a* изображена упрощенная структура пилотного канала, где W_0 отвечает двоичной $\{0, 1\}$ функции Уолша нулевого номера, состоящей из одних нулей. Поскольку операция расширения коротким кодом осуществляется для всех прямых

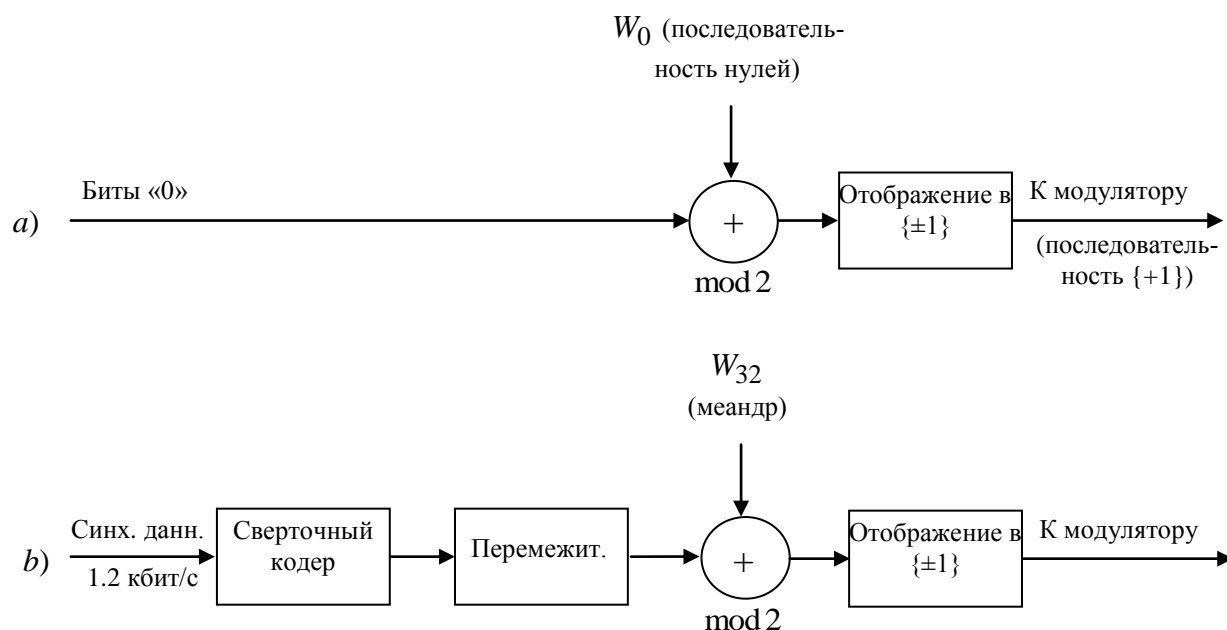


Рис.11.1. Упрощенная структура пилотного (*a*) и синхронизации (*b*) каналов.

каналов одновременно в модуляторе, т.е. после суммирования сигналов всех каналов, то указанная операция будет приведена на рисунке позднее. Проще говоря, пилотный канал, изображенный на рис. 11.1, *a*, подает на модулятор последовательность, состоящую из одних только плюс единиц.

11.3.3.2. Канал синхронизации.

Канал синхронизации также является единственным для любой конкретной БС. В качестве иллюстрации на рис. 11.1, *b* представлена его общая структура. Наряду с некоторой другой информацией по нему передаются данные, позволяющие приемнику получить специфическую для каждого пользователя маску длинного кода и, следовательно, синхронизировать свою копию длинного кода с используемым БС для шифрования и размещения бит контроля мощности (см. ниже). Исходные данные передаются со скоростью 1.2 кбит/с и структурированы в кадры, длительность которых совпадает с периодом короткого кода (26.67 мсек), причем три последовательных кадра объединяются в суперкадр, содержащий 96 бит. Суперкадры организованы в капсулы сообщения, включающие 30 CRC символов (см. пример 9.2.4) для образования индикатора качества сообщения. Сверточный код со скоростью $1/2$ и длиной кодового ограничения 9 (см. параграф 9.3) обеспечивает мощную

защиту данных от воздействия канальных помех. В поток данных не вставляются хвостовые биты, т.е. кодер и декодер не устанавливаются в нулевое состояние в конце каждого кадра. Скорость кодированного потока составляет 2.4 кбит/с, однако каждый символ представляет собой удвоенный более короткий символ (символьное повторение), так что скорость на входе последующего перемежителя составляет 4.8 кбит/с. Перемежитель используется для декорреляции пакетов ошибок, в результате чего улучшается исправляющая способность сверточного кода в отношении длительных уменьшений интенсивности сигнала из-за замираний или других коррелированных искажений сигнала (см. параграф 9.5). Перемежение охватывает один кадр из 128 кодовых символов (26.67 мс) и использует матрицу памяти размерности 16×8 . Кодированный поток записывается в нее по колонкам, а затем считывается согласно стандартно определенному образцу.

Последовательность Уолша, являющаяся 33-й строкой матрицы Адамара размерности 64, фактически канализирует данные синхронизации. Рис. 11.1, *b* снова демонстрирует первично сформированный двоичный $\{0,1\}$ прототип W_{32} , который является меандром, в котором 32 нулевым символа следуют 32 единичных. Иллюстрируемый рисунком метод прямого расширения спектра выполняется обычным образом с учетом взаимно однозначного соответствия между умножением и суммированием по модулю два в рамках вещественного $\{\pm 1\}$ и логического $\{0,1\}$ алфавитов соответственно (см. пункт 7.5.2). Таким образом, символы $\{0,1\}$ кодированного и перемеженного потока посимвольно складываются по модулю два с W_{32} , а результирующий поток затем отображается в алфавит из $\{\pm 1\}$.

11.3.3.3. Каналы вызова.

В стандарте могут быть до 7 каналов вызова, предназначенных для уведомления абонента о вызове, пришедшем из сети, ответа на запрос пользователя об установлении соединения и передачи другой информации, связанной с предоставлением доступа в сеть. Исходные данные поделены на слоты длительностью 80 мс, состоящие из четырех кадров по 20 мс. Кадры или их части образуют капсулы сообщений, каждая из которых содержит до 1184 бит, включая 30 CRC символов. Возможны две скорости передачи данных вызова: 9.6 или 4.8 кбит/с. На рис 11.2 представлена упрощенная структура канала вызова. Первым блоком является сверточный кодер со скоростью $1/2$ и длиной кодового ограниче-

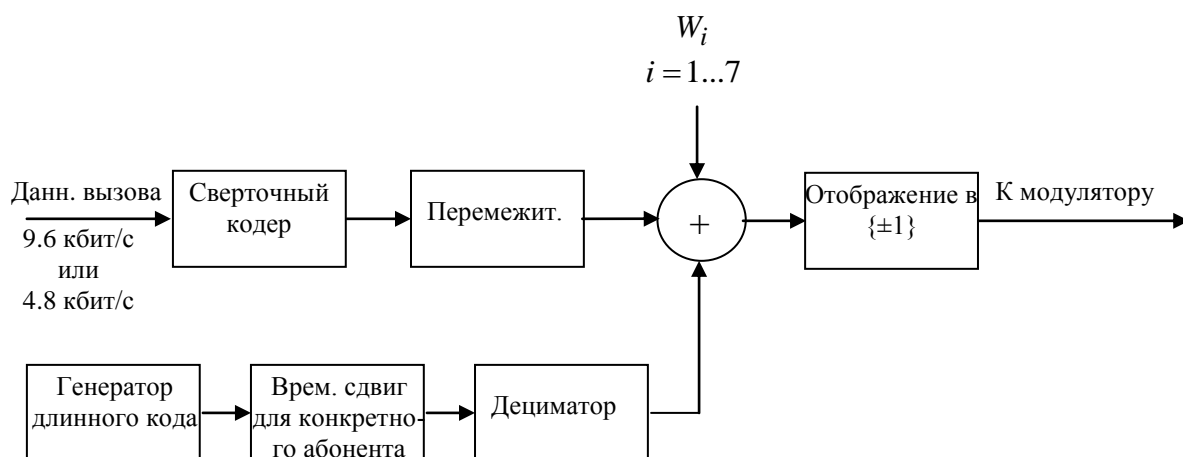


Рис.11.2. Упрощенная структура канала вызова.

ния 9. Аналогично каналу синхронизации в поток не вводятся хвостовые биты, так что отсутствует установка кодера и декодера в нулевое состояние в конце кадра. При входной скорости в 4.8 кбит/с скорость кодированного потока составляет 9.6 кбит/с, однако каж-

дый символ трактуется как удвоение более короткого (символьное повторение), так что независимо от исходной скорости данных кодированный поток обладает скоростью 19.2 кбит/с. Перемежитель оперирует с 20 мс кадрами, используя матрицу памяти размерности 24×16 , в которую кодированные символы записываются колонками. Порядок считывания из памяти устанавливается стандартом.

Следующей операцией, которой подвергается кодированный и перемеженный поток, является шифрование, реализуемое в форме скремблирования, т.е. его посимвольное суммирование по модулю два с псевдослучайной двоичной последовательностью. Последняя образуется как результат децимирования с индексом децимации 64 специфической для каждого пользователя сдвинутой копии длинного кода, т.е. выборе каждого 64-го символа. Таким образом, скорость 1.2288 Мчип/с длинного кода делится на 64, превращаясь в $1.2288/64=19.2$ Кчип/с. Как видно, шифрование является индивидуальным для любого пользователя, предотвращая несанкционированный перехват при мониторинге данных доступа, посылаемых БС.

Функции Уолша с номерами от 1 до 7 (со второй по восьмую строки матрицы Адамара) используются в качестве образующих каналов вызова. Первичный канал вызова, образованный функцией Уолша W_1 , всегда доступен, тогда как остальные могут оказаться либо не активизированными, либо используемыми как каналы трафика.

11.3.3.4. Каналы трафика.

Каналы, рассмотренные до настоящего момента, выполняют служебные функции, необходимые для начала и поддержания передачи основного сообщения пользователю. Каналы трафика ответственны за доставку основной информации: оцифрованной речи, компьютерных или мультимедийных данных и т.п. Для конкретности ограничимся передачей речи. Тогда предварительной операцией является кодирование речи, осуществляемое блоком, называемым *вокодером* (кодером речи). Оставляя в стороне обсуждение о достаточно сложных принципах работы этого устройства, отметим только, что в настоящее время в оборудовании стандарта IS-95 используется несколько типов вокодеров, выдающих речевой поток данных с номинальными скоростями 8.6 и 13.3 кбит/с. Номинальная скорость отвечает наивысшей речевой активности (когда речь говорящего непрерывна), тогда как существуют три меньших скорости, отвечающих более низкой активности. Для наибольшей скорости в 8.6 кбит/с они составляют соответственно 4.0 кбит/с, 2.0 кбит/с и 0.8 кбит/с. Вокодер постоянно отслеживает энергию аналоговой речи на интервале 20 мс кадров и сравнивает ее с тремя адаптивными порогами для установления соответствующей скорости оцифрованного выхода. Уменьшение скорости в периоды низкой активности речи сопровождается пропорциональным уменьшением излучаемой мощности сигнала, так что значение энергии, приходящейся на бит, остается постоянной. В свою очередь, более низкий уровень мощности означает уменьшенный уровень помех множественного доступа MAI для других пользователей как внутри¹, так и вне соты и, возможно, большее число абонентов, обслуживаемых данной БС (см. параграф 4.6). Поток оцифрованной речи упаковывается в 20 мс кадры, содержащие, наряду с информационными битами, также CRC символы и 8 хвостовых бит, устанавливающих в нуль сверточный кодер. В результате, множество скоростей, отвечающих номиналу в 8.6 кбит/с, трансформируется на входе канала во множество RC1 «исходных» скоростей 9.6, 4.8, 2.4 и 1.2 кбит/с².

¹ Несмотря на то, что прямой канал является синхронным и не перенасыщенным, на практике он не свободен от помех MAI: взаимные многолучевые задержки могут нарушить ортогональность различных последовательностей Уолша.

² Помимо RC1 документами стандарта IS-95 устанавливается также набор исходных скоростей RC2 для номинала 13.3 кбит/с: 14.4, 7.2, 3.6 и 1.8 кбит/с, однако после сверточного кодера кодовый поток снова характеризуется скоростью 28.8 кбит/с вследствие выкалывания (см. параграф 9.3.1).

Обратимся к блок-схеме канала трафика, представленной на рис. 11.3. Входные данные на одной из четырех ранее упомянутых скоростей поступают на сверточный кодер скорости 1/2 и длиной кодового ограничения 9. Не зависимо от входной скорости выходная скорость вследствие символьного повторения всегда равна 19.2 кбит/с, например, поток со скоростью 1.2 кбит/с кодируется в поток с истинной скоростью 2.4 кбит/с, однако в выходном потоке каждый символ интерпретируется как 8 последовательных символов, укороченных в 8 раз. Перемежитель оперирует с блоками в 20 мс (384 кодовых бита), перемешивая символы кодового потока в соответствии с образцом, определяемым спецификацией. Затем сигнал с выхода переमेжителя скремблируется аналогично операции в канале вызова для обеспечения секретности передаваемого сообщения.

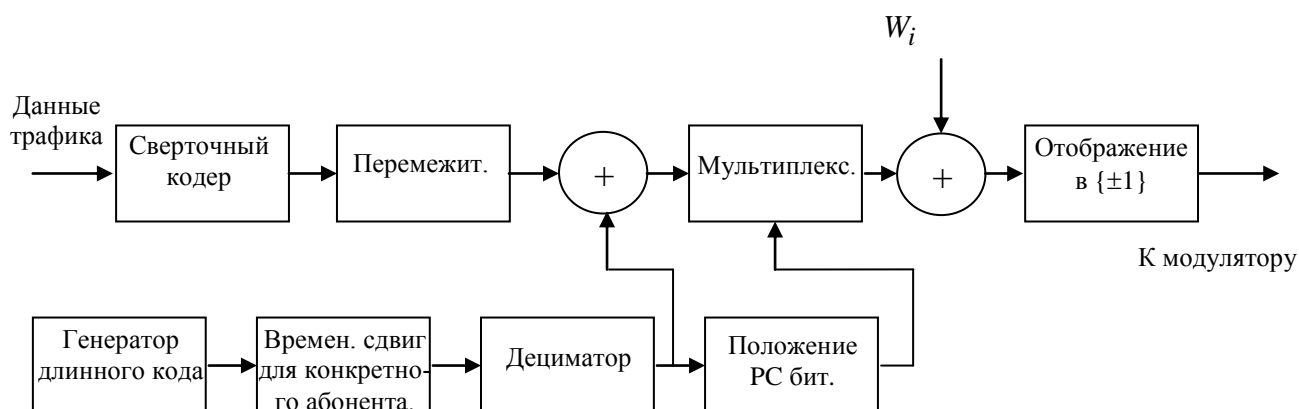


Рис.11.3. Упрощенная структура прямого канала трафика стандарта IS-95.

В параграфах 4.5-4.6 было установлено, что мощности сигналов мобильных станций должны эффективным образом контролироваться с целью преодоления проблемы близости-дальности и удержания уровня помех MAI на входе БС ниже критического уровня. Замкнутая петля регулировки мощности является одним из инструментов, решающих упомянутую задачу в стандарте IS-95. Базовая станция постоянно отслеживает интенсивность каждого принятого от МС сигнала и посылает управляющую команду МС либо на увеличение, либо на снижение излучаемой ею мощности. Команда представляет собой просто *биты регулировки мощности (power control bit)*, значение которых нулю или единице диктует соответственно увеличение или уменьшение мощности МС. Для вставки команд в сигнал прямого канала любой 20 мс кадр после перемежителя делится на 16 *групп регулировки мощности (power control groups (PCG))*, каждая из которых занимает интервал в 1.25 мс или $19.2 \cdot 10^3 \times 1.25 \cdot 10^{-3} = 24$ кодовых символа 19.2 кбит/с потока. В каждой PCG единственный бит контроля мощности заменяет два кодовых символа. Очевидно, что приемник МС, зная положение замененных символов (т.е. бит контроля мощности), исключает их из процедуры декодирования, как не имеющих отношения к содержанию сообщения. Данная процедура эквивалентна замене исходного сверточного кода кодом с выкалыванием (см. параграф 9.3), а ее негативное влияние на корректирующую способность кода полагается частично ослабленным за счет случайности позиций бит контроля мощности в пределах PCG. Псевдослучайная последовательность на выходе первого дециматора на рис. 11.3 имеет такую же скорость, как и кодовый поток, т.е. 19.2 кбит/с. На интервале одной PCG, длительностью 1.25 мс, располагаются 24 чипа этой последовательности. Последние четыре из них образуют двоичное число, причем 24-й чип отвечает наиболее значимому биту. Данное число, принимающее значение из диапазона от 0 до 15, используется в качестве указателя позиции бита контроля мощности в следующей после текущей группы PCG. Таким образом, бит контроля мощности может случайно занимать любую позицию из первых 16 в каждой группе PCG. На рис. 11.3 блоки,

реализующие позиционирование и вставку бит контроля мощности, обозначены как «Положение бит контроля мощности» и «Мультиплексор».

11.3.3.5. Модуляция в прямом канале.

На рис. 11.4 представлена блок-схема модулятора прямого канала. Выходное напряжение всех физических каналов БС первоначально взвешивается соответствующими коэффициентами усиления для осуществления управления мощностью прямого канала. Каждая МС периодически сообщает БС о достоверности принятых данных, и БС соответствующим образом регулирует уровень мощности сигнала в канале трафика, назначенном данной МС, для поддержания качества приема данных выше заданного порога. Затем

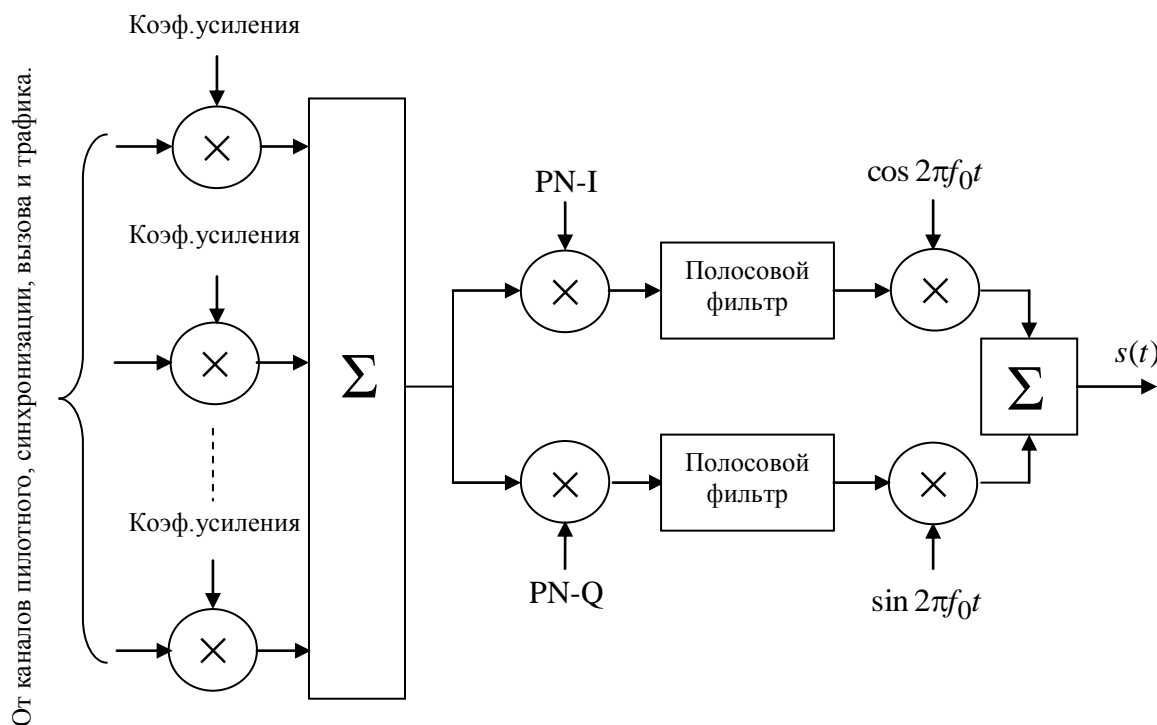


Рис.11.4. Модулятор прямого канала IS-95.

взвешенные каналные сигналы складываются в сумматоре и подаются параллельно в синфазную и квадратурную ветви модулятора для умножения с описанными выше бинарными PN-I и PN-Q кодами и формирования в частотной области полосовыми фильтрами. Умножение синфазного и квадратурного сигналов на косинусную и синусную CW компоненты частоты f_0 с их последующим суммированием завершает процедуру формирования сигнала БС и заканчивает процесс модулирования. Очевидно, что входной модулирующий сигнал в обеих ветвях является одним и тем же. Кроме того, он образован суммой множества бинарных напряжений, т.е. является многоуровневым вещественным сигналом. Временное предположение, что имеется только один физический модулирующий канал, немедленно приводит к ветвям модулятора без суммирования с другими каналами. Можно постулировать, что каждый физический канал обрабатывается именно таким образом, т.е. имеется такое множество пар ветвей модулятора, сколько присутствует каналов, и выходы всех этих параллельных модуляторов когерентно складываются. Поскольку схема на рис. 11.4 линейна, а, значит, действует принцип суперпозиции, то выходной эффект аналогичен тому, который имеет место в гипотетической схеме, т.е. с индивидуальными модулирующими каналами. По этой причине можно говорить, что в прямом канале связи стандарта IS-95 используется прямое расширение спектра, при котором бинарный поток

данных (канализированный с помощью функций Уолша) модулирует расширяющий код с QPSK (см. параграф 7.1). Поскольку скорость кодового потока на входе модулятора составляет 19.2 кбит/с, то каждый символ обладает длительностью, равной 64 чипам короткого кода. Следовательно, коэффициент расширения прямого канала равен 64.

Стоит отметить, что длинный код не играет роли в прямом расширении сигнала прямого канала, участвуя только в шифровании данных и определении местоположения бит контроля мощности. Часто говорят, что расширения прямого канала осуществляется как кодами Уолша, так и короткими PN-кодами. Концептуально, более подходящим выглядит классификация функций Уолша как каналообразующих, а PN-кодов – как расширяющих.

11.3.3.6. Обработка МС сигнала прямого канала.

Обработка сигнала в приемнике МС основывается на классических процедурах, подробно обсужденных в предыдущих разделах. При успешном поиске пилотного сигнала схема слежения за задержкой входит в захват и непрерывно отслеживает короткий код БС, находящейся в соединении. Локальная копия короткого кода, формируемая схемой DLL, используется для снятия расширения принятого сигнала. На выходе суженного пилотного канала наблюдается «чистая» СВ несущая, перенесенная на промежуточную частоту. Схема ФАПЧ согласовывает местный генератор на когерентную работу с принятым СВ сигналом, вырабатывая тем самым опорное колебание для когерентной демодуляции данных. После демодуляции и дегерережежения данные, передаваемые по каналам синхронизации, вызова и трафика, с помощью корреляторов, использующих в качестве опор последовательности Уолша, разделяются, декодируются с помощью алгоритма Витерби и используются в соответствии со своим назначением. Например, АЦП преобразует речевые данные канала трафика в напряжение, которое воспроизводится с помощью наушника.

Любой приемник МС содержит несколько (четыре и более) параллельных каналов, способных найти и отслеживать пилотный сигнал. Одной из целей подобного построения является организация RAKE приемника, материализующего преимущества многолучевого разнесения широкополосных сигналов (см. параграф 3.7). На практике, по меньшей мере, три подобных канала используются для реализации RAKE алгоритма. Другой процедурой, требующей независимых пилотных каналов в приемнике МС, является процедура передачи обслуживания (*handover*). Резервный коррелятор (или несколько) осуществляет непрерывное сканирование во временной области, стараясь обнаружить сигналы других базовых станций, характеризующихся большей интенсивностью сигнала и более пригодных для контакта. В последнем случае сеть может подать команду МС на переключение БС, которое легко выполняется, поскольку приемник уже отслеживает ее сигнал (мягкая эстафетная передача).

11.3.4. Обратный канал стандарта IS-95.

В соответствие с логическим содержанием данных, передаваемых по обратному каналу, любая МС работает по одному из двух возможных типов каналов:

- канал трафика (*traffic channel*),
- канал доступа (*access channel*).

11.3.4.1. Канал связи обратного трафика.

На рис. 11.5 представлена упрощенная структура канала обратного трафика. Поток передаваемых МС данных (оцифрованная речь от вокодера, компьютерные данные и т.п.) с вставленными CRC символами делится на 20 мс кадры, в которые затем добавляются 8 хвостовых бит для установки сверточного кодера в нулевое состояние к началу любого

кодированного кадра. В результате номинальная скорость поступления данных на входе кодера составляет 9.6 кбит/с, однако аналогично прямому каналу для пониженной речевой активности используются также три более низких скорости (4.8, 2.4 и 1.2 кбит/с). Вследствие асинхронного характера обратного канала помехи МАI (в отличие от прямого канала) будут существовать даже в гипотетическом случае отсутствия эффектов многолучевости (см. сноску на стр. 333). Это качественно оправдывает более напряженные помеховые условия обратного канала стандарта IS-95, объясняющие причины использования сверточного кода с более высокой корректирующей способностью наряду с последующей 64-

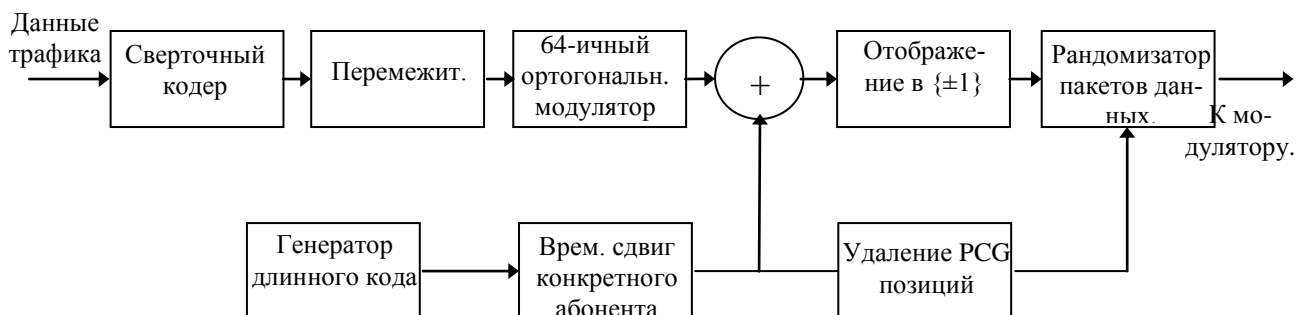


Рис.11.5. Упрощенная структура обратного канала трафика.

ичной ортогональной модуляцией. Поскольку увеличение длины кодового ограничения v_c влечет нежелательные усложнения кодека, то допустимым является то же самое значение, что и для прямого канала ($v_c = 9$), тогда как уменьшение скорости кода до $1/3$ служит вынужденной платой за улучшение дистанционных свойств. При данной скорости кода скорость выходного кодового потока составит 28.8 кбит/с независимо от скорости входного битового потока: символьное повторение, объясненное ранее для прямого канала, применяется также и в обратном канале.

Отдельный 20 мс кадр кодового потока (576 бит) делится на 16 групп регулировки мощности по 36 бит (1.25 мс) каждая. Перемежитель, оперирующий в рамках кадра, использует 32×18 матрицу, в которую кодовый поток записывается по колонкам. Считывание осуществляется построчно, причем каждая пара нечетной и следующей за ней четной строки образует PCG. Однако последовательные пары строк считываются согласно образцу, обеспечивающему соседство PCG, повторяющих друг друга (вследствие повторения кодированных символов), если исходная скорость потока данных меньше 9.6 кбит/с (т.е. 4.8, 2.4 и 1.2 кбит/с). Например, если исходная скорость составляет 4.8 кбит/с, то каждая четная PCG содержит повторение таких же перемеженных кодовых символов, что и предшествующая; для скорости потока данных 2.4 кбит/с группы с номерами $4i+2$, $4i+3$ и $4i+4$ являются копиями группы с номером $4i+1$ и т.п. Такой порядок является наиболее подходящим для процедуры последующего снижения средней излучаемой мощности пропорциональной уменьшению скорости, которая осуществлена в обратном канале несколько отлично по сравнению с реализацией в прямом канале (см. ниже).

Кодовый поток от перемежителя подается на 64-ичный ортогональный модулятор, в котором каждый его 6-символьный блок, рассматриваемый как 6-ти разрядное двоичное число, определяет один из 64 ортогональных сигналов (номер функции Уолша). Данная операция дает дополнительный выигрыш от кодирования (до трех раз в асимптотическом случае, см. параграф 2.6) к тому, который обеспечивается сверточным кодированием. Поскольку каждые 6 входных двоичных символов теперь заменяются 64 двоичными символами, то поток чипов с выхода ортогонального модулятора становится в $64/6 = 32/3$ раз быстрее (307.2 кбит/с). Отметим, что функции Уолша в обратном канале не имеют отношения к каналообразованию, а только осуществляют широкополосное ортогональное ко-

дирование передаваемых данных, обсужденное в пункте 2.7.3.

Следующим шагом по формированию канала трафика является расширение спектра, т.е. суммирование по модулю два потока бинарных символов после ортогонального модулятора со сдвинутой последовательностью длинного кода. Поскольку длинный код представляет собой поток чипов со скоростью 1.2288 Мчип/с, то 4 чипа длинного кода приходятся на один символ последовательности Уолша, или 256 чипов кода на один сигнал Уолша на выходе ортогонального модулятора. Приемник БС после снятия расширения использует 64-канальный банк корреляторов, каждый из которых настроен на одну из функций Уолша, и принимает решение в пользу такого опорного сигнала Уолша, при которой отклик на выходе коррелятора максимален. В этом случае интервал обработки охватывает полную длительность функции Уолша, т.е. 256 чипов длинного кода, и, значит, коэффициент расширения в обратном канале оказывается равным 256. Длинный код является единственным, жестко определяемым стандартом, а CDMA разделение различных пользователей обеспечивается специфическими для пользователей масками (т.е. временными сдвигами кода). Следовательно, в данном построении мы встречаемся со стратегией асинхронного варианта CDMA, которая обсуждалась в параграфе 7.4. Очевидно, что сдвиги должны быть выбраны таким образом, чтобы гарантировать от риска синхронного прихода на БС сигналов от двух МС, свободно перемещающихся во всей зоне покрытия. Величина сдвига определяется текущим идентификационным номером МС, назначаемым сетью, аналогично присвоению канальной несущей в FDMA. Параллельно с CDMA канали-

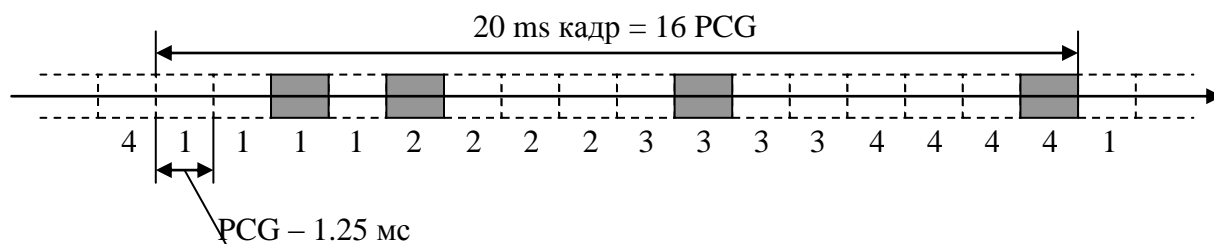


Рис.1.6. Пример образца передачи МС с исходной скорости 4.8 кбит/с.

зацией специфическая маска пользователя обеспечивает шифрование (скремблирование) потока после модулятора. Благодаря огромной длине длинного кода задача синхронизации локального генератора длинного кода, необходимая для перехвата сигнала и снятия расширения (т.е. дескремблирования и дешифрации) данных, не представляется простой для несанкционированного пользователя, который не обладает знанием пользовательской маски.

Операция отображения логических $\{0,1\}$ символов на вещественный модуляционный алфавит $\{\pm 1\}$ аналогична ранее рассмотренной и не требует дополнительных комментариев.

Одним из основных требований, предъявляемых к телефону МС, является достаточно долгая работоспособность батареи. С этой точки зрения линейный усилитель мощности, потребляющий большую среднюю мощность, является менее привлекательным, чем нелинейный (работающий в ключевом режиме). Это объясняет причину, по которой в передатчике МС снижения средней мощности добиваются не понижением мгновенной мощности, а альтернативным способом, состоящим в передаче только одной из всех повторяющихся PCG. Например, при исходной скорости в 2.4 кбит/с имеется повторенная четырежды идентичная PCG и только одна из них передается, тогда как в течение трех остальных передатчик выключен. Очевидно, что подобный режим превращает излучение МС в прерывное. Рандомизация (т.е. случайное определение) позиций отбрасываемых PCG (*рандомизация пакетов данных – data burst randomization*) способствует лучшему усреднению прерывных MAI на входе БС. Псевдослучайный образец вырезания PCG внутри

любого кадра определяется последними 14 чипами маски пользователя, т.е. сдвигом копии длинного кода, в конце предшествующего кадра. Правило, определенное в спецификации для любого конкретного значения исходной скорости, пересчитывает значения этих чипов (как двоичные числа) в позиции удаляемых PCG. Рис.11.6, на котором одинаковыми числами отмечены повторяющиеся PCG, тогда как затененные и пунктирные прямоугольники отвечают переданным и удаленным PCG, демонстрирует пример передачи образца для исходной скорости 4.8 кбит/с. Последовательность после рандомизатора пакетов данных может трактоваться как троичная с символами $\{\pm 1\}$ и нулями, соответствующими активной передаче и паузе соответственно.

11.3.4.2. Канал доступа.

МС использует канал доступа при ответе на уведомление о входящем вызове в состоянии покоя и либо при необходимости регистрации в сети, либо инициализации вызова. Процедуры деления на кадры, сверточного кодирования, ортогональной модуляции, перемежения и расширения длинным кодом в канале доступа в основном аналогичны процедурам в обратном канале трафика. Очевидно, что по данному каналу не передаются речевые данные, так что не производится регулировка скорости/мощности в соответствие с речевой активностью. Одна из специфичных черт канала доступа связана с инициализацией доступа МС. Не обладая точными сведениями об условиях распространения в обратном канале, МС начинает соединение посылкой пробных сигналов низкой интенсивности, постепенно увеличивая уровень сигнала с каждой последующей попыткой до получения подтверждения со стороны БС об установлении соединения. Пробные сигналы посылаются в пакетном режиме со случайными интервалами для уменьшения вероятности совпадения запросов от нескольких пользователей, поскольку существует вероятность, что в режиме доступа различным МС отвечает одно и то же значение маски длинного кода.

11.3.4.3. Модуляция в обратном канале.

Поскольку МС никогда не использует каналы трафика и доступа одновременно, то отсутствует необходимость суммирования канальных сигналов на входе модулятора, как это имело место в прямом канале. Выход канала трафика или доступа непосредственно параллельно подается в синфазную и квадратурную ветви, отличающиеся от изображенных на рис. 11.4 в следующих деталях. Во-первых, теперь отсутствует необходимость в идентификации БС сдвинутыми короткими кодами PN-1 и PN-2 (каждая мобильная станция обладает своим уникальным идентификатором – маской длинного кода – во всей зоне покрытия), и даже не пригодны для одновременного приема сигнала одной и той же МС несколькими базовыми во время мягкой эстафетной передачи. Таким образом, все МС используют нулевой сдвиг короткого кода. Во-вторых, имеется элемент задержки на половину длительности чипа, вставленный в квадратурную ветвь после умножителя на PN-Q. Это сделано с целью превращения QPSK в ее версию, называемую QPSK со сдвигом (O-QPSK). Последнее снова считается предпочтительным для продления работоспособности батареи. Обычный QPSK модулятор, изображенный на рис.11.4, может трактоваться как два BPSK модулятора, независимо работающие с косинусной и синусной составляющими CW несущими. В случае, когда бинарные символы в обеих ветвях одновременно изменяются на противоположные, фаза QPSK сигнала изменяется на 180° , и усилитель мощности передатчика должен обладать линейным динамическим диапазоном, равным примерно удвоенной амплитуде сигнала. В случае O-QPSK при изменении одного из двух модулирующих бинарных потоков другой остается неизменным благодаря сдвигу границ символов на половину длительности чипа. Следовательно, максимальный скачок фазы результирующего сигнала составит только 90° , смягчая тем самым требования к линейности динамического диапазона усилителя и способствуя большей работоспособности батареи.

11.3.5. Эволюция эфирного интерфейса от cdmaOne к cdma2000.

Одним из основных стимулов продвижения 3G стандартов было стремление к расширению функциональности мобильного телефона от простой трубки к терминалу, способному к высокоскоростному обмену данными с сетью, приему и отправке мультимедийной информации, доступу к Интернету и пр. Все эти новации требуют значительно более высокой скорости передачи по сравнению с системами 2-го поколения и основной скоростью, ассоциируемой с философией 3G является 2 Мбит/с. Подобное критическое увеличение скорости без компромисса с качеством услуг и числом обслуживаемых пользователей требует более широкой полосы системы. Для проектов cdma2000 характерно то, что, в отличие от концепции UMTS (см. следующий параграф), они рассматривают стандарт IS-95 в качестве отправной точки и принимают в расчет критерий обратной совместимости с ним. Из технологий, объединенных под общим названием cdma2000, один, базирующийся на MC-DS-CDMA принципе (см. пункт 10.2.1) организации прямого канала, выглядит более вероятным для принятия за основу. В данном параграфе затруднительно осветить детали этого предложения (обширное описание которого является предметом, например, [69]) вследствие их сложности и, возможно, не окончательности решения, так что далее последует лишь краткий обзор основных идей.

При трех несущих (в будущем это число может возрасти) прямой канал cdma2000 просто повторяет три раза спектр IS-95, занимая общую полосу порядка 3.75 МГц и используя в каждой из трех 1.25 МГц подполос методы расширения спектра и модуляции в чем-то отличные от применяемых в IS-95. В отличие от IS-95, где данные с BPSK манипулируют расширяющую последовательность с QPSK, в cdma2000 и данные, и расширяющий код манипулированы с помощью QPSK. Последнее позволяет при той же скорости данных удвоить длительность символов кодового потока и, следовательно, увеличить в два раза (до 128) коэффициент расширения на одну подполосу при той же чиповой скорости. Очевидно, что больший коэффициент расширения означает удвоение числа прямых физических каналов и, следовательно, потенциально большее число обслуживаемых пользователей. Но, что является даже более важным, число параллельных каналов, предоставляемых сетью некоторому пользователю, может способствовать передаче данных в число каналов раз с большей полной скоростью. Данный *многокодовый* метод, наследованный от спецификации IS-95B, является одним из основных ресурсов приближения скоростей к значениям, утверждаемым 3G концепцией.

Другим достоинством прямого канала cdma2000 является использование разнесения на передаче, в частности метода, близкого к описанному в пункте 10.3.5. Поскольку сигналы от различных передающих антенн распространяются по различным путям, то организовано несколько дополнительных пилотных каналов помимо одного, существующего в стандарте IS-95, передающих специфические для антенн пилотные сигналы от двух антенн БС, участвующих в разнесении. Это предоставляет возможность МС разделить пилотные сигналы различных антенн и строго синхронизовать опоры для сигналов разнесения, чтобы демодулировать, декодировать и соответствующим образом скомбинировать ветви разнесения.

Более сложная организация стандарта cdma2000 требует более строгой систематизации прямых логических каналов. Вместо четырех градаций в IS-95 иерархия логических каналов в cdma2000 насчитывает 10 только для первого уровня, включая общий канал управления, общий канал регулировки мощности, канал управления сообщениями, каналы быстрого вызова и др. Что касается каналов трафика, то они помимо прочих включают основные (первоначально используемые) и дополнительные каналы, которые привлекаются к многокодовой передаче в случае, когда высокая скорость передачи не достигается с использованием только основных каналов. Дополнительные каналы могут использовать либо сверточное, либо турбо-кодирование (см. параграф 9.4).

Обратный канал cdma2000 не использует MC-DS-CDMA метод, реализуя обычный

DS-CDMA с полосой в три раза более широкой (3.75 МГц). Этому значению отвечает в три раза более высокая чиповая скорость (3.6864 Мчип/с) длинного расширяющего кода. При столь большой скорости прямого расширения ряд решений стандарта IS-95 было пересмотрено и заменено более уместными. Прежде всего, используется сверточное кодирование со скоростью $1/4$ вместо $1/3$, означающее, что исходная скорость данных, равная например, 9.6 кбит/с становится 38.4 кбит/с. Использование модуляции данных с помощью QPSK и расширения такого кодового потока длинной последовательностью приводит к коэффициенту расширения, равного $2 \cdot (3.6864/38.4) \cdot 10^3 = 192$, что незначительно меньше 256 в IS-95. В то же время уменьшение скорости сверточного кода с длиной кодового ограничения 9 до величины $1/4$ приводит к некоторому дополнительному (не асимптотическому) выигрышу от кодирования. На основании этого факта проектировщики решили отказаться от ортогональной 64-ичной модуляции и сделать структуру обратного канала трафика более близкой к структуре аналогичного назначения в прямом канале.

Другим отличием служит широкий перечень логических каналов, включая обратный пилотный канал, который становится теперь обязательным, поскольку в отсутствие ортогональной модуляции БС необходимо иметь локальную когерентную опору для демодуляции QPSK данных МС. Имеются и другие новые логические каналы, часть из которых работают одновременно. Будучи образованными с помощью сигналов Уолша, упомянутые каналы затем линейно суммируются, означая, что усилитель мощности передатчика должен быть линейным, а значит, привлечение O-QPSK лишено смысла. Следовательно, используется обычный QPSK модулятор данными, при котором кодовый поток, отображенный QPSK, первоначально умножается на расширяющий код с QPSK, а затем вещественная и мнимая части произведения модулируют косинусную и синусную компоненты CW несущей, как это было объяснено в параграфе 7.1. Отметим, что аналогично прямому каналу обратные дополнительные каналы могут поддерживать обратный основной (трафика) канал (один на МС), обеспечивая многокодовую передачу на высоких скоростях, не достижимых с помощью только одного основного канала. Дополнительные каналы опять могут использовать либо сверточные, либо турбо-коды.

11.4. Эфирный интерфейс мобильной радиосвязи стандарта UMTS.

11.4.1. Предварительные замечания.

Система *Всемирной Мобильной Связи (Universal Mobile Telecommunication System (UMTS))* представляет собой широкополосный CDMA стандарт третьего поколения, проект которой был разработан Европейским телекоммуникационным сообществом. В настоящее время UMTS наряду с cdma2000 входят в т.н. IMT-2000 семейство, т.е. перечень стандартов, принятых Международным союзом связи (International Telecommunication Union (ITU)) в качестве основы для 3G (третьего поколения) систем. Существуют две версии UMTS: с *частотным дуплексом (frequency division duplex (FDD))* и *временным дуплексом (time division duplex (TDD))*. Непосредственно из названий видно, что эти системы отличаются друг от друга способом разделения прямого («вниз») и обратного («вверх») каналов. В FDD-UMTS прямой и обратный каналы занимают Неперекрывающиеся частотные полосы, тогда как в TDD-UMTS они используют различные временные слоты. Для краткости ограничимся обсуждением только FDD-UMTS, учитывая, что многие принятые решения являются общими для обеих систем.

В Европе для FDD-UMTS Международными регламентирующими документами выделены частотные полосы 1920-1980 МГц (канал «вверх») и 2010-2170 МГц (канал «вниз») с ограничением в 5 МГц на ширину полосы канала. Прямое расширение спектра (DS spreading) является основной технологией, обеспечивающей разделение физических каналов и, в частности, множественный доступ (DS-CDMA). Стандарт определяет единую и постоянную чиповую скорость в 3.84 Мчип/с в полном соответствии с ограничением на полосу канала. UMTS, как и cdma2000, является системой, в которой скорость передачи может варьироваться в очень широких пределах. Следствием этого, наряду с неизменностью чиповой скорости, является изменение коэффициента расширения при изменении скорости передачи.

В дальнейшем снова сконцентрируем наше внимание на физическом уровне системы, т.е. решениях, касающихся расширения, канализации и модуляции. Подчеркнем, что БС системы UMTS не пользуются поддержкой GPS и функционируют с собственными автономными не синхронизированными опорными генераторами. Выгодная с точки зрения экономии оборудования, подобная архитектура, однако, превращает поиск и эстафетную передачу в более сложные процедуры и определяет многие отличия в физических уровнях UMTS и cdma2000.

Аналогично параграфу 11.3, данный материал есть не более чем очень краткий обзор, однако множество недавно опубликованных книг, посвященных стандарту UMTS [92,104,114,115,120-122], помогут заинтересованному читателю в расширении своих знаний о данной системе, которая в ближайшие годы обязана стать доминирующей среди систем мобильной связи.

11.4.2. Типы каналов стандарта UMTS.

В соответствии с терминологией спецификаций UMTS различают логические, транспортные и физические каналы. На уровнях, выше физического, данные распределяются между логическими каналами на основе информационного содержания, но перед переходом на физический уровень они реструктурируются в транспортные каналы. Критерием различения транспортных каналов служит способ и формат представления данных, тогда как физические каналы (которые есть просто сигналы, переносящие информацию), как и в любой CDMA системе, различаются своими специфическими кодами. Физический уровень состоит из двух подуровней. Передаваемые данные доставляются на первый физический подуровень с верхних уровней, упакованными в транспортные каналы в соответствии с информационным содержанием. Первый подуровень включает среди других при-

соединение CRC символов для защиты блоков данных, канальное кодирование и перемежение. Канальное кодирование выполняется либо как сверточное (длина кодового ограничения 9, скорость 1/3 или 1/2) или как турбо-кодирование (образованное кодером памяти 3 и скорости 1/3). Для очень высокой скорости передачи данных возможна также не кодированная передача. Второй подуровень, т.е. линия радиосвязи, охватывает отображение транспортных каналов на физические (т.е. сигналы), передачу сигналов через среду распространения. Сигналы, принятые МС (канал «вниз») или БС (канал «вверх»), подвергаются затем всем необходимым взаимным операциям двух подуровней (демодуляции, деперемежанию, декодированию и т.п.), обратному отображению, если это необходимо, и передаче данных на более высокий уровень.

Другая классификация, применимая как к транспортным, так и физическим каналам, различает *общие* (*common*) и *выделенные* (*dedicated*) каналы. Первые из них содержат данные, относящиеся ко всей системе и используемые совместно всеми пользователями, тогда как вторые служат только для установления соединения между БС и конкретной МС.

Временная структура всех каналов строго определена стандартом. Все они состоят из кадров длительностью 10 мс (38 400 чипов), а каждый кадр, в свою очередь, поделен на 15 слотов длительностью 666.66 мкс (2560 чипов).

11.4.3. Выделенные физические каналы «вверх».

Выделенный физический канал является каналом, который сеть предоставляет МС для эксклюзивного использования, т.е. два мобильных абонента в пределах одной и той же соты никогда не используют одинаковый выделенный канал. Существуют два типа выделенных каналов «вверх»: *выделенный физический канал данных* (*dedicated physical data channel* (DPDCH)¹) и *выделенный физический канал управления* (*dedicated physical control channel* (DPCCH)), наименование которых согласуется с их функциональным назначением. Канал DPCCH используется для передачи БС служебной информации: пилотный сигнал, необходимый для оценки параметров распространения и синхронизации когерентной опоры БС, сообщение о формате данных в DPDCH, информацию обратной связи, используемую при эстафетной передаче, и команды регулировки мощностью. На одну МС всегда приходится только один DPCCH точно также, как передаваемые пользовательские данные попадают на физический уровень, упакованные в единственный транспортный канал, приходящийся на МС. Однако этот единственный транспортный канал может быть отражен на несколько (вплоть до 6) DPDCH каналов, если необходимая скорость превосходит максимальную скорость, обеспечиваемую единственным DPDCH каналом.

Мультиплексирование выделенных каналов осуществляется следующим образом (см. рис. 11.7). Канал DPCCH передается через квадратурную ветвь QPSK модулятора, тогда как первый DPDCH канал поступает в синфазную ветвь. Остальные DPDCH каналы при их использовании по возможности равномерно распределяются между ветвями. Таким образом, несмотря на применение QPSK модулятора в канале «вверх» используется BPSK модуляция данных. Скорость передачи по DPCCH всегда постоянна и равна 15 кбит/с или 10 бит/слот. Тогда каждый бит занимает интервал, равный $2560/10=256$ бит, и коэффициент расширения в DPCCH всегда равен 256. В отличие от этого коэффициент расширения в DPDCH канале варьируется в зависимости от требуемой скорости данных и может изменяться от 256 (минимальная скорость) до 4 (максимальная скорость). Таким образом минимальная скорость передачи данных равна 15 кбит/с, тогда как максимальная скорость на один DPDCH канал в 64 раза выше, т.е. 960 кбит/с. Использование 6 параллельных DPDCH каналов в многокодовом режиме передачи, описанном в предыдущем разделе, в принципе позволяет достичь максимальной скорости, равной 5760 кбит/с. Оче-

¹ Наименование и аббревиатура каналов UMTS в точности воспроизводит используемые в документах UMTS.

видно, что т.н. gross-скорость (т.е. скорость кодового потока после канального кодера) является желаемой, а исходная скорость потока данных пропорционально меньше скорости канального кода.

Каждый физический канал «вверх» обладает собственным каналообразующим кодом, умноженным на кодированные данные. Рис. 11.7 иллюстрирует ситуацию, когда все DPDCH каналы активны, используя 6 различных каналообразующих кодов KK_1, \dots, KK_6 , с обозначением KK_c , зарезервированным для каналообразующего кода DPCCCH. После умножения на канализирующий код все физические каналы взвешиваются своими коэффициентами усиления, реализуя регулировку мощности обратного канала. Фактически используются только два различных коэффициента: при любом числе DPDCH каналов их коэффициент усиления один и тот же G_d , тогда как в DPCCCH канале он равен G_c . Максимальное значение коэффициента усиления равно единице, его нулевое значение означает отключение канала, а изменение значений коэффициента осуществляется с шагом 1/15.

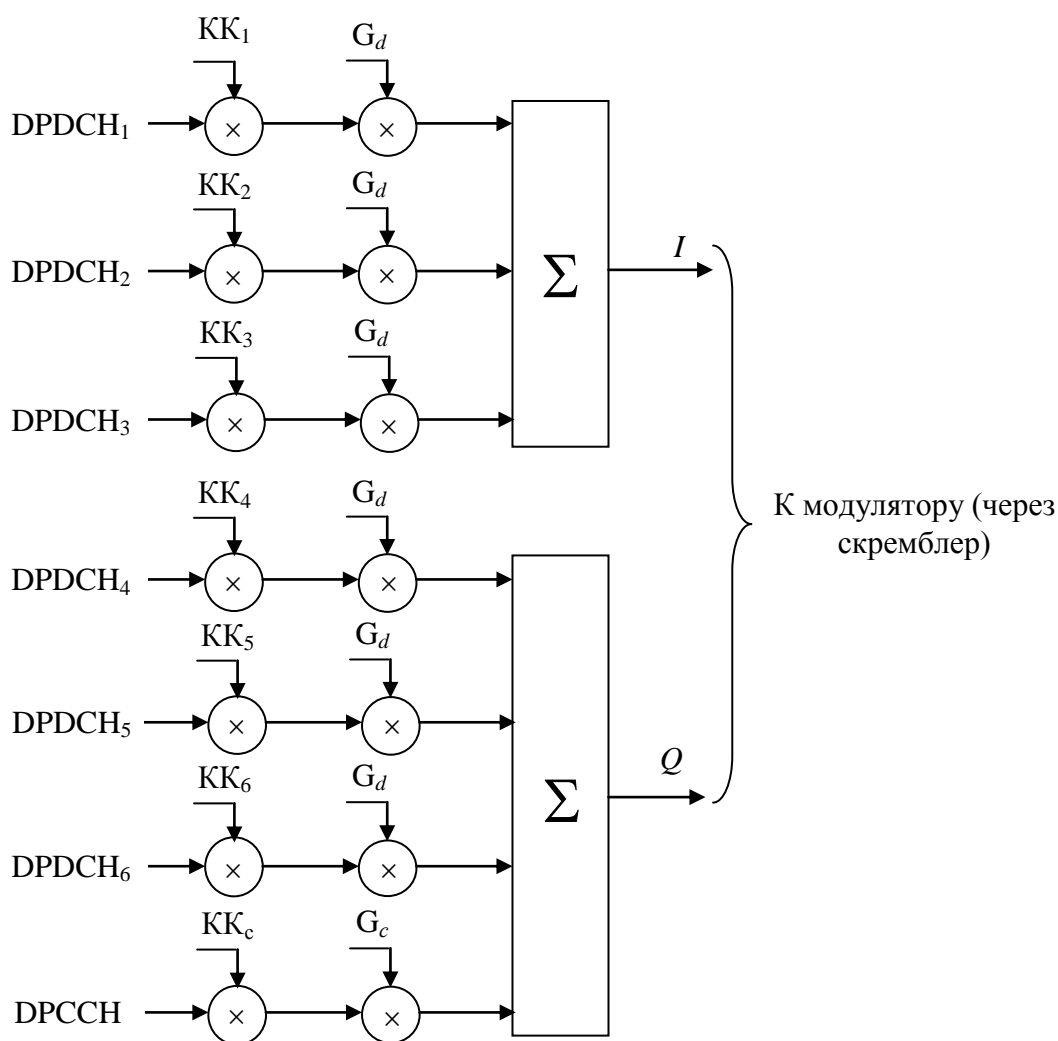


Рис.11.7 Мультиплексирование выделенных каналов «вверх».

11.4.4. Общие физические каналы «вверх».

Общие физические каналы представляют собой ресурс, которые находится в общем распоряжении всех мобильных абонентов. Существуют два типа общих каналов «вверх»: канал случайного доступа (*random access channel (RACH)*) и общий канал пакетирования

(*common packet channel* (CPCH)), название которых используется как для транспортных, так и соответствующих физических каналов. МС использует канал RACH для установления соединения (например, вызова) с сетью и для передачи коротких пакетных сообщений, а CPCH является основным ресурсом для передачи пакетов мобильной станцией.

Физический RACH (PRACH) канал содержит преамбулу из 4096 чипов, которая представляет собой 16-чиповый идентификатор МС, повторенный 256 раз. МС может начать передачу по PRACH в начале любого из 15 слотов доступа, занимающих совместно два кадра, т.е. 5120 чипов или 20 мс. Диаграмма слотов доступа задается в формате передачи БС, использующей специальный транспортный канал «вниз» (*broadcasting channel* (BCH)). В начальном состоянии у МС нет достоверных сведений об ослаблении сигнала в канале «вверх», и первая передача преамбулы реализуется при низком уровне мощности. В случае отсутствия подтверждения от БС об установлении соединения МС случайным образом выбирает новые слоты доступа и предпринимает новые попытки соединения, каждый раз увеличивая мощность сигнала. После получения от БС подтверждения МС передает собственно сообщение, занимающее один или два кадра (10 или 20 мс). В канале PRACH отсутствует замкнутая петля регулировки мощности, поскольку сеансы соединения по нему чрезвычайно коротки.

Структуры физических каналов CPCH (PCPCH) и PRACH в основном подобны, однако сегмент сообщения по каналу PCPCH может занимать большее число кадров, а сегмент преамбулы сопровождается еще одной частью: преамбулой коллизии обнаружения, помогающей выявить одновременные попытки нескольких МС воспользоваться каналом PCPCH. Вследствие большей длительности пакетов PCPCH представляется оправданным присутствие в канале замкнутой петли регулировки мощности в сравнении с PRACH.

Как и в выделенных, в общих каналах PCPCH и PRACH используется квадратурное мультиплексирование для комбинирования информационных и служебных компонентов передаваемых данных. Разделение общих и выделенных физических каналов осуществляется с помощью канализирующих кодов, рассматриваемых в следующем пункте.

11.4.5. Канализирующие коды линии «вверх».

Как было установлено выше, индивидуальная МС использует несколько типов каналов, которые должны быть разделены в приемнике БС. Поскольку все физические каналы МС привязаны к одной и той же временной шкале, то наилучшим способом организации разделения физических каналов (нескольких DPDCH, DPCCCH, PRACH, PCPCH) случит синхронное кодовое разделение. Отметим, что речь идет только о разделении каналов отдельной МС, тогда как сигналы других мобильных абонентов разделяются с помощью асинхронного варианта CDMA, аналогично реализованному в cdmaOne и cdma2000.

В документах по UMTS формат канализирующего кодирования описан с помощью двоичного графа, определяемого итерационной процедурой. На каждом шаге итерации любое кодовое слово, полученное на предыдущей итерации, порождает два новых слова удвоенной длины путем добавления либо исходного слова, либо его инверсии. Пусть \mathbf{c}_k – некоторый кодовый вектор, полученный на k -м шаге итерации. Тогда два кодовых вектора удвоенного размера $\mathbf{c}_{k+1} = (\mathbf{c}_k, \mathbf{c}_k)$ и $\mathbf{c}'_{k+1} = (\mathbf{c}_k, -\mathbf{c}_k)$ являются его потомками на следующей итерации. Таким образом, отталкиваясь от тривиального слова (1) длины один, после k итераций приходим к 2^k канализирующим кодовым словам длины $N = 2^k$. (На рис. 11.8 представлен пример построения для $k = 3$). Легко заметить, что данный алгоритм отличается от конструирования матрицы Адамара по правилу Сильвестра (см. пункт 2.7.3) только переупорядочиванием строк результирующей матрицы, так что полученные кодовые слова есть ничто иное, как функции Уолша. Несмотря на это в спецификациях по UMTS они фигурируют под специальным названием *ортогональных кодов с переменной*

величиной расширения (*orthogonal variable spreading factor (OVSF)*)).

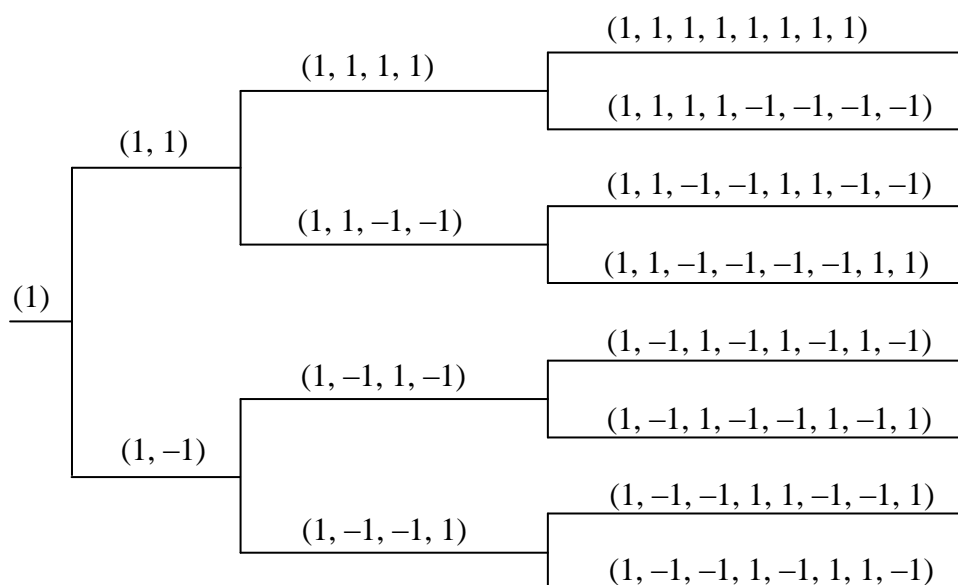


Рис.11.8. Граф OVSF кодов длины 8.

Граф OVSF кодов канала «вверх» построен с использованием $k=8$ итераций, так что кодовые слова имеют максимальную длину, равную $n = 256$ чипов. Состоящее из одних единиц слово указанной длины определено для выделенного канала управления DPSSN, в результате чего последующее скремблирование обеспечивает DPSSN каналу коэффициент расширения спектра 256, о чем упоминалось в пункте 11.4.3. Выбор каналообразующих кодов аналогичной длины для DPDCN (коэффициент расширения 256) позволяет вести передачу данных со скоростью 15 кбит/с. Если требуется более высокая скорость, то символы данных укорачиваются, т.е. уменьшается коэффициент расширения (чиповая скорость никогда не изменяется, оставаясь равной 3.84 Мбит/с). Тогда соответствующие DPDCN канализации слова берутся с промежуточной (k -й, $SF = 2^k$ – требуемое значение коэффициента расширения спектра) итерации графа OVSF кода. Если требуемая скорость достижима при единственном DPDCN канале, то номер его канализирующего слова определяется спецификацией как $SF/4 = 2^{k-2}$ при ведении отсчета по дереву сверху вниз и нумерации наивысшего слова нулевым. Подобный выбор всегда сохраняет ортогональность DPDCN и DPSSN каналов вне зависимости от коэффициента расширения в DPDCN канале. Если единственный DPDCN канал (с коэффициентом расширения 4) не может обеспечить требуемую скорость данных, то привлекается метод многокодовой передачи по нескольким DPDCN каналам, обладающих одним и тем же минимальным коэффициентом расширения 4. Имеются три кодовых слова длины 4, ортогональные друг другу и кодовому слову DPSSN канала, так что, учитывая возможность использования одного и того же слова для двух DPDCN каналов в схеме квадратурного мультиплексирования (рис. 11.7), может быть организовано до 6 многокодовых DPDCN каналов.

Способ назначения кодовых слов PRACH и PCPCH каналам, устанавливаемый спецификацией, также обеспечивает их ортогональность выделенным каналам во всем диапазоне скоростей передачи данных по DPDCN каналам.

11.4.6. Скремблирование линии «вверх».

Завершающим этапом формирования сигнала MC является его прямое расширение

с помощью специфической для каждого пользователя сигнатурой, реализующей асинхронный вариант CDMA разделения сигналов различных МС в приемнике БС. Согласно терминологии UMTS, данная операция, как и подобная в линии «вниз», называется *скремблированием*, употребление термина которой несколько отличается от используемого в документах по cdmaOne и cdma2000 (см. подпункты 11.3.3.3 и 11.3.3.4). Существует два типа скремблирующих кодов линии «вверх», определенных стандартом: длинный или короткий коды, применяемые при скремблировании выделенных каналов.

Последовательности Голда (пункт 7.5.2) длиной $L = 2^{25} - 1$, усеченные до длительности одного кадра, т.е. 38 400 чипов, используются для построения длинных скремблирующих кодов. Генераторы с линейной обратной связью, определяемой полиномами $f_1(x) = x^{25} + x^3 + 1$ и $f_2(x) = x^{25} + x^3 + x^2 + x + 1$, формируют две m -последовательности, вторая из которых может быть также получена в результате децимации первой с индексом, присущим кодам Голда. Сложение по модулю два сдвинутой копии второй m -последовательности с первой дает последовательность Голда, которая затем усечается и отображается на $\{\pm 1\}$ алфавит. «Истинная» усеченная последовательность Голда $\{c_i\}$ определяет вещественный компонент QPSK сигнатуры МС. Для получения мнимой части исходная (до усечения) последовательность Голда вначале сдвигается на 777 232 чипа, а затем подвергается усечению и отображению с целью построения бинарной $\{\pm 1\}$ последовательности $\{c'_i\}$. Далее, инверсией каждого четного элемента последней последовательности заменяется следующий нечетный элемент, а полученный результат посимвольно перемножается с $\{c_i\}$. Последняя операция вдвое уменьшает число элементов в результирующей QPSK последовательности, полярность которых противоположна предшествующему. Действительно, соотношение, описывающее полученную таким способом QPSK сигнатуру, представимо в виде

$$a_i = c(i) + j(-1)^i c(i)c'(2\lfloor i/2 \rfloor), \quad (11.1)$$

где $c(i)$ и $c'(i)$ – просто более подходящее обозначение для элементов c_i и c'_i . Теперь каждый может увидеть, что при переходе от четной позиции $i = 2l$ к следующей нечетной $i = 2l + 1$ вещественная и мнимая части не могут изменяться одновременно, означая, что максимальные фазовые скачки возможны только на $\pm 90^\circ$ и никогда на 180° , так что чиповая полярность может изменяться в противоположную только во время переходе от нечетного к четному чипам. Как правило, уменьшение частоты противоположных переходов желательно с точки зрения продления работоспособности батареи, поскольку усилитель мощности в эти моменты рассеивает большую энергию.

Завершающий шаг QPSK скремблирования состоит в умножении мультиплексированного сигнала МС, взятого из ветвей схемы на рис. 11.7, со скремблирующей последовательностью. Эта операция, осуществляемая обычным QPSK модулятором, была обстоятельно обсуждена в пункте 7.1.2.

Скремблирующая последовательность точно синхронизирована с временной шкалой МС, начинаясь в любом кадре с одного и того же символа (чипа).

Отметим, что главное достоинство кодов Голда – минимаксные периодические корреляционные свойства – не могут объяснить их выбор для скремблирования линии «вверх», поскольку усечение радикально ухудшает корреляционные свойства. Альтернативным объяснением может служить легкость формирования множества (не менее чем $2^{25} + 1$) псевдослучайных последовательностей.

Короткие скремблирующие последовательности обладают периодом 256 и планируются к использованию в случае, когда приемник БС готов к выполнению многокодового алгоритма (см. параграф 10.1). В случае асинхронного CDMA сложность этого типа приемника обычно возрастает с ростом длины расширяющего кода. Правило формирова-

ния коротких скремблирующих кодов, устанавливаемое спецификацией, включает операции генерирования четвертичной линейной рекуррентной последовательности длиной 255, ее суммирования по модулю четыре с двумя двоичными рекуррентными последовательностями длин 51 и 85, расширения результирующей последовательности на один элемент до длины 256, отображения четвертичных символов на QPSK алфавит и преобразования полученной комплексной последовательности по правилу (11.1).

11.4.7. Отображение транспортных каналов «вниз» на физические каналы.

Информация, передаваемая сетью некоторой конкретной МС на транспортном уровне, организована в виде единственного выделенного канала, который затем отображается двумя физическими каналами «вниз», подобными в линии «вверх»: *выделенные каналы данных* (*dedicated physical data channel (DPDCH)*) и *управления* (*dedicated physical control channel (DPCCH)*).

Перечень общих транспортных каналов линии «вниз» много шире по сравнению с линией «вверх». В частности, он содержит уже упоминавшийся канал оповещения BCH, по которому передаются параметры сети или соты, используемые всеми МС; *прямой канал доступа* (*forward access channel (FACH)*), используемый БС для отсылки командных данных МС, положение которой известно; *канал вызова* (*paging channel (PCH)*), посредством которого БС передает команды МС с неизвестным местоположением и т.п. На физическом уровне *первичный общий физический канал управления* (*primary common control physical channel (P-CCPCH)*) передает данные BCH канала, канал FACH отображается во *вторичный общий физический канал управления* (*secondary common control physical channel (S-CCPCH)*), передача данных PCH частично осуществляется снова S-CCPCH каналом и частично *каналом синхронизации* (*synchronization channel (SCH)*). Еще имеется *общий пилотный канал* (*common pilot channel (CPICH)*), транслирующий сигнал немодулированных данных, которые используются приемником для оценки параметров канала. Полная схема соответствия между транспортными и физическими каналами линии «вниз» устанавливается спецификацией, которая может быть найдена в [92,121].

11.4.8. Формат физических каналов линии «вниз».

Механизм мультиплексирования выделенных каналов данных и управления линии «вниз» DPDCH и DPCCH отличается от используемого в линии «вверх»: каждый слот разбит на части и каждая часть отведена либо данным DPDCH канала, либо DPCCH канала. При подобной организации передатчик БС работает в прерывистом режиме, что обусловлено паузами в потоке данных во время низкой речевой активности. Прерывистость излучения в работе передатчика МС может вызвать помехи, сильно затрагивающие соседние электронные устройства (например, слуховой аппарат). О данном явлении неоднократно сообщалось при использовании телефона стандарта GSM. Однако для БС прерывистость излучения не является столь значимым фактором, поскольку маловероятно, чтобы приборы, о которых говорилось выше, находились достаточно близко от антенны БС.

Другим отличием формата физического канала линии «вниз» служит применение QPSK модуляция данных (в отличие от BPSK в линии «вверх»). Во-первых, демультимплексор (DMUX) расщепляет битовый поток после первого физического подуровня (канальное кодирование, перемежение и т.п.) на два, содержащих четные и нечетные биты исходного потока соответственно. Оба потока, отображенные на $\{\pm 1\}$ алфавит, умножаются на двоичный каналообразующий код (CC), создавая вещественную и мнимую части комплексного канализирующего сигнала данных, что демонстрирует рис. 11.9. Завершающим шагом является умножение данного QPSK сигнала данных на QPSK расширяющий (скремблирующий) код и QPSK модуляция несущей, аналогично осуществляемой в линии «вверх».

11.4.9. Канализирующие коды линии «вниз».

Мультиплексирование во времени каналов данных и управления, рассмотренное в предыдущем пункте, означает, что для передачи обоих этих информационных потоков необходим только один *выделенный физический канал* (*dedicated physical channel (DPCH)*). Если единственный данный канал удовлетворяет требованиям скорости передачи данных, то БС формирует его с помощью специфической для данного пользователя кодовой последовательности, которая, конечно, является единственной в соте (или секторе) и не может быть использована для соединения с любой другой МС. Сценарий многокодовой пе-

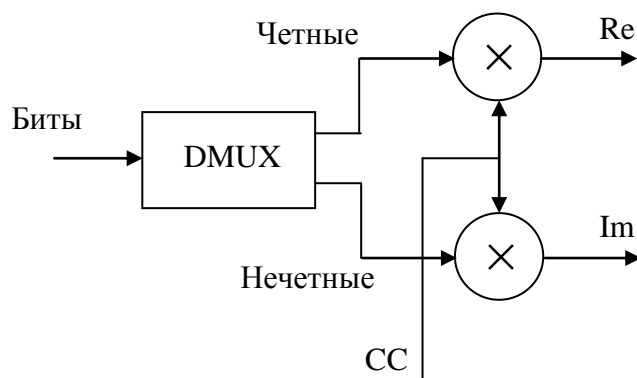


Рис.11.9. Демультимплексирование и канализация данных в линии вверх.

редачи возникает тогда, когда единственный физический канал не способен обеспечить передачу данных с требуемой скоростью. Тогда БС организует несколько параллельных физических каналов для соединения с той же МС. Эти каналы всегда строятся с одинаковым коэффициентом расширения спектра, но, очевидно, с различными канализирующими кодами, которые не могут использоваться БС при соединении с другими мобильными. Поскольку отсутствует необходимость в повторении командных сообщений по всем многокодовым каналам, то информация управления пересылается МС только по одному из них.

В подобных условиях каналообразующие коды линий «вниз» и «вверх» принадлежат одному и тому же типу. Используются семейство функций Уолша или OVSF графа на рис. 11.8 с коэффициентом расширения спектра из диапазона от 4 до 512. Некоторые канализирующие последовательности не доступны для DPCH канала, поскольку предназначены для общих каналов, как CPICH. Например, при минимальном значении коэффициента расширения только три кодовые последовательности могут обслуживать DPCH канал, обеспечивая максимально возможную скорость передачи данных 2.88 Мбит/с. В предположении скорости канального кода 1/2 и наличии неизбежного заголовка (команд управления и т.д.) данная скорость не удовлетворяет требованиям, предъявляемым к 3G системам, по величине «истинной» скорости передачи данных (до 2 Мбит/с), что является одной из причин стимулирования спецификацией высокоскоростной некодированной передачи данных.

При работе многих пользователей с разными скоростями их канализирующие коды должны сохранять ортогональность, несмотря на различие в величине коэффициента расширения спектра или, что то же самое, в длине кода. Как следует из рис. 11.8, две функции Уолша различной длины ортогональны на интервале минимальной длины, если и только если ни одна из них не является потомком другой. Тогда становится понятным, что в отличие от линии «вверх», где любая МС, выделяемая от других своим уникальным скремблирующим законом, имеет в эксклюзивном распоряжении все множество канализирующих кодов, управление канализированием в линии «вниз» является гораздо более сложной задачей. Действительно, разным МС следует назначать различные подмножества

последовательностей Уолша, не содержащие потомков последовательностей, обслуживающих в текущее время другие МС. Данная проблема динамического распределения ресурса решается на более высоком уровне протоколом сети.

11.4.10. Скремблирующие коды линии «вниз».

Как указывалось ранее, скремблирующие коды линии «вниз» обеспечивают разделение сигналов различных БС. Каждая скремблирующая последовательность базируется на последовательности Голда длины $L = 2^{18} - 1 = 262\,143$. Два LFSR генератора с обратной связью, задаваемой полиномами $f_1(x) = x^{18} + x^7 + 1$ и $f_2(x) = x^{18} + x^{10} + x^7 + x^5 + 1$, формируют исходные m -последовательности, связанные друг с другом, как того требует построение ансамбля Голда (см. пункт 7.5.2). Хотя имеется $2^{18} + 1 = 262\,145$ последовательностей Голда указанной длины, спецификация ограничивает число используемых величиной $2^{13} = 8\,192$. Затем от каждой разрешенной последовательности отрезаются два сегмента длиной 38 400: исходный и сдвинутый на $2^{17} = 131\,072$ чипа, отображаемые далее в $\{\pm 1\}$ алфавит по обычному правилу (6.15). Результирующие бинарные последовательности представляют собой вещественную и мнимую части QPSK скремблирующего кода линии «вниз».

Спецификацией устанавливается строгая иерархия скремблирующих последовательностей. Множество всех скремблирующих кодов делится на 512 подмножеств, каждое из которых содержит один первичный и 15 вторичных кодов. В свою очередь, все 512 первичных кодов поделены на 64 группы по 8 последовательностей в каждой. Конкретной БС назначается только один уникальный первичный код. Для некоторых физических каналов разрешено использовать только первичный код, тогда как для других можно применять как первичный, так и вторичный коды.

11.4.11. Канал синхронизации.

11.4.11.1. Общая структура.

Канал синхронизации (SCH) играет исключительно важную роль в структуре сети, обеспечивая начальный поиск соты и согласование временной шкалы МС с границами кадров и слотов принятого сигнала БС. Ни канализирующие, ни скремблирующие коды не принимают участия в формировании канала синхронизации, поскольку его сигнал должен быть обнаружен и обработан, прежде чем МС будет осведомлена о скремблирующем коде контактирующей с ней БС. Подчеркнем еще раз, что временные шкалы базовых станций сети UMTS не синхронизированы между собой и что при переключении МС на другую базовую (например, в ходе эстафетной передачи) некоторое время должен осуществляться начальный поиск соты. Последнее служит обоснованием одного из ключевых положений проектирования архитектуры SCH канала, заключающегося в проведении двухэтапной процедуры поиска, которая потенциально способна обеспечить экономию времени поиска по сравнению с обычным последовательным поиском (см. параграфы 8.2, 8.3). Действительно, SCH канал образован парой из первичного и вторичного каналов синхронизации, сигналы которых – первичный и вторичный коды синхронизации – используются на первом и втором этапах поиска соответственно. Оба сигнала занимают начальные 256 чипов в каждом слоте. Последовательность первичного кода синхронизации идентична не только во всех слотах, но и для всех базовых станций сети. Указанный факт делает невозможным установление синхронизации с первичным сигналом заранее определенной БС; определение БС, с которой установлен контакт МС после завершения первого этапа синхронизации, становится ясным только по завершению второго этапа. После нахождения первич-

ного сигнала синхронизации МС знает границы слотов, но не кадров. На втором этапе МС должна снять эту неопределенность, проверяя все 15 (по числу слотов в кадре) возможных значений рассогласования между шкалой МС и принятым сигналом БС для всех возможных версий вторичного кода синхронизации. Для способствования решению этой задачи вторичный код синхронизации имеет период, равный одному кадру (15 слотам или 38 400 чипам), и кодовую структуру, специфичную для данной БС. Кроме того, вторичный код синхронизации жестко связан с группой скремблирующих кодов (одной из 64), назначенных базовой станции. Таким образом, после завершения второго этапа поиска МС опознает группу скремблирующих кодов захваченного сигнала БС и, тестируя 8 гипотез относительно возможной последовательности первичного скремблирующего кода, окончательно определяет, какая из них используется контактируемой БС.

11.4.11.2. Первичный код синхронизации.

Первичный код синхронизации (primary synchronization code (PSC)) определен спецификацией как 16-элементная последовательность

$$\mathbf{a} = (1, 1, 1, 1, 1, 1, -1, -1, 1, -1, 1, -1, 1, -1, -1, 1),$$

повторяющаяся 16 раз с прямой или инвертированной полярностью, образуя 256-элементную бинарную последовательность

$$\mathbf{c}_{ps} = (\mathbf{a}, \mathbf{a}, \mathbf{a}, -\mathbf{a}, -\mathbf{a}, \mathbf{a}, -\mathbf{a}, -\mathbf{a}, \mathbf{a}, \mathbf{a}, -\mathbf{a}, \mathbf{a}, -\mathbf{a}, \mathbf{a}, \mathbf{a}).$$

Каждая БС непосредственно передает одну и ту же последовательность \mathbf{c}_{ps} , реализуя PSC и предоставляя МС возможность осуществить первый этап поиска соты. Поскольку сигнал PSC предназначен для измерения времени в обстановке многолучевости, то он должен обладать хорошими автокорреляционными свойствами (см. параграф 6.1). Несмотря на периодичность PSC большая часть его периода свободна, так что уровень боковых лепестков аперiodической АКФ является адекватным критерием качества PSC. Как указывалось в примере 6.10.2 (см. также рис. 6.16), аперiodическая АКФ PSC системы UMTS достаточно далека от идеала, что, возможно, является неизбежной платой за простоту аппаратной реализации, к которой стремились разработчики.

11.4.11.3. Вторичный код синхронизации.

Вторичные коды синхронизации (secondary synchronization codes (SSC)) строятся на основе 16-элементной последовательности \mathbf{b} , повторяющей последовательность \mathbf{a} в первых 8-ми элементах, и с инверсией \mathbf{a} (т.е. $-\mathbf{a}$) – в остальных:

$$\mathbf{b} = (1, 1, 1, 1, 1, 1, -1, -1, -1, 1, -1, 1, -1, 1, 1, -1).$$

Данная последовательность повторяется 16 раз с изменением полярности или без него, образуя 256-элементную последовательность вида

$$\mathbf{z} = (\mathbf{b}, \mathbf{b}, \mathbf{b}, -\mathbf{b}, \mathbf{b}, \mathbf{b}, -\mathbf{b}, -\mathbf{b}, \mathbf{b}, -\mathbf{b}, \mathbf{b}, -\mathbf{b}, -\mathbf{b}, -\mathbf{b}, -\mathbf{b}, -\mathbf{b}).$$

Затем последовательность \mathbf{z} посимвольно перемножается с каждой 16-й строкой построенной по алгоритму Сильвестра матрицы Адамара (см. пункт 2.7.3) размерности 256 для образования 16 ортогональных символов 256-ричного алфавита, которые используются в дальнейшем для построения кодовых слов SSC. Алгоритм кодирования SSC должен обеспечивать низкий уровень всех неправильных корреляций, т.е. между всеми 15 циклическими сдвигами любого одного слова, а также между любыми временными сдвигами различных слов. В спецификации UMTS все 64 кодовые слова SSC представлены в виде таблицы, однако простая проверка устанавливает, что эти слова взяты из (15,3) 256-ричного

кода Рида-Соломона. Минимальное хэммингово расстояние любого кода Рида-Соломона определено как на единицу большее числа проверочных символов [31,33], т.е. в нашем случае равно 13. Оно гарантирует, что возможно не более двух совпадений 256-ричных символов между любым SSC словом и любой из 14-ти циклически сдвинутых копий, разнесенных друг относительно друга на целое число слотов. Другими словами, нормированная периодическая АКФ любого SSC кодового слова при указанных сдвигах имеет уровень боковых лепестков не выше $2/15$. Аналогичное утверждение справедливо для любых двух кодовых слов, назначенных различным БС, при взаимных сдвигах, равных целому числу слотов. В то же время, поскольку БС сети не привязаны к общей шкале, передаваемые ими коды синхронизации скользят друг относительно друга и, значит, низкий уровень корреляции должен сохраняться при произвольном временном сдвиге, а не только равном целому числу слотов. Оптимальность расстояния кода Рида-Соломона не может обеспечить низкий уровень корреляции при произвольных сдвигах SSC последовательностей, и данная проблема заслуживает дальнейших исследований.

В качестве общего резюме к параграфам 11.3-11.4 снова подчеркнем, что данный краткий обзор 3G систем и стандартов касался только решений, затрагивающих содержание книги, т.е. касающихся практической реализации широкополосности и идей CDMA. Уже имеется большое количество книг и статей, еще значительно большее ожидается, которые объяснят и прояснят различные аспекты систем 3-го и дальнейших поколений, охватят все уровни системного протокола. В частности, заинтересованного читателя можно отослать к следующей библиографии [19,69,92,104,114,115,120-122] и непосредственно к 3G спецификациям.

Список литературы.

Глава 1.

1. Carlson, A. B. *Communication Systems*, McGraw-Hill, New York, 1986.
2. Sklar, B. *Digital Communications*, Prentice-Hall, Upper Saddle River, NJ, 2001.
3. Dixon, R. C. *Spread Spectrum Systems with Commercial Applications*, John Wiley & Sons, 1994.
4. Haykin, S. *Communication Systems*, John Wiley & Sons, 2001.
5. Ziemer, R. E., and Peterson, R. L. *Introduction to Digital Communication*, Prentice-Hall, Upper Saddle River, NJ, 2001.
6. Ziemer, R. E., Peterson, R. L., and Borth, D. E. *Introduction to Spread Spectrum Communications*, Prentice-Hall, Englewood Cliffs, NJ, 1995.
7. Proakis, J. G. *Digital Communications*, McGraw-Hill, New York, 2001.
8. Proakis, J., and Salehi, M. *Communication Systems Engineering*, Prentice-Hall, Upper Saddle River, NJ, 2002.
9. Simon M. K., Omura, J. K., Scholtz, R. A., and Levitt, B. K. *Spread Spectrum Communication Handbook*, McGraw-Hill, New York, 1994.
10. Scholtz, R. A. "The origins of spread spectrum communications", *IEEE Trans. Commun.*, vol. 30, pp. 822-854, May, 1982.

Глава 2.

11. Conway, J. H., and Sloane, N. J. A. *Sphere Packings, Lattices and Groups*, Springer-Verlag, New York, 1998.
12. Sloane, N. J. A. *Spherical codes*, Nice arrangements of points on a sphere in various dimensions, <http://www.research.att.com/~njas/packings/index.html>.

Глава 3.

13. Adams, R.A. *Calculus*, Addison Wesley Longman, 1999.
14. Leon-Garcia, A. *Probability and Random Processes for Electrical Engineering*, Addison Wesley, 1994.
15. Freeman, R. L. *Radio System Design for Telecommunications*, John Wiley & Sons, 1997.
16. Berg, O., Berg, T., Haavik, S., Hjelmstad, J. and Skaug, R. *Spread Spectrum in Mobile Communication*, IEE, London, 1998.
17. Lee, W. C. Y. *Mobile Communications Engineering*, McGraw-Hill, New York, 1997.
18. Lee, J. S. and Miller, L.E. *CDMA Systems Engineering Handbook*, Artech House, 1998.
19. Glisic, S. *Adaptive WCDMA: Theory and Practice*, John Wiley & Sons, 2003.

Глава 4.

20. Kim, K.I. "CDMA cellular engineering issues", *IEEE Trans. Veh. Tech.*, vol. 42, pp. 345-350, 1993.
21. Gradstein, I and Ryzhik, I. *Table of Integrals, Series, and Products*, Academic Press, 1980.

Глава 6.

22. Eliahou, S., and Kervaire, M. "Barker sequences and difference sets", *L'Enseignement Mathématique*, vol. 38, pp. 345-382, 1992.
23. Friese, M. "Polyphase Barker sequences up to length 36", *IEEE Trans. Inform. Theory*, vol. 42, pp. 1248-1250, 1996.
24. Brenner, A.F. "Polyphase Barker sequences up to length 45 with small alphabets", *Electron. Lett.*, vol. 34, pp. 1576-1577, 1998.

25. Lindner, J. "Binary sequences up to lengths 40 with best possible autocorrelation function", *Electron. Lett.*, vol. 11, p. 507, 1975.
26. Cohen, M., Fox, M.R., and Baden, J.M. "Minimum peak sidelobe compression codes", in *Proc. IEEE Int. Radar Conf.*, 1990, pp. 633-639.
27. Deng, X., and Fan, P. "New binary sequences with good aperiodic autocorrelation obtained by evolutionary algorithm", *IEEE Commun. Lett.*, vol. 3, pp. 288-290, 1999.
28. Baumert, L.D. *Cyclic Difference Sets*, Springer-Verlag, 1971.
29. Schmidt, B. "Cyclotomic integers and finite geometry", *J. Amer. Math. Soc.*, vol. 12, pp. 929-952, 1999.
30. Gilbert, W.J., and Nicholson, W.K. *Modern Algebra with Applications*, John Wiley & Sons, 2004.
31. Bossert, M. *Channel Coding for Telecommunications*, John Wiley and Sons, 1999.
32. Lidl, R., and Niederreiter, H. *Introduction to Finite Fields and Their Applications*, Cambridge: Cambridge University Press, 1994.
33. Blahut, R. E. *Algebraic Codes for Data Transmission*, Cambridge: Cambridge University Press, 2003.
34. Sverdlick, M. B. *Optimal Discrete Signals*, Moscow: Sov. Radio, 1975, (In Russian).
35. Lüke, H. D., Schotten, H.D., and Hadinejad-Mahram, H. "Binary and quadriphase sequences with optimal autocorrelation properties: a survey", *IEEE Trans. Inform. Theory*, vol. 49, pp. 3271-3282, 2003.
36. Lüke, H. D. *Korrelationsignale*, Berlin: Springer-Verlag, 1992 (in German).
37. Fan, P., and Darnell, M. *Sequence Design for Communication Applications*, John Wiley & Sons, 1996.
38. Amiantov, I. N. *Selected Issues of Statistical Communications Theory*, Moscow: Sov. Radio, 1971, (In Russian).
39. Levanon, N., and Mozeson, E. *Radar Signals*, John Wiley & Sons, 2004.
40. Ipatov, V. P. "Ternary sequences with ideal periodic autocorrelation properties", *Radio Eng. Elect. Physics*, vol. 24, no. 10, pp. 75-79, 1979.
41. Ipatov, V. P. "Contributory to the theory of ternary sequences with perfect periodic autocorrelation properties", *Radio Eng. Elect. Physics*, vol. 25, no. 4, pp. 31-34, 1980.
42. Zierler, N. "Linear recurring sequences", *J. Soc. Appl. Math.*, vol. 7, pp. 31-48, 1959.
43. Hoholdt, T., and Justesen, J. "Ternary sequences with perfect periodic autocorrelation", *IEEE Trans. Inform. Theory*, vol. 29, pp. 597-600, 1983.
44. Ackroyd, M. H., and Ghani, F. "Optimal mismatched filter for sidelobe suppression", *IEEE Trans. Aerosp. Electron. Syst.*, vol. 9, pp. 214-218, 1973.
45. Rihaczek, A. W., and Golden, R.M. "Range sidelobe suppression for Barker codes", *IEEE Trans. Aerosp. Electron. Syst.*, vol. 7, pp. 1087-1092, 1971.
46. Ipatov, V. P., "Total suppression of sidelobes of periodic correlation functions of phase manipulated signals", *Radio Eng. Elect. Physics*, vol. 22, no. 8, pp. 42-47, 1977.
47. Ipatov, V. P., "Choice of periodical PSK signal and filter combination", *Radioelectron. a. Commun. Syst. (Radioelektronika)*, vol. 21, no.4, pp. 49-55, 1978.
48. Ipatov, V. P., "Synthesis of a binary periodic signal-filter pair", *Radioelectron. a. Commun. Syst. (Radioelektronika)*, vol. 23, no.4, pp. 46-51, 1980.
49. Ipatov, V. P. "Binary periodic sequences with low sidelobe suppression loss", *Radioelectron. a. Commun. Syst. (Radioelektronika)*, vol. 23, no.1, pp. 15-19, 1980.
50. Ipatov, V. P., and Fedorov, B.V. "Regular binary sequences with small losses in suppressing sidelobes", *Radioelectron. a. Commun. Syst. (Radioelektronika)*, vol. 27, no.3, pp. 29-33, 1984.
51. Blokhuis, A., and Tiersma, H. J. "Bounds for the size of radar arrays", *IEEE Trans. Inform. Theory*, vol. 34, pp. 164-167, 1988.
52. Hamkins, J., and Zeger, K. "Improved bounds on maximum size binary radar array", *IEEE Trans. Inform. Theory*, vol. 43, pp. 997-1000, 1997.

53. Golomb, S. W., and Taylor, H. "Two-dimensional synchronization patterns for minimum ambiguity", *IEEE Trans. Inform. Theory*, vol. 28, pp. 600-604, 1982.
54. Golomb, S. W., and Taylor, H. "Construction and properties of Costas arrays", *Proc. IEEE*, vol. 72, pp. 1143-1163, 1984.

Глава 7.

55. Chatschik, B. "An overview of Bluetooth wireless technology", *IEEE Commun. Magazine*, vol. 39, pp. 86-94, Dec., 2001.
56. Ross, J. A. F., and Taylor, D. P. "Vector assignment scheme for $N+M$ users in N -dimensional global additive channel", *Electron. Lett.*, vol. 28, pp. 1634-1636, 1992.
57. Paavola, J., and Ipatov, V. P. "Binary CDMA signatures for $N+M$ users in N -dimensional global signal space", *Electron. Lett.*, vol. 39, pp. 738-740, 2003.
58. Djonin, D., and Bhargava V, "New results on low complexity detectors for oversaturated CDMA systems", in *Proc. Globecom 2001*, San Antonio, USA, November 2001, pp.846-850.
59. Paavola, J., and Ipatov, V. P. "Oversaturating synchronous CDMA systems on the signature per user basis". *Proc. 5th European Mobile Communications Conf*, Glasgow, Scotland, UK, 2003, pp.427-430.
60. Learned, R. E., Willsky, A. S., and Boroson, D.M. "Low complexity optimal joint detection for oversaturated multiple access communications", *IEEE Trans. Signal Processing*, vol. 45, pp. 113-123, 1997.
61. Shi, Z., and Schlegel, C. "Spreading code construction for CDMA", *IEEE Commun. Lett.*, vol. 7, pp. 4-6, 2003.
62. Welch, L. R. "Lower bound on the maximum cross-correlation of signals", *IEEE Trans. Inform. Theory*, vol. 20, pp. 397-399, 1974.
63. Rupp, M., and Massey, J. L. "Optimum sequence multisets for synchronous code-division multiple-access channels", *IEEE Trans. Inform. Theory*, vol. 40, pp. 1261-1266, 1994.
64. Karystinos, G. N., and Pados, D. A. "New bounds on the total squared correlation and optimum design of DS-CDMA binary signature sets", *IEEE Trans. Commun.*, vol. 51, pp. 48-51, 2003.
65. Ipatov, V. P. "On the Karystinos-Pados bounds and optimal binary DS-CDMA signature ensembles", *IEEE Commun. Lett.*, vol. 8, pp. 81-83, 2004.
66. Stark, H, and Woods, J. W. *Probability and Random processes with Applications to Signal Processing*, Prentice-Hall, Upper Saddle River, NJ, 2002.
67. Sarwate, D. V., and Pursley, M. B. "Crosscorrelation properties of pseudorandom and related sequences", *Proc. IEEE*, vol. 68, pp. 593-619, 1980.
68. Sidelnikov, V.M. "On the mutual correlation of sequences", *Soviet Math. Dokl.*, vol.12, pp. 197-201, 1971.
69. Korovajczuk L., et al. *Designing cdma2000 Systems*, John Wiley & Sons, 2004.
70. Ipatov, V. P. *Periodic Discrete Signals with Optimal Correlation Properties*, Moscow: Radio I Sviaz, 1992 (in Russian).
71. Goldberg, B.-G. "Code division multiplexing by frequency shifted biphas modulated M-sequences", *IEEE Trans. Aerosp. Electron. Syst.*, vol. 17, 303-304, 1981.
72. Gold, R. Optimal binary sequences for spread spectrum multiplexing, *IEEE Trans. Inform. Theory*, vol. 13, pp. 619-621, 1967.
73. Kasami, T. "Weight distribution formula for some class of cyclic codes", Coordinated Science Lab., Univ. Illinois, Urbana, Tech. Rep. R-285, April 1966.
74. Kamaletdinov, B. Zh. "An optimal ensemble of binary sequences based on the union of the ensembles of Kasami and bent-function sequences", *Problems of Inform. Transmission*, vol. 24, pp. 167-169, 1988.
75. Olzen, J. D., Scholtz, R. A., and Welch, L.R. "Bent-function sequences", *IEEE Trans. Inform. Theory*, vol. 28, pp. 858-864, 1982.

76. Kamaletdinov, B. Zh. "Optimal sets of binary sequences", *Problems of Inform. Transmission*, vol. 32, pp. 171-175, 1996.

Глава 8.

77. Glisic, S., and Vucetic, B. *Spread Spectrum CDMA Systems for Wireless Communications*, Artech House, 1997.

78. Ipatov V. P., Kazarinov, Yu. M. (Editor), Kolomensky, Yu. A., Uljanitsky, Yu. D. *Acquisition, Detection and Parameter Measuring of Signals in Radionavigation Systems*, Moscow: Sov. Radio, 1975 (In Russian).

79. Holmes, J. K., and Chen, C. C. "Acquisition time performance of PN spread spectrum systems", *IEEE Trans. Commun.*, vol. 25, pp. 778-784, 1977.

80. Polydoros, A., and Weber, C. L. "A unified approach to serial search spread-spectrum code acquisition – Part I: General theory", *IEEE Trans. Commun.*, vol. 32, pp. 542-549, 1984.

81. Polydoros, A., and Weber, C. L. "A unified approach to serial search spread-spectrum code acquisition – Part II: A matched filter receiver", *IEEE Trans. Commun.*, vol. 32, pp. 550-560, 1984.

82. Di Carlo, D.M., and Weber, C. L. "Multiple dwell serial search: Performance and application in direct sequence code acquisition", *IEEE Trans. Commun.*, vol. 31, pp. 650-659, 1983.

83. Viterbi, A. J. *CDMA: Principles of Spread Spectrum Communication*, Addison-Wesley, 1995.

84. Zigangirov, K. Sh. *Theory of Code Division Multiple Access Communication*, John Wiley & Sons, 2004.

85. Wald, A. *Sequential Analysis*, New York: Wiley, 1947.

86. Stiffler, J. J. *Theory of Synchronous Communications*, Prentice-Hall, Englewood Cliffs, NJ, 1971.

87. Golomb, S.W. et al, *Digital Communications with Space Applications*, Prentice-Hall, Englewood Cliffs, NJ, 1964.

88. Van Trees, H. L. *Detection, Estimation and Modulation Theory, Part II, Nonlinear Modulation Theory*, John Wiley & Sons, 2002.

89. Lindsey, W. C. *Synchronization Systems in Communication and Control*, Prentice-Hall, Englewood Cliffs, NJ, 1972.

90. Viterbi, A. J. *Principles of Coherent Communications*, McGraw-Hill, New York, 1966.

Глава 9.

91. Clark G. C., and Cain, J. B. *Error-Correcting Coding for Digital Communications*, New York: Plenum, 1988.

92. Castro, J. P., *The UMTS Network and Radio Access Technology. An Interface Techniques for Future Mobile Systems*, John Wiley & Sons, 2001.

93. Viterbi, A. J. and Omura, J. K. *Principles of Digital Communications and Coding*, McGraw-Hill, New York, 1979.

94. Heller, J. A., and Jacobs, I. W. "Viterbi decoding for satellite and space communications", *IEEE Trans. Commun. Technol.*, vol. 19, pp. 835-848, 1971.

95. Berrou, C., Glavieux, A., and Thitimajshima, "Near Shannon limit error-correcting coding and decoding: turbo codes", in *Proc. IEEE Int. Commun. Conf., ICC'93*, Geneva, Switzerland, 1993, pp. 1064-1070.

96. Berrou, C., and Glavieux, A. "Near optimum error correcting coding and decoding: turbo codes", *IEEE Trans. Commun.*, vol. 44, pp. 1261-1271, 1996.

97. Schlegel, C.B., and Pérez, L.C. *Trellis and Turbo Coding*, IEEE Press, 2004.

98. Bahl, L. R., Cocke, J., Jelinek, F, and Raviv, J. "Optimal decoding of linear codes for minimizing symbol error rate", *IEEE Trans. Inform. Theory*, vol. 20, pp. 284-287, 1974.

Глава 10.

99. Madhow, U., and Honig, M. L. "MMSE interference suppression for direct-sequence spread spectrum CDMA", *IEEE Trans. Commun.*, vol. 42, pp. 3178-3188, 1994.
100. Wang, X., and Poor, V. H. *Wireless Communication Systems. Advanced Techniques for Signal Reception*, Prentice-Hall, Upper Saddle River, NJ, 2004.
101. Honig, M., and Tatsanis, M. K. "Adaptive techniques for multiuser CDMA receivers". *IEEE Signal Process. Magazine*, vol. 17, pp. 49-61, May 2000.
102. Verdú, S. *Multiuser Detection*, Cambridge: Cambridge University Press, 1998.
103. Lu, W. W. (Editor). *Broadband Wireless Mobile. 3G and Beyond*, John Wiley & Sons, 2002.
104. Tanner, R., and Woodward, J. *WCDMA – Requirements and Practical Design*, John Wiley & Sons, 2004.
105. Fazel, K., and Kaiser, S. *Multi-Carrier and Spread Spectrum Systems*, John Wiley & Sons, 2003.
106. Hara, S., and Prasad, R. "Overview of multicarrier CDMA", *IEEE Commun. Magazine*, vol. 35, pp. 126-133, Dec., 1997.
107. Popović, B. M. "Spreading sequences for multicarrier CDMA systems", *IEEE Trans. Commun.*, vol. 47, pp. 918-926, 1999.
108. Telatar, E. "Capacity of multi-antenna Gaussian channels", *European Transactions on Telecommunications*, vol. 10, pp. 585-595, 1999.
109. Foschini, G.J., and Gans, M.J. "On limits of wireless communications in a fading environment when using multiple antennas", *Wireless Personal Communications*, vol. 6, pp. 311-335, 1998.
110. Vucetic, B., and Yuan, J. *Space-Time Coding*, John Wiley & Sons, 2003.
111. Alamouti, S. M. "A simple transmit diversity technique for wireless communications", *IEEE Journal Select. Areas. Commun.*, vol. 16, pp. 1451-1458, 1998.
112. Tarokn, V., Jafarkhani, H., and Calderbank, A. R. "Space-time block codes from orthogonal designs", *IEEE Trans. Inform. Theory*, vol. 45, pp. 1456-1467, 1999.
113. Tarokn, V., Naguib, A. Seshardi, N., and Calderbank, A. R. "Space-time codes for high data rate wireless communication: performance criteria in the presence of channel estimation errors, mobility, and multiple paths", *IEEE Trans. Commun.*, vol. 47, pp. 199-207, 1996.
114. Holma, A., and Toskala, A. (Editors). *WCDMA for UMTS. Radio Access for Third Generation Mobile Communications*, John Wiley & Sons, 2001.
115. Korhonen, J. *Introduction to 3G Mobile Communications*, Artech House, 2001.
116. Hockwald, B., Marzetta, T. L., and Papadias, C.B. "A transmitter diversity scheme for wideband CDMA systems based on space-time spreading", *IEEE Journal Select. Areas. Commun.*, vol. 19, pp. 48-60, 2001.

Глава 11.

117. Hoffman-Wellenhoff, B., Lichtenegger, H., and Collins, J. *Global Positioning System: Theory and Practice*, Springer-Verlag, New York, Wien, 2001.
118. El-Rabbani, A. *Introduction to GPS: the Global Positioning System*, Artech House, 2002.
119. Farrel, J., and Barth, M. *The Global Positioning System and Inertial Navigation*, McGraw-Hill, New York, 1999.
120. Steele, R., Lee, Ch., and Gould, P. *GSM, cdmaOne and 3G Systems*, John Wiley & Sons, 2001.
121. Karim, M.R., and Sarraf, R. *W-CDMA and cdma2000 for 3G Mobile Networks*, McGraw-Hill, New York, 2002.
122. Walke, B., Seidenberg, P., and Althoff, M.P. *UMTS: The Fundamentals*, John Wiley & Sons, 2003.

Список обозначений.

- A_r – effective area of the receiving antenna (эффективная площадь приемной антенны),
 A – signal amplitude (амплитуда сигнала),
 \mathbf{A} – signature matrix (матрица сигнатур),
 $\mathbf{a} = (a_0, a_1, \dots, a_{N-1})$ – vector of complex amplitude of APSK signal (вектор комплексных амплитуд APSK сигнала),
 $a_{k,i}$ – i -th symbol of code sequence of k -th APSK signal (i -й символ кодовой последовательности k -го APSK сигнала),
 B – length of error burst (длина пакета ошибок),
 B_c – channel coherence bandwidth (канальная полоса когерентности),
 $\dot{B}(t)$ – data stream (поток данных),
 $b_{k,i}$ – i -th information symbol of k -th user (i -й информационный символ k -го пользователя),
 \mathbf{b}_k – data sequence of k -th user (последовательность данных k -го пользователя),
 \mathbf{C} – signature correlation matrix (корреляционная матрица сигнатур),
 D – space distance between antennas (расстояние в пространстве между антеннами),
 d – decimation index (индекс децимации),
 $d(\mathbf{s}_0, \mathbf{s}_1)$ – distance (Euclidean) between signal vectors (расстояние (Евклидово) между сигнальными векторами),
 $d_H(\mathbf{u}, \mathbf{v})$ – Hamming distance between vectors (расстояние Хэмминга между векторами),
 d_H – code distance (кодированное расстояние),
 d_i – element of linear sequence on a finite field (элемент линейной последовательности над конечным полем),
 E – signal energy (энергия сигнала),
 $e(\tau)$ – error signal of discriminator (сигнал ошибки дискриминатора),
 F – frequency shift (частотный сдвиг),
 F_e – envelope frequency spread (протяженность по частоте огибающей),
 G – coding gain (выигрыш от кодирования),
 G_a – asymptotic coding gain (асимптотический выигрыш от кодирования),
 G_d – diversity gain (выигрыш от разнесения, коэффициент разнесения),
 G_t, G_r – power gains of transmitting and receiving antennas (коэффициенты усиления по мощности передающей и приемной антенн),
 \mathbf{G} – matrix of user's amplitudes (матрица амплитуд пользователей),
 $GF(p)$ – finite field of order p (конечное поле порядка p),
 $h(t)$ – filter pulse response (импульсный отклик фильтра),
 H_i, \hat{H}_j – hypothesis and decision (гипотеза и решение),
 \dot{H}_{ij} – channel fading coefficient (канальный коэффициент замираний),
 I – length of interleaver (длина перемежения),
 L – length of sequence (длина последовательности),
 L_0, L_1, \dots – number of elements 0, 1, ... on the period of m -sequence (число элементов 0, 1, ... на периоде m -последовательности),
 M – number of signals, alphabet size (число сигналов, объем алфавита),
 M_c – number of carriers in multicarrier system (число несущих в системе со многими несущими),
 m – number of encoded bits in orthogonal signaling, sequence shift (число кодированных бит

при ортогональной сигнализации, сдвиг последовательности),
 N – discrete signal length (processing gain) (длина дискретного сигнала (выигрыш от обработки)),
 N_I, N_J – extra power spectrum density of interference or jammer (дополнительная спектральная плотность мощности помехи),
 N_S – spectrum density of signal (спектральная плотность сигнала),
 N_0 – one side spectral density of white noise (односторонняя спектральная плотность белого шума),
 N_+, N_-, N_p – numbers of plus ones, minus ones and pauses (zeros) in a binary or ternary sequence (число плюс единиц, минус единиц и пауз (нулей) в бинарной и троичной последовательности),
 n – sequence memory (память последовательности),
 $n(t)$ – normal (Gaussian) noise (нормальный (гауссовский) шум),
 (n, k) – length and number of bits of a linear code (длина и число бит (информационных символов) линейного кода),
 n_d – number of diversity branches (число ветвей разнесения),
 n_{\min} – number of signal pairs with the minimum distance (число сигнальных пар, находящихся на минимальном расстоянии),
 n_s – signal space dimension (размерность сигнального пространства),
 n_T, n_R – number of transmit and receive MIMO antennas (число передающих и приемных антенн в схеме MIMO),
 P_b – bit error probability (вероятность ошибки на бит),
 P_e – complete (unconditional) error probability (полная (безусловная) вероятность ошибки),
 P – signal power (мощность сигнала),
 p_f, p_d – probabilities of false alarm and detection (вероятности ложной тревоги и правильного обнаружения),
 p – order of Galois field (порядок поля Галуа),
 $p[y(t)|s(t)]$ – channel transition probability, likelihood function (канальная переходная вероятность, функция правдоподобия),
 p_{ij} – probability of mistaking the i -th signal for the j -th one (вероятность перепутывания i -го сигнала с j -м),
 $Q(\cdot)$ – complementary error function (дополнительная функция ошибок),
 q – signal-to-noise ratio (отношение сигнал-шум),
 q_b – signal-to-noise ratio per bit (отношение сигнал-шум на бит),
 q_I – signal-to-interference ratio (отношение сигнал-помеха),
 q_J – signal-to-jammer ratio (отношение сигнал-помеха),
 q_r – resultant SNR (результатирующее SNR),
 R – data rate (скорость поступления данных),
 R_c – code rate (скорость кода),
 $R(\tau)$ – autocorrelation function of the signal (автокорреляционная функция сигнала),
 $\dot{R}(\tau)$ – autocorrelation function of complex envelope (автокорреляционная функция комплексной огибающей),
 $R(m), R_a(m), R_p(m)$ – autocorrelation functions of code sequence (автокорреляционная функция кодовой последовательности),
 $R_{kl}(m)$ – cross-correlation function of k -th and l -th code sequences (взаимно корреляционная

функция k -й и l -й кодовых последовательностей),
 $r(t)$ – filter response (отклик фильтра),
 S – signal set (множество сигналов),
 $S(t)$ – signal real envelope (вещественная огибающая сигнала),
 $S_I(t), S_Q(t)$ – signal quadrature components (квадратурные компоненты сигнала),
 S_d – discriminator slope (крутизна дискриминатора),
 $\dot{S}(t)$ – signal complex envelope (комплексная огибающая сигнала),
 $\dot{S}_{ff}(t)$ – signal complex envelope at the filter output (комплексная огибающая сигнала на выходе фильтра),
 SP – signal plane (сигнальная плоскость),
 $s(t)$ – signal (сигнал),
 $s_k(t)$ – k -th signature (k -я сигнатура),
 $s_{\perp}(t)$ – Hilbert transform of $s(t)$ (преобразование Гильберта от сигнала $s(t)$),
 \mathbf{s} – signal vector (сигнальный вектор),
 TSC – total squared correlation (полная сумма квадратов корреляций),
 T – signal duration (длительность сигнала),
 T_b – data bit duration (длительность бита данных),
 T_d – dwell time per search cell (время анализа на ячейку поиска),
 T_{ob} – observation interval (интервал наблюдения),
 T_p – symbol duration (длительность символа),
 T_s – sampling interval (период взятия отсчетов),
 T_t – total system time resource (общий временной ресурс системы),
 t_d, t_c – number of detected and corrected errors (число обнаруживаемых и исправляемых ошибок),
 U – code (код),
 u_d – output detector voltage (напряжение на выходе детектора),
 $u(z)$ – code polynomial (кодировый полином),
 $\mathbf{u} = (u_0, u_1, \dots, u_{n-1})$ – codeword (code vector) (кодировое слово (кодировый вектор)),
 W – signal bandwidth (ширина полосы сигнала),
 W_t – total bandwidth allocated to the system (общая полоса, отводимая системе),
 w – hypothesis a priori probability (априорная вероятность),
 w_i – weights of combining (веса комбинирования),
 \bar{x} – expectation of x (математическое ожидание x),
 $\dot{Y}(t)$ – complex envelope of observation (комплексная огибающая наблюдения),
 $y(t)$ – observation (наблюдение),
 \mathbf{y} – observation vector (вектор наблюдения),
 z – correlation (корреляция),
 $\mathbf{z} = (z_1, z_2, \dots, z_K)$ – vector of correlations (вектор корреляций),
 α – primitive element of finite field (примитивный элемент конечного поля),
 $\gamma(t)$ – angle modulation law (закон угловой модуляции),
 γ – energy loss in SLSF (энергетические потери в SLSF),
 δ – distance attenuation exponent (экспонента ослабления от расстояния),
 Δ – chip repetition interval (период повторения чипов (элементарных символов)),
 Δ_c – chip duration (длительность чипа),

ε – error of parameter estimation (ошибка оценки параметра),
 λ – estimated parameter (оцениваемый параметр),
 λ_w – wavelength (длина волны),
 v – peak-factor (пик-фактор),
 θ – angle polar coordinate (угол в полярных координатах),
 θ_n – noise temperature (шумовая температура),
 ρ_{ij} – correlation coefficient of i -th and j -th signals (коэффициент корреляции i -го и j -го сигналов),
 $\dot{\rho}_{ij}$ – correlation coefficient of i -th and j -th complex envelopes (коэффициент корреляции i -й и j -й комплексных огибающих),
 $\rho(\lambda)$ – correlation coefficient in dependence on parameter λ (коэффициент корреляции в зависимости от параметра λ)
 $\rho(\tau)$ – normalized ACF (нормированная АКФ),
 $\rho_0(\tau)$ – normalized envelope of a bandpass signal ACF (нормированная АКФ огибающей полосного сигнала);
 $\rho(m)$ – normalized ACF of code sequence (нормированная АКФ кодовой последовательности),
 $\rho_a(m), \rho_p(m)$ – normalized aperiodic and periodic ACF (нормированная аperiodическая и периодическая АКФ),
 $\dot{\rho}_c(\tau)$ – ACF of one chip (АКФ одиночного чипа),
 $\rho_{a,kl}(m), \rho_{p,kl}(m)$ – normalized aperiodic and periodic CCF (нормированная аperiodическая и периодическая ВКФ),
 $\sigma^2 = D\{x\}$ – variance of x (дисперсия x),
 τ – time delay (задержка во времени),
 τ_c – correlation spread (время корреляции),
 τ_{ds} – delay spread (задержка распространения),
 τ_{\min} – minimum delay between paths (минимальная задержка между путями распространения),
 ζ – alternative symbol for a correlation (альтернативный символ для обозначения корреляции),
 (\mathbf{u}, \mathbf{v}) – inner product of vectors (скалярное произведение векторов),
 $\|\mathbf{s}\|$ – norm of vector \mathbf{s} (норма вектора \mathbf{s}).

Список сокращений.

ACF – autocorrelation function (автокорреляционная функция (АКФ)),
AF – ambiguity function (функция неопределенности (ФН)),
APSK – amplitude-phase shift keying (амплитудно–фазовая модуляция (АФМ)),
ASK – amplitude shift keying (амплитудная модуляция (АМ)),
AWGN – additive Gaussian white noise (аддитивный белый гауссовский шум (АБГШ)),
BCH – broadcast channel (вещательный канал),
BPSK – binary phase shift keying (бинарная фазовая модуляция (БФМ)),
BS – base station (базовая станция (БС)),
BSC – binary symmetric channel (двоичного симметричного канала (ДСК)),
CCF – cross-correlation function (взаимная корреляционная функция (ВКФ)),
CDMA – code division multiple access (множественный доступ с кодовым разделением (МДКР)),
CPCH – common packet channel (общий канал пакетирования),
CPICH – common pilot channel (общий пилотный канал),
CRC – cyclic redundancy code (циклический избыточный код),
CW – continuous wave (непрерывное колебание),
DDFT – direct DFT (прямое ДПФ (ПДПФ)),
DFT – discrete Fourier transform (дискретное преобразование Фурье (ДПФ)),
DLL – delay-lock loop (петля захвата по задержки)
DPCCH – dedicated physical control channel (выделенный физический канал управления),
DPSCH – dedicated physical channel (выделенный физический канал),
DPDCH – dedicated physical data channel (выделенный физический канал данных),
DS – direct sequence (прямая последовательность),
EMC – electromagnetic compatibility (электромагнитная совместимость (ЭМС)),
FACH – forward access channel (прямой канал доступа),
FDD – frequency division duplex (дуплекс с частотным разделением),
FDMA – frequency division multiple access (множественный доступ с частотным разделением (МДЧР)),
FH – frequency hopping (прыгающая частота),
FFH – fast FH (быстрая FH),
FIR – finite impulse response (фильтр с конечным импульсным откликом (КИО)),
FSK – frequency shift keying (частотная модуляция (ЧМ)),
GNSS – global navigation satellite system (глобальная навигационная спутниковая система (ГНСС)),
GPS – global positioning system (система глобального позиционирования),
IDFT – inverse DFT (обратного ДПФ (ОДПФ)),
IIR – infinite impulse response (фильтр с бесконечным импульсным откликом (БИО)),
ISI – intersymbol interference (межсимвольная интерференция (МСИ)),
LFM – linear frequency modulation (линейная частотная модуляция (ЛЧМ)),
LFSR – linear feedback shift register (регистр сдвига с линейной обратной связью),
LOS – line-of-site (линия прямой видимости),
MAI – multiple-access interference (помеха множественного доступа),
MAP – maximum a posteriori probability (максимум апостериорной вероятности (МАП)),
MC – multicarrier modulation (модуляция с многими несущими);
MC-CDMA – multicarrier CDMA (CDMA со многими несущими),
MC-DS-CDMA – multicarrier direct sequence CDMA (CDMA со многими несущими и прямым расширением спектра),
MF – matched filter (согласованный фильтр (СФ)),
MIMO – multiple-input multiple out system (системы с множественным входом – множественным выходом),

ML – maximum likelihood (максимальное правдоподобие (МП)),
MMSE – minimum mean-square error (минимум среднеквадратической ошибки (МСКО)),
MS – mobile station (мобильная станция (МС)),
OFDM – orthogonal frequency division multiplexing (мультиплексирование с ортогональным частотным разделением),
OVSF – orthogonal variable spreading factor (codes) (ортогональные коды с переменной величиной расширения),
PAM – pulse amplitude modulation (импульсно-кодовая модуляция (ИКМ)),
PCG – power control group (группа регулировки мощности),
PCH – paging channel (канал вызова),
PCPCH – physical common packet channel (физический общий канал пакетирования),
PDF – probability density function (плотность распределения вероятности),
PRACH – physical random access channel (физический канал случайного доступа),
PSC – primary synchronization code (первичный код синхронизации),
PSK – phase shift keying (фазовая модуляция (ФМ)),
RACH – random access channel (канал случайного доступа),
QAM – quadrature amplitude modulation (квадратурная амплитудная модуляция (КАМ)),
QPSK – quadrature PSK (квадратурная ФМ (КФМ)),
SCH – synchronization channel (канал синхронизации),
SLSF – sidelobe suppression filter (фильтр подавления боковых лепестков (ФПБЛ)),
SINR – signal to interference plus noise ratio (отношением сигнал-(помеха плюс шум)),
SIR – signal-to-interference ratio (отношение сигнал-помеха),
SNR – signal-to-noise ratio (отношение сигнал-шум),
SSC – secondary synchronization code (вторичный код синхронизации),
TDD – time division duplex (дуплекс с временным разделением),
TDMA – time division multiple access (множественный доступ с временным разделением (МДВР)),
TSC – total squared correlation (полный квадрат корреляции),
UMTS – universal mobile telecommunication system (система всемирной мобильной связи),
VCO – voltage controlled oscillator (генератор управляемый напряжением (ГУН)).