

Министерство образования и науки Российской Федерации

САНКТ–ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ
ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

**Приоритетный национальный проект «Образование»
Национальный исследовательский университет**

*Н.В. Богач, Г.Б. Гублер, В.Е. Евдокимов, З.В. Куляшова,
С.А. Перепелица, Т.З. Хабурзания*

ОБРАБОТКА СИГНАЛОВ В ИНФОРМАЦИОННЫХ СИСТЕМАХ

*Рекомендовано Учебно-методическим объединением по
университетскому политехническому образованию в качестве
учебного пособия для студентов высших учебных заведений,
обучающихся по направлениям подготовки и специальностям
техники и технологии*

Санкт-Петербург
Издательство Политехнического университета
2010

УДК 621.391.26(075)
ББК 32.811.3я7
О-23

Рецензенты:

Рецензенты:

Доктор технических наук,

Г. А. Кондрашкова

Кандидат технических наук, доцент

Санкт-Петербургского государственного политехнического
университета С. Э. Сараджишвили

Авторы:

*Богач Н.В., Гублер Г.Б., Евдокимов В.Е., Куляшова З.В.,
Перепелица С.А., Хабурзания Т.З..*

Обработка сигналов в информационных системах: учеб. пособие /
Н.В. Богач [и др.]. – СПб.: Изд-во Политехн. ун-та, 2010. – 222 с.

ISBN 978-5-7422-1845-6

Рассматриваются задачи обработки сигналов, необходимые для проектирования информационных системах. Предполагается наличие у студентов базовой подготовки по упомянутым разделам, так как рассматриваемые задачи выходят за пределы одной учебной дисциплины. Авторы стремятся продемонстрировать тесные междисциплинарные связи внутри реальных задач обработки сигналов. Мы надеемся тем самым мотивировать у студентов, проходящих подготовку по магистерским программам на старших курсах, обращать внимание на необходимость комплексного изучения специальных предметов. Наряду с общетеоретическими сведениями в пособии представлены результаты исследований авторов по таким важным вопросам как современные методы аналого-цифрового преобразования, дискретное вейвлет-преобразования, фазовая автоподстройка частоты, обратные задачи при измерениях и управлении.

Предназначено для студентов, обучающихся по направлению 220100 «Системный анализ и управление». Также может быть использовано специалистами в области разработки и применения различных информационных систем.

Работа выполнена в рамках реализации программы развития национального исследовательского университета «Модернизация и развитие политехнического университета как университета нового типа, интегрирующего мультидисциплинарные научные исследования и надотраслевые технологии мирового уровня с целью повышения конкурентоспособности национальной экономики»

Печатается по решению редакционно-издательского совета
Санкт-Петербургского государственного политехнического университета.

ISBN 978-5-7422-2987-2

© Санкт-Петербургский государственный
политехнический университет, 2010

СОДЕРЖАНИЕ

Список принятых сокращений.....	7
Предисловие.....	8
Введение.....	10
1. Представление сигналов в информационных системах.....	13
1.1. Дискретный сигнал	14
1.1.1. Цифровой единичный импульс	14
1.1.2. Цифровой единичный скачок	16
1.1.3. Цифровая синусоидальная последовательность.....	17
1.1.4. Цифровая экспоненциальная последовательность.....	18
1.1.5. Периодичность последовательностей.....	18
1.2. Дискретизации по времени	18
2. Линейные системы	20
2.1. Определение линейной системы	20
2.2. Свойство инвариантности во времени.....	20
2.3. Импульсная характеристика	21
2.4. Физически реализуемая система	23
2.5. Устойчивость линейной системы.....	23
2.6. Примеры анализа линейных систем на устойчивость и физическую реализуемость	24
2.7. Системы с коенчной и бесконечной импульсными характеристиками.....	25
2.8. Каскадное соединение линейных систем инвариантных времени .	26
2.9. Описание линейных систем при помощи разностных уравнений..	27
3. Методы частотно-временного анализа сигналов. Разложение сигналов в базисе	30
3.1. Общие положения	30
3.2. Описание линейных систем в частотной области. Преобразование Фурье	31
3.2.1. Свёртка как матрично-векторная операция	31
3.2.2. Собственные функции линейного оператора и линейной системы	32
3.2.3. Синусоидальные функции для описания линейных систем.....	33

3.2.4. Комплексная экспонента – собственная функция линейных систем	34
3.2.5. Пример вычисления частотных характеристик систем	35
3.2.6. Представление сигналов при помощи преобразования Фурье	36
3.3. Свойства преобразования Фурье	37
3.3.1. Условия существования преобразования Фурье цифровой последовательности	37
3.3.2. Представление цифровой последовательности как суммы чётной и нечётной последовательностей.....	37
3.3.3. Свойства преобразования Фурье, основанные на симметрии..	38
3.3.4. Теоремы о преобразовании Фурье дискретных сигналов	40
3.4. Кратковременное преобразование Фурье.....	45
3.5. Разложение в биортогональном базисе	46
3.6. Использование преобразования Фурье для анализа процессов аналого-цифрового и цифро-аналогового преобразований. Теорема Шеннона-Котельникова-Найквиста	48
3.6.1. Описание процесса дискретизации сигнала в частотной области при помощи преобразования Фурье	48
3.6.2. Восстановление непрерывного сигнала	50
3.6.3. Спектр цифровой последовательности.....	53
3.7. Z-преобразование	54
3.7.1. Ограничения на использование преобразования Фурье для анализа систем и сигналов	54
3.7.2. Затухающие комплексные экспоненты – собственные функции линейных систем	55
3.7.3. Z-преобразование	56
3.7.4. Обратное Z-преобразование	61
3.7.5. Свойства Z-преобразования	62
3.8. Анализ линейных систем	63
3.8.1. Свойства Z-преобразования односторонних и двусторонних последовательностей	63
3.8.2. Линейные системы, описываемые линейным разностным уравнением.....	66
3.8.3. Устойчивость и физическая реализуемость	68
3.8.4. Простейшее звено с одним "нулём"	69

3.8.5. Все пропускающие системы	70
3.8.6. Связь амплитудно-частотной и фазо-частотной характеристик	72
3.8.7. Системы с минимальной фазовой задержкой	74
3.9. Современные методы частотно-временного анализа	78
3.9.1. Разложение в вейвлет-базисе	80
3.9.2. Вейвлет-преобразование как средство частотно-временного анализа	84
3.9.3. Обобщенный алгоритм частотно-временного вейвлет-анализа	87
3.9.4. Базисные функции вейвлет-преобразования	88
4. Операции преобразования сигналов	91
4.1. Фильтрация	91
4.1.1. Проектирование фильтров	91
4.1.2. Исследование частотных характеристик банка квадратурно-зеркальных фильтров	99
4.2. Модуляция	110
4.2.1. Модуляция аналоговых сигналов	110
4.2.2. Методы демодуляции	125
4.2.3. Фазовая автоподстройка частоты	129
4.2.4. Модуляция (манипуляция) цифровых сигналов	133
4.3. Технологии расширения спектра	146
4.3.1. Метод частотных скачков	147
4.3.2. Метод прямой последовательности	148
4.3.3. Сравнение методов частотных скачков и прямой последовательности	148
4.3.4. Ортогональное частотное разделение	149
5. Обзор современных методов аналого-цифрового преобразования	151
5.1. Классификация устройств аналого-цифрового преобразования по фактору точность/быстродействие	151
5.2. Субполосные АЦП с временным разделением сигнала	152
5.3. Улучшение точностных характеристик быстрых АЦП. Субполосные АЦП с частотным разделением сигнала	155
6. Построение узла АЦП на основе сигма-дельта преобразования	158
6.1. Применение передискретизации при аналого-цифровом преобразовании	160

6.2. Формирование спектра шума квантования с помощью обратной связи.....	162
6.3. Формирование спектра шума квантования с помощью усредняющих окон.....	165
7. Дискретное вейвлет-преобразование	167
7.1. Быстрый алгоритм дискретного вейвлет-преобразования	168
7.2. Средства обработки сигналов, основанные на вейвлет-преобразовании	174
7.2.1. Применение вейвлет-преобразования для сжатия сигналов ..	174
7.2.2. Удаление шумов при вейвлет-преобразовании	177
8. Обратные задачи при измерениях и управлении	182
8.1. Общая часть	182
8.2. Формулировка метода регуляризации А.Н. Тихонова в пространстве Фурье-образов.....	188
8.3. Принцип минимального модуля.....	192
8.4. Нейросетевой подход решения обратной задачи	197
8.5. Модифицированный оптимальный фильтр Винера дает регуляризованное решение	202
8.6. Возможности построения физически реализуемого обратного фильтра.....	208
8.7. Дробно-рациональная аппроксимация и факторизация	211
8.8. Пример цифрового моделирования для нормирования погрешностей.....	212
Библиографический список.....	220

СПИСОК ПРИНЯТЫХ СОКРАЩЕНИЙ

АМ – амплитудная модуляция;
АМП – амплитудная манипуляция;
АМ-ПН – амплитудная модуляция с подавлением частоты;
АЦП – аналого-цифровое преобразование;
АЧХ – амплитудно-частотная характеристика;
БИХ – бесконечная импульсная характеристика;
ВП – вейвлет-преобразование;
ДВП – дискретное вейвлет-преобразование;
ИХ – импульсная характеристика;
КИХ – конечная импульсная характеристика;
КЧХ – комплексная импульсная характеристика;
ЛС – линейная система;
ЛЧМ – линейная частотная модуляция;
ОПФ – обратное преобразование Фурье;
ППФ – прямое преобразование Фурье;
ПФ – преобразование Фурье;
ФВЧ – фильтр высоких частот;
ФМ – фазовая модуляция;
ФНЧ – фильтр низких частот;
ФЧХ – фазо- частотная характеристика;
ЦАП – цифро-аналоговое преобразование;
ЦОС – цифровая обработка сигналов;
ЧИМ – частотно-импульсная модуляция;
ЧМ – частотная модуляция;
ЧХ – частотная характеристика.

ПРЕДИСЛОВИЕ

В пособии описываются задачи обработки сигналов, возникающие в информационных системах. В настоящее время опубликовано достаточное количество учебной и научной литературы, посвященной различным разделам данной теме [4, 10, 11]. Задачи обработки сигналов выходят за пределы одной учебной дисциплины, что предполагает необходимость наличия базовой подготовки у студента по основным разделам цифровой обработки сигналов. Авторы стремятся продемонстрировать тесные междисциплинарные связи внутри реальных задач обработки сигналов. Мы надеемся тем самым мотивировать студентов, проходящих подготовку по магистерским программам на старших курсах, обращать внимание на необходимость комплексного изучения специальных предметов. Это, в свою очередь, будет способствовать формированию системного подхода к задачам обработки сигналов у будущих выпускников, расширению их кругозора, и созданию прочной базы для развития их профессиональных навыков.

В книге две части: первая – это теоретические основы обработки сигналов, куда включены необходимые разделы из теории сигналов, теории аналого-цифрового преобразования, теории частотно-временного преобразования, теории телекоммуникаций, фильтрации. Вторая часть содержит описания технологий обработки сигналов на примерах реальных научных задач, где читателю предлагается убедиться в преимуществах междисциплинарного подхода.

Идея написания пособия появилась благодаря книге В. С. Гутникова [4], в которой рассматривается внедрение методов цифровой обработки сигналов в практику создания измерительных систем. Важные практические результаты, полученные Валентином Сергеевичем и его учениками, вошли в первую часть настоящего пособия [2].

Другим важным направлением работы кафедры, проводимым Г. Н. Солопченко, является развитие основ теории и практики обеспечения единства измерений, внедрение современных методов обработки результатов измерений, отвечающих требованиям нормативной метрологии. Вторая часть настоящего пособия включает научные разработки Геннадия Николаевича и его учеников [12, 14], что позволит студентам, изучающим данный курс получить представление о многообразии задач современной обработки сигналов.

При подготовке пособия были учтены предложения и пожелания коллег и студентов. Авторы выражают благодарность А. Е. Осадчему, читавшему курс цифровой обработки сигналов.



Гутников Валентин Сергеевич, профессор, доктор технических наук, действительный член Метрологической академии, заслуженный деятель науки Российской Федерации, заведующий кафедрой измерительных информационных технологий с 1997 г. по 2005 г. Валентин Сергеевич известен как крупнейший специалист по измерительной электронной схемотехнике, автор нескольких книг по применению микроэлектроники в измерительной технике, которые стали настольными у инженеров, преподавателей и студентов всей страны. Научные труды В. С. Гутникова включают более 250 наименований, в том числе более 100 авторских свидетельств на изобретения. Число подготовленных им кандидатов наук приближается к тридцати, из них девять стали ведущими доцентами кафедры.

ВВЕДЕНИЕ

Аналитическое исследование поведения информационной системы основано на построении адекватной математической модели, отражающей характеристики элементов системы и возможные способы их взаимодействия. Понятие функции определяемой с помощью математической записи в виде $y(x)$ позволяет применять математический аппарат функций в качестве базовой основы построения моделей технических систем. Особое значение функциональная математика имеет в технике связи, где временные функции вида $s(t)$, $v(f)$ и так далее, используемые для передачи информации, получили название сигналов. При анализе сигналов предметом исследований является сигнал как математический объект, представленный в виде некоторой функции $f(t)$, где t – независимая переменная любой физической природы (время, перемещение, частота). Функции, служащие для описания реальных сигналов, всегда вещественны. Понятие "сигнал" широко используется в информационных системах и обычно обозначает физический процесс, являющийся материальным носителем информационного сообщения – изменение какого-либо параметра носителя во времени, в пространстве или в зависимости от других аргументов. Термин "сигнал" также может означать смысловое содержание определенного физического состояния или процесса, например, сигналы светофора, звуковые сигналы и тому подобное.

Сигнал часто отождествляется с понятиями "данные" и "информация". Действительно, эти понятия взаимосвязаны, но относятся к разным категориям.

В настоящее время мировая наука склоняется к точке зрения, что информация, наряду с материей и энергией, принадлежит к фундаментальным философским категориям естествознания и относится к одному из свойств объективного мира. В то время как "данные" представляют собой только совокупность фактов,

результатов наблюдений, измерений, представленных в формализованном виде, количественном или качественном [4].

С точки зрения информационных систем важным является следующее соображение: *сигнал выступает средством материализации данных*; благодаря сигналам данные могут передаваться и обрабатываться с целью извлечения информации [4]. В процессе передачи сигналы могут менять свои свойства (их можно усиливать, фильтровать, модулировать, то есть целенаправленно изменять их спектр и так далее) тогда как данные (следовательно, информация), составляющие сообщение, которое переносит сигнал, остаются *неизменными* на всем пути от источника до приемника. Итак, резюмируя сказанное выше, можно заключить, что:

Информация – это мера изменения неопределенности субъективных представлений о физическом мире – объекте, системе или среде.

Данные – это элемент информации о физическом объекте, системе или среде, полученный в результате измерения и преобразования сигналов и обычно представленный в *хорошо формализованном виде*.

Сигнал – это *функция*, несущая сообщение (данные) о физических свойствах, состоянии или поведении какой-либо физической системы, объекта или среды.

Таким образом, можно сформулировать основное требование к процессам обработки сигналов в информационных системах: процессы обработки сигналов *не должны изменять количества информации в системе* (не порождать новых данных и достоверно передавать имеющиеся).

Настоящее пособие предназначено для студентов, изучающих основные принципы проектирования информационных систем, неотъемлемой частью которых является обработка и преобразование сигналов. Так как к изучению программы могут приступать студенты, имеющие различный уровень подготовки, то в пособие включен

раздел теоретических основ цифровой обработки сигналов, базирующихся на [10, 11]. И вместе с тем, инновационный характер исследования позволяет акцентировать внимание студентов на широко внедряемые в практику проектирования информационных систем методов обработки сигналов, используемых аналогово-цифровых преобразователей на основе сигма-дельта и вейвлет преобразованиях [3]; не менее важным является и изучение методов, применяемых при обработке результатов измерения в сложных информационных системах, относящимся к классу обратных задач [8, 12, 15].

Пособие состоит из восьми разделов.

Первая часть состоит из четырех разделов, где рассматриваются теоретические вопросы обработки сигналов, такие как представление сигналов в информационных системах, их преобразования линейными системами, методы частотно-временного анализа и разложение сигналов в различных функциональных базисах. Приведены примеры основных операций преобразования сигналов, таких как фильтрация, модуляция, расширение спектра.

Инновационный характер современных методов обработки сигналов представлен в разделах 5–8 и содержит обзор современных методов аналого-цифрового преобразования, построения сигма-дельта АЦП, алгоритмов дискретного вейвлет-преобразования. Отдельно рассмотрены обратные задачи применительно к использованию их в методах обработки результатов измерений, значимость которых связана с необходимостью метрологической аттестации каналов получения обработки информации.

1. ПРЕДСТАВЛЕНИЕ СИГНАЛОВ В ИНФОРМАЦИОННЫХ СИСТЕМАХ

Прежде чем приступать к рассмотрению задач анализа сигналов, выделим некоторые классы сигналов, которые будут часто встречаться нам в дальнейшем. В наиболее общей формулировке сигнал — это зависимость одной величины от другой (то есть с математической точки зрения сигнал является *функцией*). Физическая природа сигнала может быть весьма различной. Очень часто это напряжение, несколько реже — ток, возможны и многие другие физические величины.

В данной книге подразумевается (если иное не оговорено специально), что независимой переменной в математическом представлении сигнала выступает время, хотя в отдельных примерах независимая переменная в сигнале фактически временем не является.

Независимая переменная в математической реализации сигнала может быть как непрерывной, так и дискретной. Сигналы с *непрерывным временем* определяются на непрерывных промежутках, зависят от непрерывной переменной и обычно называются *аналоговыми*. Сигналы с *дискретным временем* или просто дискретные сигналы определены в дискретные моменты времени и, как следствие, имеют независимую переменную с дискретными значениями. Тем самым они отождествляются с числовыми последовательностями. Кроме независимой переменной, дискретным или непрерывным может быть уровень (величина) сигнала. К *цифровым* сигналам относят те, у которых дискретны как переменная, так и уровень.

Аналогично и системы обработки сигналов могут быть разбиты на два класса. *Системой, непрерывной по времени*, или просто *непрерывной системой*, принято называть такую систему, у которой как входной сигнал, так и реакция на него — непрерывные сигналы, а *дискретной по времени (дискретной)* называют систему, у которой

входной и выходной сигналы дискретны. *Цифровой* системой называют систему, перерабатывающую цифровой сигнал в цифровой же. Таким образом, цифровая обработка сигналов имеет дело с преобразованием сигналов, дискретных как по времени, так и по уровню. Основное внимание в этом пособии уделено дискретным (а не цифровым) сигналам и системам.

1.1. ДИСКРЕТНЫЙ СИГНАЛ

Дискретный сигнал обычно получают из непрерывного (аналогового) сигнала путем применения операций квантования по уровню и дискретизации по времени. Операция квантования заключается в преобразовании дискретного сигнала $s(t_n)$ в цифровой сигнал $s(n) = s_n \approx s(t_n)$, $n = 0, 1, 2, \dots, N$, как правило, кодированный в двоичной системе счисления. Амплитудные значения сигнала определяются как $y_n = Q_k[y(n\Delta t)]$, где Q_k – функция квантования с числом уровней квантования k , при этом интервалы квантования могут быть как с равномерным распределением, так и с неравномерным, например – логарифмическим. Процесс преобразования бесконечных по значениям аналоговых отсчетов в конечное число дискретных значений называется квантованием по уровню, а возникающие при квантовании ошибки округления отсчетов (отбрасываемые значения) – шумами или ошибками квантования.

Величина, обратная шагу дискретизации, $f_\delta = 1/\Delta t$, называется *частотой дискретизации (sampling frequency)*. Если дискретный сигнал получен *дискретизацией (sampling)* аналогового сигнала, то он представляет собой последовательность отсчетов, значения которых равны значениям исходного сигнала по координатам $n\Delta t$. Рассмотрим некоторые примеры дискретных сигналов.

1.1.1. Цифровой единичный импульс

Элементарным сигналом является *цифровой единичный импульс (ЦЕИ)*. Мы будем обозначать этот сигнал как $\delta[n - n_0]$. Формальное

определение этого сигнала и его графическое изображение приведены на рис. 1.1.



Рис. 1.1. Цифровой единичный импульс

ЦЕИ используется для представления произвольных дискретных последовательностей. Например, запишем следующую последовательность, изображенную на рис. 1.2, при помощи взвешенной комбинации ЦЕИ с соответствующими задержками. Обратите внимание, что нам надо записать формулу последовательности $x[n]$ так, чтобы, подставив значение n , мы получили бы значение сигнала.

$$x[n] = 1\delta[n - 1] + 2\delta[n - 2] + 3\delta[n - 3] + 4\delta[n - 4] + 5\delta[n - 5] + 4\delta[n - 6] + 3\delta[n - 7] + 2\delta[n - 8] + 1\delta[n - 9] + (-1)\delta[n - 10]$$

В более общем виде мы можем записать $x[n] = \sum_{k=-\infty}^{\infty} x[k]\delta[n - k]$.

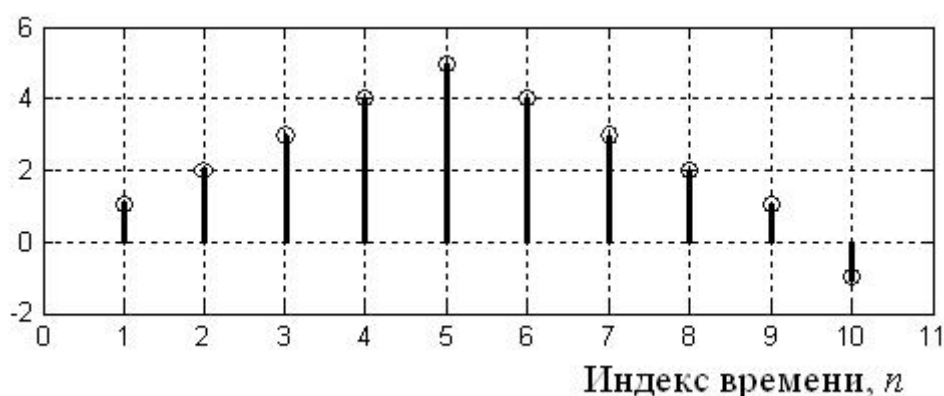


Рис. 1.2. Пример цифровой последовательности x

ЦЕИ является дискретным аналогом **дельта-функции Дирака** $\delta(t)$, широко используемой при анализе аналоговых линейных систем. В соответствии с определением: $\delta(t) = \begin{cases} \infty, & t = 0 \\ 0, & t \neq 0 \end{cases}$.

Замечательным свойством этой функции является способность "вырезать" значение произвольной функции, соответствующее фиксированному значению аргумента. Формально это свойство

$$\text{записывается как } \int_{-\infty}^{\infty} g(t) \cdot \delta(t - t_0) dt = g(t_0).$$

Его возможно проиллюстрировать следующим образом (рис. 1.3).

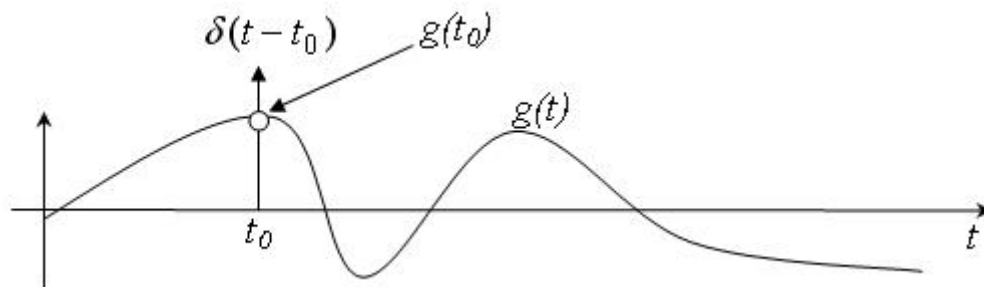


Рис. 1.3. Свойство дельта-функции Дирака

Данное свойство нетрудно доказать, если воспользоваться другим определением дельта-функции Дирака: $\delta(t - t_0) = \lim_{\varepsilon \rightarrow 0} d_{t_0}(t)$,

где

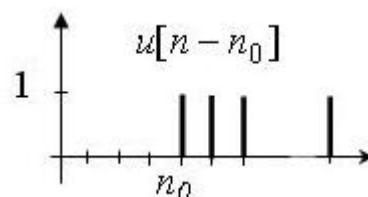
$$d_{t_0}(t) = \begin{cases} 1/2\varepsilon, & -\varepsilon \leq t - t_0 \leq \varepsilon \\ 0 & \text{иначе} \end{cases}.$$

Несмотря на то, что $\delta(t - t_0)|_{t=t_0} = \infty$,

$$\int_{-\infty}^{\infty} \delta(t - t_0) dt = \int_{t_0 - \varepsilon}^{t_0 + \varepsilon} \delta(t - t_0) dt = 1.$$

1.1.2. Цифровой единичный скачок

$$u[n - n_0] = \begin{cases} 1, & n \geq n_0 \\ 0, & \text{иначе} \end{cases}.$$



Цифровой единичный скачок (ЦЕС) можно записать в виде бесконечной суммы ЦЕИ:

$$u[n - n_0] = \sum_{k=n_0}^{\infty} \delta[n - k].$$

В свою очередь ЦЕИ является линейной комбинацией двух ЦЕС:

$$\delta[n - n_0] = u[n - n_0] - u[n - n_0 - 1].$$

1.1.3. Цифровая синусоидальная последовательность

Основная польза цифровой синусоидальной последовательности (ЦСП) заключается в том, что линейные инвариантные системы (ЛИС), о которых мы поговорим позже, изменяют лишь амплитуду и фазу цифровой синусоиды. Это свойство позволяет использовать ЦСП для описания ЛИС.

Аналитически ЦСП записывается как $x[n] = A \sin(\omega_0 n + \phi)$. Три параметра A , ω и ϕ определяют амплитуду, частоту и фазу ЦСП. Пример такой последовательности приведен на рис. 1.4.

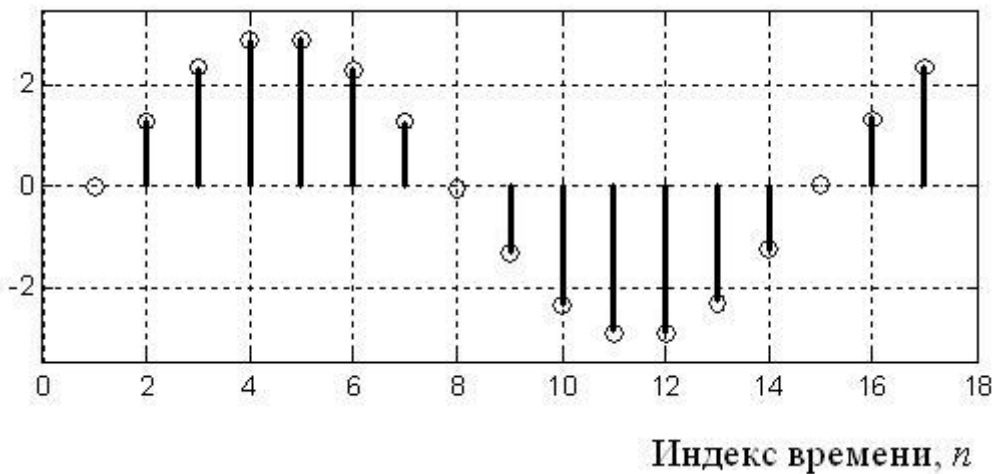


Рис. 1.4. Цифровая синусоидальная последовательность

ЦСП с частотами, отличающимися на целое число 2π , неотличимы друг от друга:

$$A \cdot \sin(\omega_0 n + \phi) = A \cdot \sin((\omega_0 + 2\pi r)n + \phi) = A \cdot \sin(\omega_0 n + \phi + 2\pi r n).$$

При рассмотрении ЦИС нас будут интересовать лишь частоты в диапазоне $\omega \in [0, 2\pi)$.

1.1.4. Цифровая экспоненциальная последовательность

Ещё одна очень важная последовательность — цифровая экспоненциальная последовательность (ЦЭП): $x[n] = Ae^{cn}$. Она особенно актуальна при комплексном показателе: $x[n] = Ae^{j\omega n} = A\cos(\omega n) + jA\sin(\omega n)$ (формула Эйлера).

1.1.5. Периодичность последовательностей

Последовательность $x[n]$ периодична, если существует такое целое число N , что $x[n + N] = x[n]$ для любых n . N тогда называется периодом последовательности.

Протестируем ЦСП на периодичность:

$$\begin{aligned} A\sin(\omega n + \varphi) &= A\sin(\omega(n + N) + \varphi) = \\ &= A\sin(\omega n + \varphi + \omega N) \rightarrow \omega N = 2\pi k, k \in Z. \end{aligned}$$

Таким образом период $N = \frac{2\pi}{\omega}$.

1.2. ДИСКРЕТИЗАЦИИ ПО ВРЕМЕНИ

Дискретизация осуществляет преобразование аналоговых сигналов (функций), непрерывных по аргументу, в функции мгновенных значений сигналов по дискретному аргументу. Дискретизация обычно производится с постоянным шагом по аргументу (*равномерная дискретизация*), при этом $s(t) \Rightarrow s(n\Delta t)$, где значения $s(n\Delta t)$ представляют собой отсчеты функции $s(t)$ в моменты времени $t = n\Delta t$, $n = 0, 1, 2, \dots, N$. Частота, с которой берутся значения аналогового сигнала, называется *частотой дискретизации*. Операция восстановления аналогового сигнала из его дискретного представления обратна операции дискретизации и представляет, по существу, интерполяцию данных. Однако, т.к. операция дискретизации приводит к изменению частотных свойств сигнала, то при интерполяции требуется учесть влияние этих изменений, ввиду этого оба процесса будут подробно описаны в разделе частотно-временных преобразований.

Дискретизация сигналов может приводить к определенной потере информации о поведении сигналов в промежутках между отсчетами. Однако существуют условия, определенные теоремой Шеннона-Котельникова, согласно которым аналоговый сигнал с ограниченным частотным спектром может быть без потерь информации преобразован в дискретный сигнал, и затем абсолютно точно восстановлен по значениям своих дискретных отсчетов [10].

Математическое описание (1.1) и алгоритмическая блок-схема устройства дискретизации по времени (рис. 1.5):

$$s_T(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT) = \begin{cases} \infty, & t = nT, n \in Z \\ 0 & \end{cases} \quad (1.1)$$

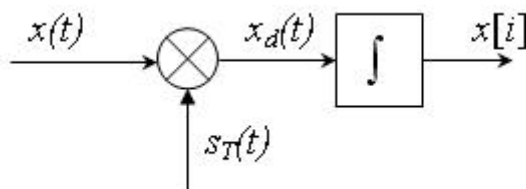


Рис. 1.5. Блок-схема устройства дискретизации по времени

Устройство дискретизации сигнала состоит из устройства выделения сигнала, контролируемого сигналом-стробом $S_T(t)$, и устройства преобразования импульсов с бесконечно большой амплитудой и бесконечно короткой длительностью в последовательность цифровых импульсов с амплитудой, равной значению сигнала в моменты времени nT . Это устройство изображено на схеме в виде интегрирующего звена.

Дискретизирующий строб-сигнал $S_T(t)$ может быть представлен как сумма цифровых единичных импульсов с различной задержкой,

т.е. $S_T(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT)$. Тогда дискретизированная по времени версия

сигнала $x(t)$ может быть записана как

$$x_d(t) = x(t) \cdot s_T(t) = x(t) \sum_{n=-\infty}^{\infty} \delta(t - nT) = \sum_{n=-\infty}^{\infty} \delta(t - nT) x(nT). \quad (1.2)$$

2. ЛИНЕЙНЫЕ СИСТЕМЫ

2.1. ОПРЕДЕЛЕНИЕ ЛИНЕЙНОЙ СИСТЕМЫ

Преобразование $T\{\bullet\}$ называется линейным, если $T\{ax_1 + bx_2\} = aT\{x_1\} + bT\{x_2\}$. Система, осуществляющая линейное преобразование, называется линейной системой, сокращённо ЛС.

Ошибочно думать, что, например, система, работающая в соответствии с уравнением $y[n] = T\{x[n]\} = 1 + x[n]$, в котором нет никаких нелинейных операторов, линейна. Проверим это:

$$\begin{aligned} T\{ax_1 + bx_2\} &= 1 + ax_1 + bx_2 \neq \\ &\neq aT\{x_1\} + bT\{x_2\} = a(1 + x_1) + b(1 + x_2) = a + b + ax_1 + bx_2. \end{aligned}$$

Определение линейности даётся на основании принципа суперпозиции, $T\left\{\sum_n a_n x_n\right\} = \sum_n a_n T\{x_n\}$. Пусть $x[n] = \sum_k a_k \delta[n-k]$, где $\delta[n]$ – ЦЕИ. Таким образом,

$$T\{x[n]\} = T\left\{\sum_k a_k \delta[n-k]\right\} = \sum_k a_k T\{\delta[n-k]\}.$$

Теперь представим себе, что мы знаем отклик нашей системы на ЦЕИ и что, если $T\{\delta[n]\} = y[n]$, то $T\{\delta[n-n_0]\} = y[n-n_0]$. Тогда отклик системы на произвольное воздействие может быть вычислен как взвешенная сумма (суперпозиция) сдвинутых во времени откликов на единичное воздействие. Весовые коэффициенты в этой сумме составляют *импульсную характеристику (ИХ)*, а операция вычисления отклика называется *дискретной свёрткой*.

2.2. СВОЙСТВО ИНВАРИАНТНОСТИ ВО ВРЕМЕНИ

ЛС называется инвариантной во времени, если параметры системы постоянны и не являются функциями времени.

Двусторонним следствием этого определения является свойство системы, инвариантной во времени:

$$T\{x[n]\} = y[n] \rightarrow T\{x[n - n_0]\} = y[n - n_0]. \quad (2.1)$$

С другой стороны, если свойство (2.1) соблюдается для любой последовательности $x[n]$, то система инвариантна во времени. Мы будем сокращённо называть линейную систему, инвариантную во времени, ЛСИВ или ЛСПП.

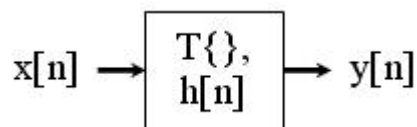
Примером системы, не инвариантной во времени, служит компрессор, работающий в соответствии с уравнением $y[n] = x[Mn]$ и осуществляющий преобразование входного сигнала посредством сохранения лишь каждого M -го отсчёта последнего. Проверим это:

$$\begin{aligned} y[n] = T\{x[n]\} = x[Mn] &\rightarrow y[n - n_0] = x[Mn - n_0] \neq \\ &\neq T\{x[n - n_0]\} = x[M(n - n_0)], \forall M \neq 1. \end{aligned}$$

2.3. ИМПУЛЬСНАЯ ХАРАКТЕРИСТИКА

Рассмотрим систему [10], работающую в соответствии с уравнением $y[n] = \sum_{k=-M_1}^{M_2} \frac{1}{M_1 + M_2 + 1} x[n - k]$ и вычисляющую скользящее среднее.

На рис. 2.1 изображены выход системы $y[n]$, вход $x[n]$ и механизм вычисления выхода на основании входа для $M_1 = M_2 = 1$. Здесь видно, что мы как бы "скользим" вдоль входного сигнала окном коэффициентов размером $1/3$ и получаем выход в виде взвешенной суммы отсчётов входного. Например, в соответствии с формулой, для того, чтобы получить *выходной* отсчёт номер 3, мы суммируем 2-ой, 3-ий и 4-ый отсчёты *входного* сигнала. В общем случае для n -го отсчёта получаем $y[n] = \frac{1}{3} x[n - 1] + \frac{1}{3} x[n] + \frac{1}{3} x[n + 1]$.



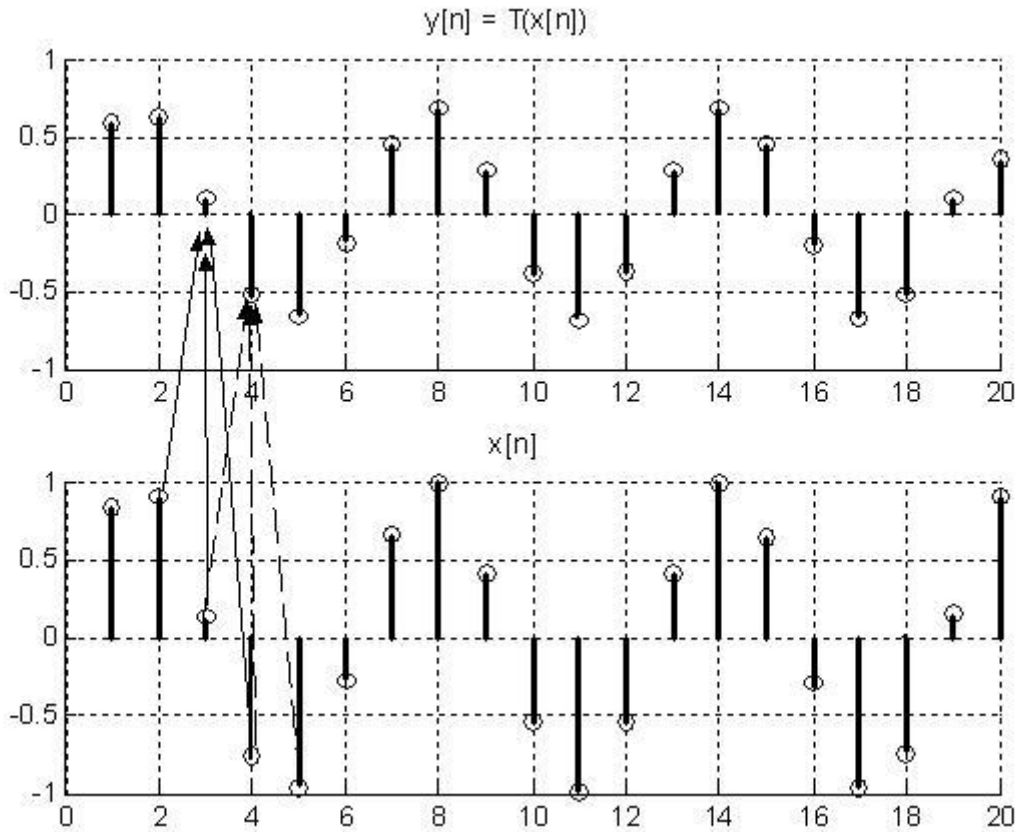


Рис. 2.1. Механизм вычисления выходной характеристики на основании входной

Описанная процедура вычисления выходного сигнала, выполненная для каждого из отсчётов последнего, представляет собой частный случай операции, называемой *дискретной свёрткой*. В дискретной свёртке участвуют две цифровые последовательности, $\{x[n]\}$ и $\{h[n]\}$. В приведённом примере системы, вычисляющей скользящее среднее, последовательность $\{h[-1], h[0], h[1]\} = \left\{\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right\}$.

Более общее выражение операции свёртки может быть записано как

$$y[n] = \sum_{k=-\infty}^{\infty} h[k]x[n-k] = \sum_{k=-\infty}^{\infty} h[n-k]x[k]. \quad (2.2)$$

Из определения видно, что свёртка симметрична по отношению к двум своим операндам. Второе определение получается из первого простой заменой переменных.

Воспользуемся операцией свёртки и подсчитаем отклик системы на воздействие цифровым единичным импульсом, т.е. $x[n] = \delta[n]$. Нетрудно видеть, что

$$y[n] = T\{\delta[n]\} = \sum_{k=-\infty}^{\infty} h[k]\delta[n-k] = \sum_{k=-\infty}^{\infty} h[k]\delta[k-n] = h[n]$$

на основании вырезающего свойства ЦЕИ.

Таким образом, последовательность $h[n]$, участвующая при вычислении отклика линейной системы, называется *импульсной характеристикой* (ИХ) системы, поскольку представляет собой отклик системы на единичный импульс.

2.4. ФИЗИЧЕСКИ РЕАЛИЗУЕМАЯ СИСТЕМА

На примере со скользящим средним мы вычисляли выход системы в момент времени n в соответствии со следующим выражением:

$$y[n] = \frac{1}{3}x[n-1] + \frac{1}{3}x[n] + \frac{1}{3}x[n+1] = h[1]x[n-1] + h[0]x[n] + h[-1]x[n+1].$$

Такая система физически нереализуема, так как для вычисления отклика в момент времени n требуется значение входного сигнала в момент времени $n+1$, т.е. из будущего! Проанализировав уравнение дискретной свёртки возможно сделать вывод, что ЛС физически реализуема, если $h[n] = 0, \forall n < 0$. В самом деле, тогда

$$y[n] = \sum_{k=-\infty}^{\infty} h[k]x[n-k] = \sum_{k=0}^{\infty} h[k]x[n-k] \text{ и при вычислении отклика}$$

используются только прошлые или текущий отсчёты входного сигнала.

2.5. УСТОЙЧИВОСТЬ ЛИНЕЙНОЙ СИСТЕМЫ

ЛС называется устойчивой, если отклик на ограниченное по амплитуде воздействие ограничен по амплитуде. ЛС устойчива тогда

и только тогда, когда её импульсная характеристика абсолютно суммируема, иными словами ЛС устойчива, если $S = \sum_{k=-\infty}^{\infty} |h[k]| < \infty$.

Если $S < \infty$, то выход ЛС ограничен по амплитуде при условии ограниченного входного воздействия; в самом деле, если $x[n] < B_x, \forall n$ то,

$$|y[n]| = \left| \sum_{k=-\infty}^{\infty} h[k]x[n-k] \right| \leq \sum_{k=-\infty}^{\infty} |h[k]| |x[n-k]| \leq B_x \sum_{k=-\infty}^{\infty} |h[k]| = B_x S.$$

Теперь покажем, что если $S = \infty$, то отклик системы на ограниченное по амплитуде воздействие неограничен. Продемонстрируйте это сами: рассмотрите отклик такой системы $y[0]$ на последовательность

$$x[n] = \begin{cases} \frac{h^*[n]}{|h^*[-n]|}, & h[n] \neq 0, \\ 0, & h[n] = 0. \end{cases} \quad y[0] = \sum_{-\infty}^{\infty} \dots = S = \infty.$$

2.6. ПРИМЕРЫ АНАЛИЗА ЛИНЕЙНЫХ СИСТЕМ НА УСТОЙЧИВОСТЬ И ФИЗИЧЕСКУЮ РЕАЛИЗУЕМОСТЬ

Идеальная задержка:

$$h[n] = \delta[n - n_d] \quad \text{устойчива} \quad \begin{array}{l} \text{физически} \\ \text{реализуема} \end{array}$$

Скользящее среднее:

$$h[n] = \frac{1}{M_1 + M_2 + 1} \sum_{k=-M_1}^{M_2} \delta[n-k] \quad \text{устойчива} \quad \begin{array}{l} \text{физически} \\ \text{нереализуема} \end{array}$$

Аккумулятор (дискретный интегратор):

$$h[n] = \sum_{k=-\infty}^n \delta[k] \quad \text{неустойчива} \quad \begin{array}{l} \text{физически} \\ \text{реализуема} \end{array}$$

Правая разность:

$$h[n] = \delta[n+1] - \delta[n] \quad \text{устойчива} \quad \text{физически нереализуема}$$

Левая разность:

$$h[n] = \delta[n] - \delta[n-1] \quad \text{устойчива} \quad \begin{array}{l} \text{физически} \\ \text{реализуема} \end{array}$$

Указанные и другие примеры подробно рассмотрены в [10,11].

2.7. СИСТЕМЫ С КОНЕЧНОЙ И БЕСКОНЕЧНОЙ ИМПУЛЬСНЫМИ ХАРАКТЕРИСТИКАМИ

Цифровые фильтры принято классифицировать на фильтры с конечной импульсной характеристикой (КИХ) [4] и фильтры с бесконечной импульсной характеристикой (БИХ) [4].

Система с *конечной импульсной характеристикой* имеет конечное число ненулевых элементов в своей импульсной характеристике, например, скользящее среднее. Структурная система представлена на рис. 2.2.

Только КИХ фильтры принципиально могут иметь фазовую характеристику, линейно зависящую от частоты [3].

Алгоритм дискретной фильтрации приведен ниже:

$$y(nT_2) = \sum_{m=0}^{N-1} a_m[(n-m)T_2],$$

где nT_2 – текущее значение квантованного времени.

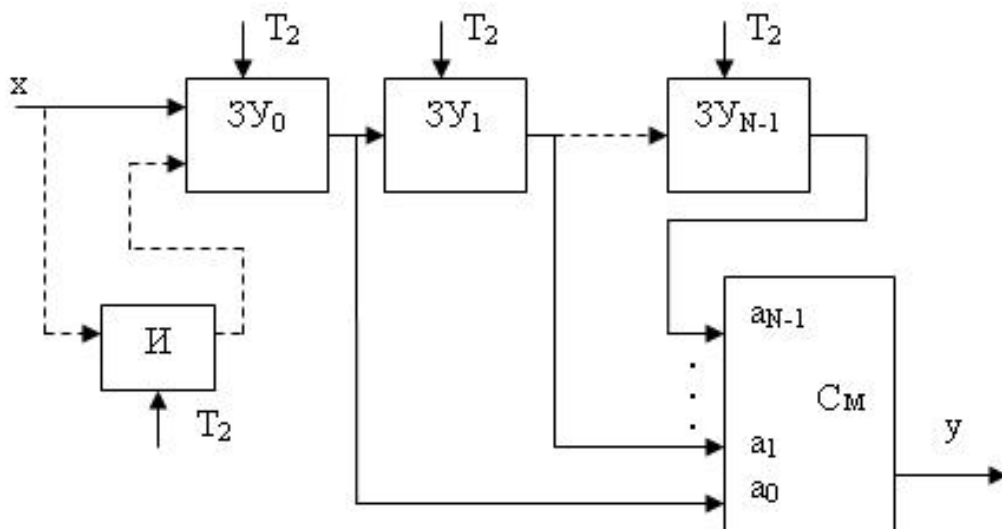


Рис. 2.2. Структурная схема КИХ-фильтра

Соответственно, система с *бесконечной импульсной характеристикой* имеет бесконечное число ненулевых элементов в своей импульсной характеристике, например, аккумулятор (дискретный интегратор). Такие системы мы будем сокращённо называть *БИХ* системами. Структура приведена на рис. 2.3.

Передаточная функция БИХ-фильтра имеет следующий вид:

$$G(z) = \frac{\sum_{i=0}^{N-1} a_i z^{-i}}{1 + \sum_{i=1}^{m-1} b_i z^{-i}}$$

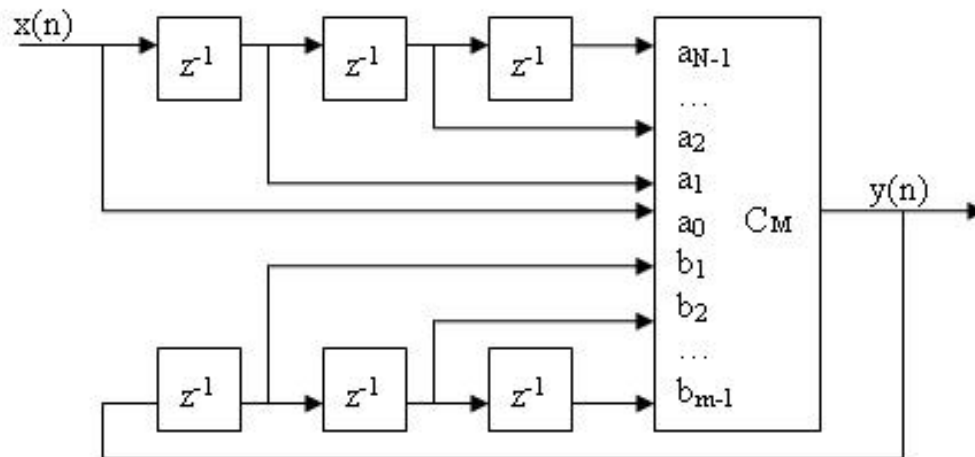


Рис. 2.3. Общая структура БИХ-фильтра

Примером устойчивой БИХ системы может служить таковая с импульсной характеристикой $h[n] = a^n u(n)$ при $|a| < 1$. Тогда

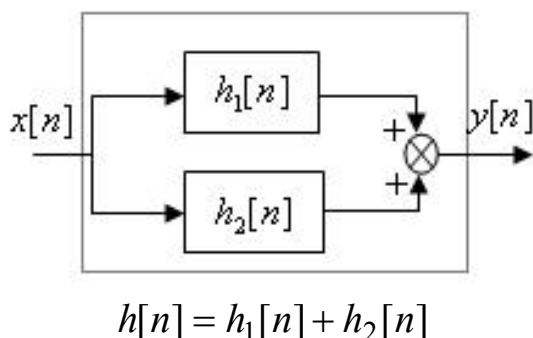
$$S = \sum_{n=0}^{\infty} |h(n)| = \sum_{n=0}^{\infty} |a^n| = \frac{1}{1-|a|} < \infty \quad \forall a : |a| < 1.$$

2.8. КАСКАДНОЕ СОЕДИНЕНИЕ ЛИНЕЙНЫХ СИСТЕМ ИНВАРИАНТНЫХ ВРЕМЕНИ

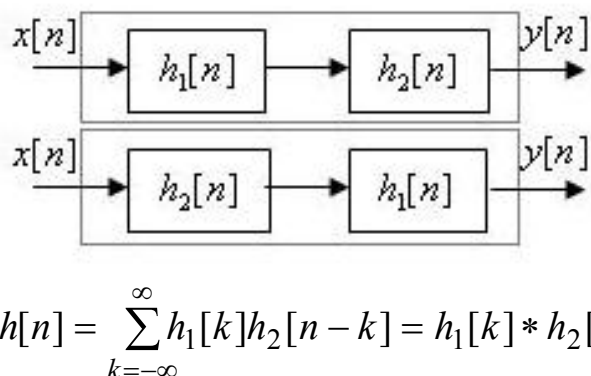
При последовательном соединении систем реакция первой системы подается на вход второй, отклик второй – на третий и т.д. Отклик последней системы служит реакцией всей системы.

При параллельном соединении системы имеют общий вход, а их выходные последовательности складываются и дают реакцию всего соединения.

Параллельное



Последовательное



Вычислим импульсную характеристику последовательно соединённых систем левой разности и аккумулятора:

$$h_1[n] = u[n],$$

$$h_2[n] = \delta[n] - \delta[n-1],$$

$$h[n] = h_1[n] * h_2[n] = u[n] - u[n-1] = \delta[n].$$

Таким образом, эти две системы осуществляют взаимно обратные преобразования. Это и не удивительно, так как аналогами этих двух систем, работающими в непрерывном времени являются дифференциатор и интегратор!

Вообще говоря, импульсная характеристика ЛСИВ системы, обратной данной системе с ИХ $h[n]$, может быть найдена как решение уравнения $h[n] * h_i[n] = \delta[n]$.

В большинстве ситуаций найти решение такого уравнения непросто. Аппарат Z-преобразования позволяет легко справиться с этой задачей (см. раздел 3.7).

2.9. ОПИСАНИЕ ЛИНЕЙНЫХ СИСТЕМ ПРИ ПОМОЩИ РАЗНОСТНЫХ УРАВНЕНИЙ

Важный класс ЛСИВ может быть описан при помощи разностных уравнений N -го порядка с постоянными коэффициентами,

общий вид которых может быть записан как

$$\sum_{k=0}^N a_k y[n-k] = \sum_{k=0}^M b_k x[n-k].$$

Рассмотрим аккумулятор: $y[n] = \sum_{k=-\infty}^n x[n-k]$. В этом виде

уравнение не является разностным уравнением конечного порядка, но запишем выражение для

$$y[n-1] = \sum_{k=-\infty}^{n-1} x[n-k] = y[n] - x[n], \quad (2.3)$$

тогда получаем $y[n] - y[n-1] = x[n]$, следовательно $a_0 = 1, a_1 = -1, b_0 = 1, M = 0, N = 1$.

Попытка описать аккумулятор при помощи разностного уравнения привела нас к существенно более экономичной схеме (рис. 2.4) реализации такой системы по сравнению с определением.

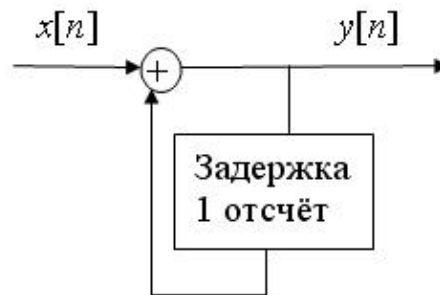


Рис. 2.4. Схема реализации аккумулятора

Пример

Рассмотрим физически реализуемый вариант системы, вычисляющей скользящее среднее с импульсной характеристикой

$$h[n] = \frac{1}{M_2 + 1} (u[n] - u[n - M_2 - 1]) \text{ и соответственно осуществляющей}$$

$$\text{преобразование } y[n] = \frac{1}{M_2 + 1} \sum_{k=0}^{M_2} x[n-k].$$

Попробуем записать выражение для $h[n]$ в виде свёртки $h[n] = \frac{1}{M_2 + 1} (\delta[n] - \delta[n - M - 1]) * u[n]$. Таким образом, скользящее среднее может быть реализовано как последовательно соединённые аккумулятор и система левой разности с задержкой $M_2 + 1$ (рис. 2.5).

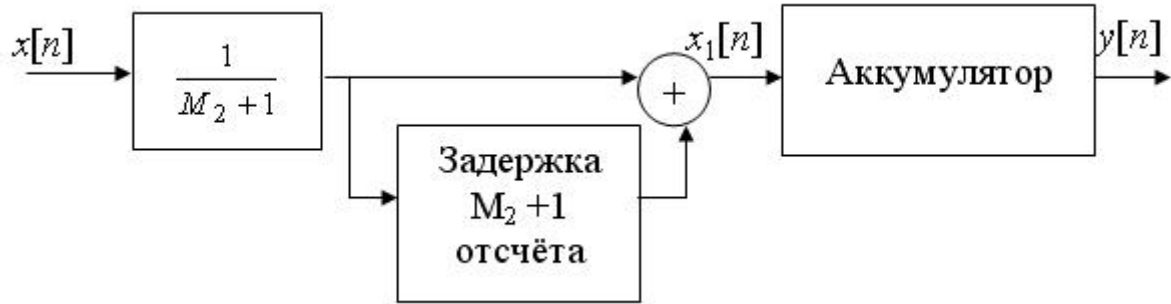


Рис. 2.5. Схема реализации скользящего среднего

Тогда $x_1[n] = \frac{1}{M_2 + 1} (x[n] - x[n - M_2 - 1])$ и в соответствии с (2.3), мы можем записать следующее выражение, описывающее работу аккумулятора $y[n] - y[n - 1] = x_1[n]$. Объединяя эти два выражения, мы получим разностное уравнение $y[n] - y[n - 1] = \frac{1}{M_2 + 1} (x[n] - x[n - M_2 - 1])$, описывающее работу физически реализуемого вычислителя скользящего среднего.

Полученная форма реализации этой системы $y[n] = \frac{1}{M_2 + 1} (x[n] - x[n - M_2 - 1]) + y[n - 1]$ существенно более экономична по сравнению с реализацией на основании определения системы $y[n] = \frac{1}{M_2 + 1} \sum_{k=0}^{M_2} x[n - k]$, что становится особенно заметным при больших значениях M_2 .

3. МЕТОДЫ ЧАСТОТНО-ВРЕМЕННОГО АНАЛИЗА СИГНАЛОВ. РАЗЛОЖЕНИЕ СИГНАЛОВ В БАЗИСЕ

3.1. ОБЩИЕ ПОЛОЖЕНИЯ

Разложение (анализ) сигнала отражает общий подход к исследованию физического объекта x путем его представления в виде суммы элементарных блоков $x_i : x = \sum x_i$. Для сигналов или изображений это могут быть гармонические функции (базис Фурье-анализа, дискретное косинусное преобразование), гармонические функции вида $\sin x/x$ (базис Шеннона-Котельникова), прямоугольные функции (базис Хаара), некоторые функции специального вида (вейвлет-анализ). Принцип моделирования реального сигнала аналитической функцией некоторого (возможно бесконечного) числа переменных $f(t_1, t_2, \dots, t_n)$ позволяет применить аппарат функционального анализа к процессам передачи и обработки информации. В частности, рассматривая функцию $f(t_1, t_2, \dots, t_n)$ как элемент некоторого пространства функций, можно провести ее разложение в базисе этого пространства [1].

В общем случае, базисом бесконечномерного линейного пространства $\mathfrak{R}$, на котором определено понятие скалярного произведения функций (векторов), т.е. Гильбертова пространства, называется набор линейно независимых функций (векторов) $\{v_i\}$ такой, что для любой $f \in \mathfrak{R}$ существует единственная линейная комбинация векторов $\{v_i\}$.

$$f = c_1 v_1 + c_2 v_2 + \dots + c_n v_n + \dots \quad (3.1)$$

где $\{c_i\}$ – коэффициенты разложения, v_i – компоненты базиса, указывающие направление в пространстве.

Коэффициенты разложения определяются как скалярное произведение сигнала f и базисных функций:

$$c_1 = \langle f, v_i \rangle \quad (3.2)$$

и (3.1) можно переписать в виде:

$$f = \sum_i (< f, v_i > v_i). \quad (3.3)$$

В том случае, если исследуемый сигнал аппроксимируется кусочно-непрерывной дифференцируемой функцией, т.е. отвечает условиям известной теоремы Дирихле, для $f(t_1, t_2, \dots, t_n)$ существует преобразование Фурье – разложение в базисе комплексной экспоненты. Помимо базиса Фурье существует значительное число других базисов разложения.

Общей целью при применении разложения по базису является представление сигнала в виде, удобном для исследования и обработки, а также приближение представления сигнала к оптимальному путем минимизации числа компонент разложения при максимальном качестве аппроксимации. Очевидно, что результативность применения определенного набора базисных функций зависит от природы сигнала.

3.2. ОПИСАНИЕ ЛИНЕЙНЫХ СИСТЕМ В ЧАСТОТНОЙ ОБЛАСТИ. ПРЕОБРАЗОВАНИЕ ФУРЬЕ

3.2.1. Свёртка как матрично-векторная операция

Операцию свёртки (2.2) цифровых последовательностей $x[n]$ и $h[n]$ можно представить в виде умножения вектора и матрицы. Будем считать здесь, что $h[n]$ соответствует ИХ некоторой ЛС, а $x[n]$ – это входной сигнал. Наша задача – вычислить выход системы $y[n]$ при помощи свёртки, т.е. $y[n] = h[n] * x[n]$. Рассмотрим ситуацию, когда ИХ $h[n]$ содержит конечное число ненулевых отсчётов и $h[n] = 0, \forall n < 0$, т.е. соответствует линейной КИХ системе. И входная последовательность $x[n]$ также не содержит ненулевых элементов на отрицательной части оси абсцисс.

$$y = Hx$$

$$\begin{bmatrix} y[0] \\ y[1] \\ y[2] \\ \vdots \\ \vdots \\ y[N-1] \end{bmatrix} = \begin{bmatrix} h[0] & 0 & 0 & 0 & \cdots & 0 \\ h[1] & h[0] & 0 & 0 & \ddots & 0 \\ h[2] & h[1] & h[0] & 0 & \ddots & 0 \\ \vdots & h[2] & h[1] & h[0] & \ddots & 0 \\ h[N-2] & \vdots & \vdots & \vdots & \ddots & 0 \\ h[N-1] & h[N-2] & \dots & h[2] & h[1] & h[0] \end{bmatrix} \begin{bmatrix} x[0] \\ x[1] \\ x[2] \\ \vdots \\ \vdots \\ x[N-1] \end{bmatrix} = \\
= \begin{bmatrix} h[0]x[0] \\ h[1]x[0] + h[1]x[0] \\ h[2]x[0] + h[1]x[0] + h[0]x[0] \\ \vdots \\ \vdots \\ h[N-1]x[0] + \dots + h[0]x[N-1] \end{bmatrix}$$

Нетрудно убедиться, что данное матричное уравнение эквивалентно оператору свёртки $y[n] = \sum_{k=0}^N x[n-k]h[k]$, вычисленному на конечном интервале $[0, N]$. Основной целью данного упражнения было провести параллель между линейными операторами и линейными системами. Обратите внимание на ниже-треугольную структуру матрицы $\mathbf{H}$, соответствующую физически реализуемой системе.

Таким образом, мы убедились в том, что существует однозначное соответствие между линейными системами и линейными операторами. Это очень важно, так как такая связь открывает нам новый взгляд на вещи с использованием изящного и простого аппарата линейной алгебры.

3.2.2. Собственные функции линейного оператора и линейной системы

Вспомним из курса линейной алгебры определение собственного вектора линейного оператора. Вектор $\mathbf{e}$ называется собственным вектором матрицы $\mathbf{H}$, если найдётся такое число λ , что $\mathbf{H}\mathbf{e} = \lambda\mathbf{e}$. Число λ в этом случае называется собственным числом. Из

определения видно, что преобразование, осуществляемое линейной системой, над её собственным вектором сводится к шкалированию последнего по амплитуде. Предположим, что ЛС имеет N собственных векторов $\mathbf{e}_i$ с соответствующим набором собственных чисел $\lambda_i, i = 1, N$. Тогда, если мы можем представить наш входной сигнал (вектор) как линейную комбинацию собственных векторов системы, то мы сможем вычислить отклик системы на такое входное воздействие как взвешенную линейную комбинацию всё тех же собственных векторов.

Предположим наш входной сигнал (вектор) может быть представлен как линейная комбинация собственных векторов ЛС:

$x[n] = \sum_{k=1}^N a_k \mathbf{e}_k$. Тогда отклик $y[n]$ нашей ЛС на входное воздействие

$x[n]$ может быть подсчитан как $y[n] = \sum_{k=1}^N \lambda_k a_k \mathbf{e}_k$.

3.2.3. Синусоидальные функции для описания линейных систем

Из цифрового сигнала $x[n] = A \sin(\omega n)$ на входе ЛС мы получали выход в форме $y[n] = \Lambda A \sin(\omega n + \varphi)$. Теоретически, если предварительно оценить для всех частот ω_i значения Λ_i и φ_i , соответствующие данной системе, то, представив произвольное входное воздействие как линейную комбинацию синусоид, можно вычислить отклик системы как линейную комбинацию тех же синусоид, сдвинутых в соответствии с предварительно оцененными значениями фазы φ_i для каждой из синусоид, принимающих участие в описании (разложении) входного сигнала. Зная (ω_i, Λ_i) , мы можем вычислить выход ЛС на любое входное воздействие, представленное

как линейная комбинация синусоид $x[n] = \sum_{k=1}^N a_k \sin(\omega_k n - \theta_k)$ как

суперпозицию $y[n] = \sum_{k=1}^N \Lambda_k a_k \sin(\omega_k n - \theta_k - \varphi_k)$.

Однако достаточно очевидно, что функция $\sin(\omega n)$ не является собственной функцией линейной системы, так как $\sin(\omega n) \neq \Lambda \sin(\omega n - \varphi)$ для произвольного значения φ . Вместо синусоиды давайте рассмотрим комплексную экспоненту $e^{j\omega t} = \cos(\omega t) + j \sin(\omega t)$.

3.2.4. Комплексная экспонента – собственная функция линейных систем

Посмотрим для начала, что происходит, когда комплексная экспонента умножается на произвольное комплексное число $\Lambda = a + jb = \sqrt{a^2 + b^2} e^{j \tan^{-1}(b/a)} = |\Lambda| e^{j\angle\Lambda}$. Тогда

$$\Lambda e^{j\omega t} = |\Lambda| e^{j\angle\Lambda} e^{j\omega t} = |\Lambda| e^{j(\omega t + \angle\Lambda)} = |\Lambda| (\cos(\omega t + \angle\Lambda) + j \sin(\omega t + \angle\Lambda)).$$

Таким образом, сдвиг по фазе в гармонических функциях, составляющих комплексную экспоненту, достигается посредством умножения на комплексное число.

Применим теперь операцию свёртки и вычислим отклик ЛС с ИХ $h[n]$ на входное воздействие $x[n] = e^{j\omega n}$:

$$y[n] = \sum_{k=-\infty}^{\infty} h[k] e^{j\omega(n-k)} = \left(\sum_{k=-\infty}^{\infty} h[k] e^{-j\omega k} \right) e^{j\omega n} = H(e^{j\omega}) e^{j\omega n}.$$

Мы видим, что комплексная экспонента при проходе через ЛС просто умножается на комплексное число и таким образом является собственной функцией ЛС. Множитель $H(e^{j\omega})$ является функцией частоты и называется *частотной характеристикой системы* (ЧХ). Поскольку $H(e^{j\omega})$ – величина комплексная, мы будем изображать её при помощи двух графиков. Первый из них соответствует *амплитудно-частотной характеристике* (АЧХ), вычисляемой как

абсолютное значение комплексной величины $H(e^{j\omega})$, второй – фазовой частотной характеристике (ФЧХ), являющейся аргументом комплексной величины $H(e^{j\omega})$. Пример АЧХ и ФЧХ для фильтра Баттерворта показан на следующем рис. 3.1.

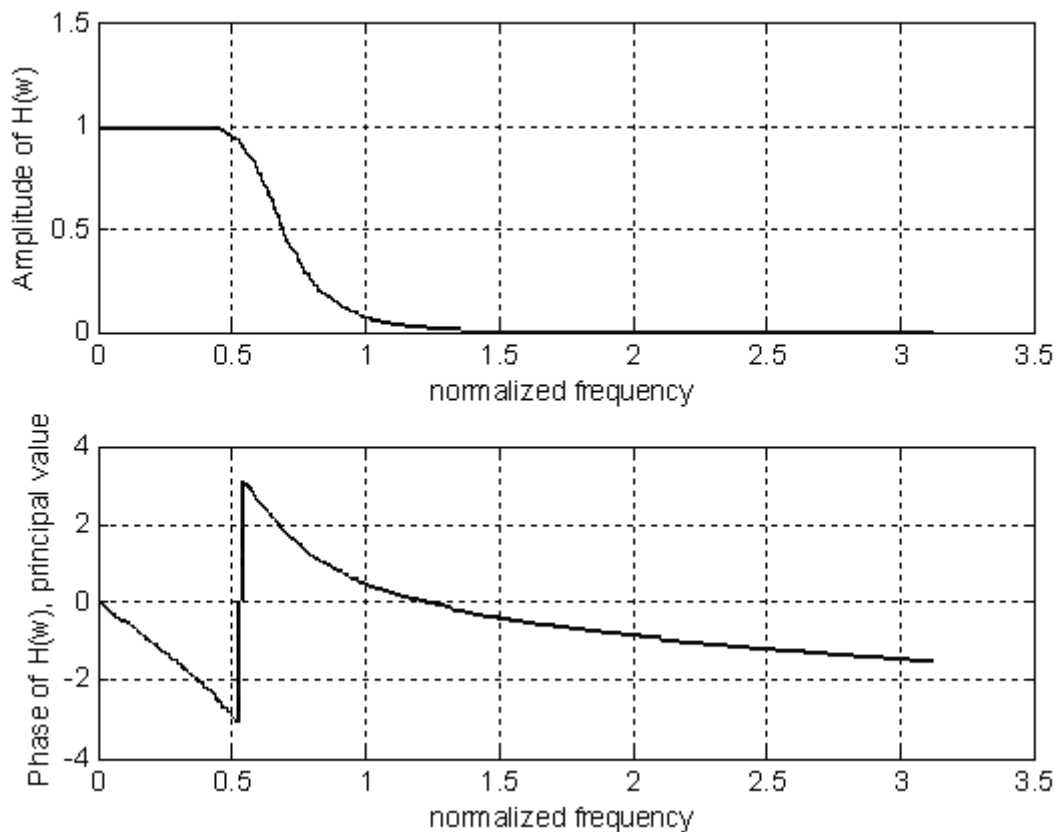


Рис. 3.1. АЧХ и ФЧХ для фильтра Баттерворта

3.2.5. Пример вычисления частотных характеристик систем

Идеальная задержка $y[n] = T\{x[n]\} = x[n - n_d]$

с ИХ $h[n] = \delta[n - n_d]$,

$$H(e^{j\omega}) = \sum_{k=-\infty}^{\infty} \delta[n - n_d] e^{j\omega n} = e^{-j\omega n_d}, \quad |H(\omega)| = 1, \quad \arg(H(\omega)) = -\omega n_d.$$

$$\text{Скользящее среднее с ИХ } h[n] = \frac{1}{2M+1} \sum_{k=-M}^M \delta(n-k),$$

$$H(e^{j\omega}) = \sum_{n=-\infty}^{\infty} \frac{1}{2M+1} \sum_{k=-M}^M \delta[n-k] e^{j\omega n} = \sum_{k=-M}^M \frac{1}{2M+1} e^{-j\omega k}.$$

3.2.6. Представление сигналов при помощи преобразования Фурье

Необходимо представлять сигналы как линейную комбинацию собственных функций ЛС. Действительно, большое число цифровых сигналов может быть представлено как линейная комбинация комплексных экспонент, являющихся, в свою очередь, собственными функциями любой линейной системы. Итак,

$$x[n] = \frac{1}{2\pi} \int_{-x}^x X(e^{j\omega}) e^{j\omega n} d\omega. \quad (3.4)$$

Частотно-зависимые коэффициенты $X(e^{j\omega})$ называются Фурье представлением сигнала и могут быть найдены как

$$x[n] = \sum_{n=-\infty}^{\infty} x[n] e^{-j\omega n}. \quad (3.5)$$

Формула (3.5) называется *Преобразованием Фурье* (ПФ), а предыдущая (3.4) – *Обратным преобразованием Фурье* (ОПФ). Преобразование Фурье представляет сигнал в виде линейной комбинации "синусоид" с бесконечно малой амплитудой $\frac{1}{2\pi} X(e^{j\omega}) d\omega$, что следует из определения интеграла как предела конечной суммы.

В общем случае, $X(e^{j\omega})$ — комплексная функция и может быть представлена как

$$X(e^{j\omega}) = X_R(e^{j\omega}) + jX_I(e^{j\omega}) = |X(e^{j\omega})| e^{i\angle X(e^{j\omega})}.$$

Величины $|X(e^{j\omega})|$ и $\angle X(e^{j\omega}) = \tan^{-1} \frac{X_I(e^{j\omega})}{X_R(e^{j\omega})}$ называются

соответственно амплитудой и фазой преобразования Фурье.

Проверим формулы преобразования Фурье.

$$\begin{aligned} \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j\omega}) e^{j\omega n} d\omega &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{m=-\infty}^{\infty} x[m] e^{-j\omega m} e^{j\omega n} d\omega = \frac{1}{2\pi} \sum_{m=-\infty}^{\infty} x[m] \int_{-\pi}^{\pi} e^{-j\omega(n-m)} d\omega = \\ &= \sum_{m=-\infty}^{\infty} x[m] \frac{\sin \pi(n-m)}{\pi(n-m)} = \frac{1}{2\pi} \sum_{m=-\infty}^{\infty} x[m] \delta[n-m] = x[n]. \end{aligned}$$

3.3. СВОЙСТВА ПРЕОБРАЗОВАНИЯ ФУРЬЕ

3.3.1. Условия существования преобразования Фурье цифровой последовательности

Математическое определение существования ПФ $X(e^{j\omega})$ цифровой последовательности $x[n]$: $X(e^{j\omega})$ существует, если $X(e^{j\omega}) < \infty \quad \forall \omega$.

Достаточное условие существования ПФ последовательности $x[n]$:

$$\left| X(e^{j\omega}) \right| = \left| \sum_{n=-\infty}^{\infty} x[n] e^{-j\omega n} \right| \leq \sum_{n=-\infty}^{\infty} |x[n]| |e^{-j\omega n}| = \sum_{n=-\infty}^{\infty} |x[n]| < \infty$$

Иными словами, у любой абсолютно суммируемой последовательности $x[n]$ существует ПФ $F\{x[n]\} = X(e^{j\omega})$.

В применении к ЛС данный вывод эквивалентен утверждению: любая устойчивая ЛС с импульсной характеристикой $h[n]$ имеет передаточную характеристику $H(e^{j\omega})$. Причина тому то, что устойчивость ЛС накладывает условие абсолютной суммируемости на ИХ системы, что в свою очередь определяет существования ПФ последней.

3.3.2. Представление цифровой последовательности как суммы чётной и нечётной последовательностей

Любую цифровую последовательность $x[n]$, в том числе и комплексную, можно представить как $x[n] = x_1[n] + x_2[n]$, где $x_1[n]$ — нечётная последовательность, такая, что $x_1[n] = -x_1^*[-n]$ и $x_2[n]$ — чётная последовательность, такая, что $x_2[n] = x_2^*[-n]$.

Найдём последовательности $x_1[n]$ и $x_2[n]$. Рассмотрим следующую сумму

$x[n] + x^*[-n] = x_1[n] + x_2[n] + x_1^*[-n] + x_2^*[-n] = 2x_2[n]$, учитывая свойства чётной и нечётной последовательностей. Отсюда мы заключаем, что $x_2[n] = \frac{1}{2}(x[n] + x^*[-n])$.

Рассмотрев разность $x[n] - x^*[-n]$, можно найти следующее выражение для нечётной последовательности $x_1[n] = \frac{1}{2}(x[n] - x^*[-n])$.

3.3.3. Свойства преобразования Фурье, основанные на симметрии

Следующие свойства ПФ очень полезны для быстрого нахождения выражений ПФ, а также для разработки тестов для проверки реализованных алгоритмов ПФ.

1. $F\{x^*[n]\} = X^*(e^{-j\omega})$ Докажем:

$$F\{x^*[n]\} = \sum_{n=-\infty}^{\infty} x^*[n]e^{-j\omega n} = \left(\sum_{n=-\infty}^{\infty} x[n]e^{-j(-\omega)n} \right)^* = X^*(e^{-j\omega})$$

2. $F\{x^*[-n]\} = X^*(e^{j\omega})$ Докажем:

$$F\{x^*[-n]\} = \sum_{n=-\infty}^{\infty} x^*[-n]e^{-j\omega n} = \left(\sum_{n=-\infty}^{\infty} x[-n]e^{-j\omega(-n)} \right)^* = X^*(e^{j\omega})$$

3. $F\{\text{Re}\{x[n]\}\} = X_2(e^{j\omega})$, где $X_2(e^{j\omega})$ – чётная составляющая $X(e^{j\omega})$. Докажем:

$$\begin{aligned} F\{\text{Re}\{x[n]\}\} &= \sum_{n=-\infty}^{\infty} \frac{1}{2}(x[n] + x^*[n])e^{-j\omega n} = \\ &= \frac{1}{2}(X(e^{j\omega}) + X^*(e^{-j\omega})) = X_2(e^{j\omega}). \end{aligned}$$

4. $F\{j \text{Im}\{x[n]\}\} = X_1(e^{j\omega})$, где $X_1(e^{j\omega})$ – нечётная составляющая $X(e^{j\omega})$. Докажем:

$$\begin{aligned}
F\{j \operatorname{Im}\{x[n]\}\} &= \sum_{n=-\infty}^{\infty} j \frac{1}{2} (x[n] - x^*[n]) e^{-j\omega n} = \\
&= j \frac{1}{2} (X(e^{j\omega}) - X^*(e^{-j\omega})) = X_1(e^{-j\omega}).
\end{aligned}$$

5. $F\{x_2[n]\} = \operatorname{Re}\{X(e^{j\omega})\}$, где $x_2[n]$ – чётная составляющая $x[n]$.
Докажем:

$$\begin{aligned}
F\{x_2[n]\} &= \sum_{n=-\infty}^{\infty} \frac{1}{2} (x[n] + x^*[-n]) e^{-j\omega n} = \\
&= \frac{1}{2} (X(e^{j\omega}) + X^*(e^{j\omega})) = \operatorname{Re}\{X(e^{j\omega})\}
\end{aligned}$$

6. Не трудно доказать, что $F\{x_1[n]\} = j \operatorname{Im}\{X(e^{j\omega})\}$, где $x_1[n]$ – чётная составляющая $x[n]$.

Обратите внимание на зеркальную симметричность свойств 3 и 5, а также свойств 4 и 6.

Следующие свойства ПФ применимы только к ПФ действительных последовательностей.

$$7. \forall x[n]: \operatorname{Im}\{x[n]\} = 0 \quad X(e^{j\omega}) = X^*(e^{-j\omega}).$$

Свойство 7 является прямым следствием свойства 3 и свойства чётных последовательностей.

$$8. \forall x[n]: \operatorname{Im}\{x[n]\} = 0 \quad \operatorname{Re}\{X(e^{j\omega})\} = \operatorname{Re}\{X(e^{-j\omega})\}.$$

$$9. \forall x[n]: \operatorname{Im}\{x[n]\} = 0 \quad \operatorname{Im}\{X(e^{j\omega})\} = -\operatorname{Im}\{X^*(e^{-j\omega})\}.$$

Для действительных последовательностей амплитудный спектр чётный, а фазовый спектр нечётный. Об этом нам говорят свойства 10 и 11, соответственно

$$10. \forall x[n]: \operatorname{Im}\{x[n]\} = 0 \quad |X(e^{j\omega})| = |X(e^{-j\omega})|.$$

$$11. \forall x[n]: \operatorname{Im}\{x[n]\} = 0 \quad \operatorname{Arg}\{X(e^{j\omega})\} = -\operatorname{Arg}\{X(e^{-j\omega})\}.$$

В качестве иллюстрации рассмотрите графики амплитудного и фазового спектра последовательности $x[n] = 0.9^n u[n]$ (рис. 3.2). Как

мы видим, амплитудный спектр (Magnitude) – чётная функция частоты, а фазовый спектр (Phase) – нечётная.

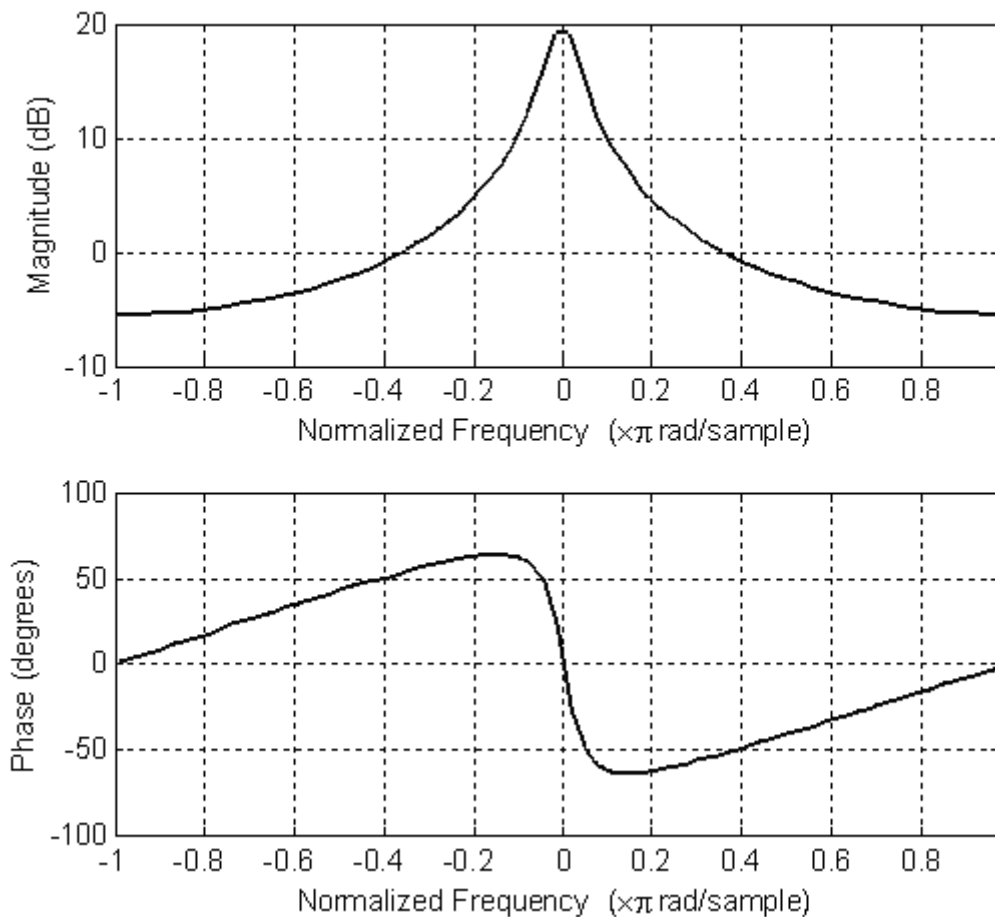


Рис. 3.2. Графики амплитудного и фазового спектра последовательности $x[n] = 0,9^n u[n]$.

3.3.4. Теоремы о преобразовании Фурье дискретных сигналов

1. Преобразование Фурье линейно,

$$F\{ax[n] + by[n]\} = aF\{x[n]\} + bF\{y[n]\}.$$

Докажите это свойство самостоятельно, воспользовавшись просто определением преобразования Фурье (3.5).

2. ПФ последовательности, задержанной на n_d отсчётов,

$$F\{x[n - n_d]\} = e^{-j\omega n_d} X(e^{j\omega}).$$

Это легко доказать, воспользовавшись определением преобразования Фурье:

$$F\{x[n - n_d]\} = \sum_{n=-\infty}^{\infty} x[n - n_d] e^{-j\omega n} = e^{-j\omega n_d} \sum_{n=-\infty}^{\infty} x[n - n_d] e^{-j\omega(n - n_d)} = e^{-j\omega n_d} F\{x[n]\},$$

где последний переход получается заменой $m = n - n_d$ и рассмотрением бесконечной суммы по индексу m .

3. ПФ модулированной последовательности

$$F\{e^{j\omega_0 n} x[n]\} = X(e^{j(\omega - \omega_0)}).$$

Не трудно доказать это свойство, опираясь на приёмы, использованные для доказательства предыдущей теоремы.

4. ПФ свёртки двух последовательностей

$$F\{x[n] * y[n]\} = F\{x[n]\} F\{y[n]\} = X(e^{j\omega}) Y(e^{j\omega}).$$

Докажем это свойство, используя определения свёртки (2.2) и ПФ (3.5):

$$\begin{aligned} F\{x[n] * y[n]\} &= F\left\{\sum_{k=-\infty}^{\infty} x[k] y[n - k]\right\} = \sum_{n=-\infty}^{\infty} \left(\sum_{k=-\infty}^{\infty} x[k] y[n - k]\right) \cdot e^{-j\omega n} \\ &= \sum_{k=-\infty}^{\infty} x[k] \sum_{n=-\infty}^{\infty} y[n - k] \cdot e^{-j\omega n} = \sum_{k=-\infty}^{\infty} x[k] e^{-j\omega k} Y(e^{j\omega}) = Y(e^{j\omega}) \sum_{k=-\infty}^{\infty} x[k] \cdot e^{-j\omega k} = Y(e^{j\omega}) X(e^{j\omega}). \end{aligned}$$

5. ПФ произведения двух последовательностей

$$F\{x[n] y[n]\} = \frac{1}{2\pi} F\{x[n]\} * F\{y[n]\} = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j(\omega - \theta)}) Y(e^{j\theta}) d\theta.$$

6. Теорема Парсеваля, постулирующая сохранение энергии

$$\sum_{n=-\infty}^{\infty} |x[n]|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |X(e^{j\omega})|^2 d\omega.$$

Многие из этих формул легко выводятся на основании свойств и теорем преобразования Фурье.

Т а б л и ц а 3.1

**Выражения для преобразования Фурье
для некоторых базовых цифровых последовательностей**

	Последовательность $x[n]$	Преобразование Фурье $F\{x[n]\}$
1.	$\delta[n]$	1
2.	$\delta[n - n_0]$	$e^{-j\omega n_0}$
3.	1, для $-\infty < n < \infty$	$\sum_{k=-\infty}^{\infty} 2\pi\delta(\omega + 2\pi k)$
4.	$a^n u[n]$, для $ a < 1$	$\frac{1}{1 - ae^{-j\omega}}$
5.	$u[n]$	$\frac{1}{1 - e^{-j\omega}} + \sum_{k=-\infty}^{\infty} \pi\delta(\omega + 2\pi k)$
6.	$\frac{\sin \omega_c n}{\pi n}$	$X(e^{j\omega}) = \begin{cases} 1, & \omega < \omega_c, \\ 0, & \omega_c < \omega < \pi. \end{cases}$
7.	$x[n]$	$\frac{\sin\left(\frac{\omega(M+1)}{2}\right)}{\sin\left(\frac{\omega}{2}\right)} e^{-j\omega M/2}$
8.	$e^{j\omega_0 n}$	$\sum_{k=-\infty}^{\infty} 2\pi\delta(\omega - \omega_0 + 2\pi k)$

Пример 1

$$F\{\delta[n - n_0]\} = \text{п.3.3.4 теорема 2} = e^{-j\omega n_0} F\{\delta[n]\} = e^{-j\omega n_0} 1 = e^{-j\omega n_0}.$$

Пример 2

$$\begin{aligned}
 F\{e^{j\omega_0 n}\} &= F\{e^{j\omega_0 n} 1\} = \text{п. 3.3.4 теорема 3 и табл. 3.1, строка 3} \\
 &= \sum_{k=-\infty}^{\infty} 2\pi\delta(\omega + 2\pi k - \omega_0).
 \end{aligned}$$

Пример 3

$$F\{a^n u[n]\} = \sum_{n=-\infty}^{\infty} a^n u[n] e^{-j\omega n} = \sum_{n=0}^{\infty} a^n e^{-j\omega n} = \sum_{n=0}^{\infty} (ae^{-j\omega})^n = \frac{1}{1 - ae^{-j\omega}}, \quad |ae^{-j\omega}| = |a| < 1$$

Пример 4 (вычисления обратного преобразования Фурье)

$$X(e^{j\omega}) = \begin{cases} e^{-j\omega n_d}, & \omega_c < |\omega|, \\ 0, & \omega_c < |\omega| \leq \pi. \end{cases}$$

Графическое изображение соответствующих амплитудного и фазовых спектров приведено на рис. 3.3.

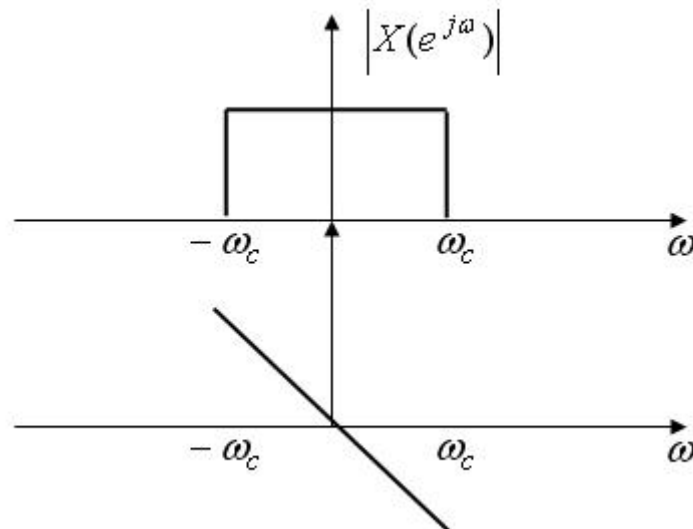


Рис. 3.3. Амплитудный и фазовый спектр обратного преобразования Фурье

Найдем последовательность, имеющую такое ПФ. Для этого воспользуемся выражением для обратного преобразования Фурье (3.4):

$$\begin{aligned}
 x[n] &= \frac{1}{2\pi} \int_{-\omega_c}^{\omega_c} e^{-j\omega n_d} e^{-j\omega n} d\omega = \frac{1}{2\pi} \int_{-\omega_c}^{\omega_c} e^{-j\omega(n-n_d)} d\omega = \\
 &= \frac{1}{2\pi} \left(\left. \frac{1}{-j(n-n_d)} e^{-j\omega(n-n_d)} \right|_{\omega=\omega_c} - \left. \frac{1}{-j(n-n_d)} e^{-j\omega(n-n_d)} \right|_{\omega=-\omega_c} \right) = \\
 &= \frac{e^{j\omega_c(n-n_d)} - e^{-j\omega_c(n-n_d)}}{j2\pi(n-n_d)} = \frac{\sin(\omega_c(n-n_d))}{\pi(n-n_d)}
 \end{aligned}$$

График этой последовательности для $n_d = 5$ приведён на рис. 3.4.

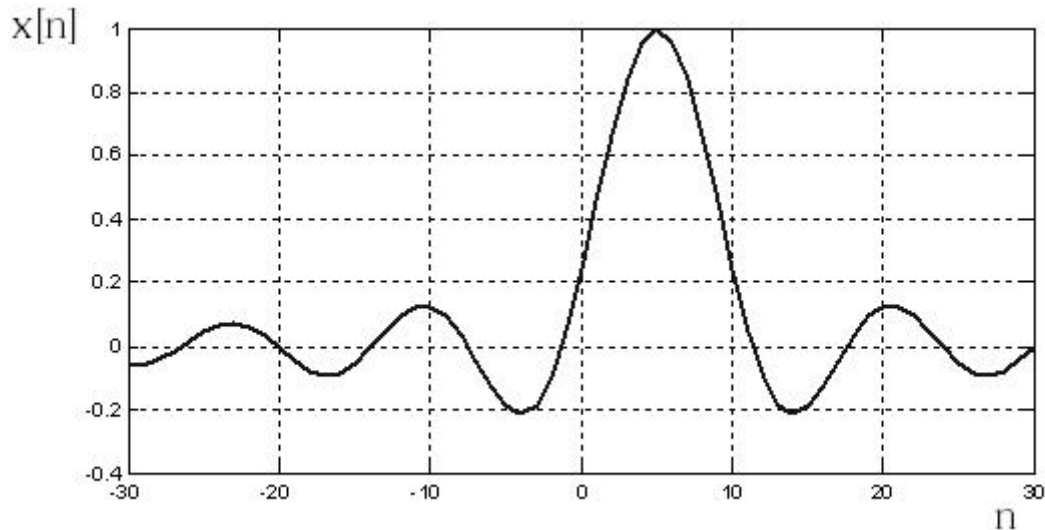


Рис. 3.4. Последовательность, имеющая заданное ПФ

Применение прямого (ППФ) и обратного (ОПФ) преобразования Фурье является наиболее широко известным методом исследования сигналов. ППФ позволяет определить диапазон частот, присутствующих в сигнале, и перейти от временных характеристик сигнала к его частотным характеристикам. Преобразование Фурье сигнала является разложением по гармоническим функциям всех частот в диапазоне от $-\infty$ до ∞ .

Поведение сигнала в частотной области несет важную информацию о сигнале. В частности, ППФ необходимо для эффективного выделения полезного сигнала из смеси сигнала и шума

(фильтрации). Однако, прямое преобразование Фурье имеет низкое разрешение по времени: параметр t является постоянным. Аналогично, обратное преобразование Фурье имеет столь же низкое разрешение по частоте: f является константой. Иными словами, ППФ и ОПФ не дают возможности установить время присутствия данной частоты в сигнале и частотный состав сигнала в заданный момент времени. Если, при этом, спектр сигнала не изменяется во времени, то исследование с помощью преобразования Фурье дает хорошие результаты.

Однако большую группу сигналов составляют сигналы, имеющие ряд нестационарных особенностей: краевые эффекты, паузы, шумы коммутации, выбросы, разрывы и т. д. Обнаружение нарушений непрерывности и гладкости сигнала нередко составляет основную задачу исследования, поэтому для изучения нестационарных сигналов требуется локализация во времени частотно-временного анализа Фурье.

3.4. КРАТКОВРЕМЕННОЕ ПРЕОБРАЗОВАНИЕ ФУРЬЕ

Метод локализации преобразования Фурье, предложенный в 1946 г. Д. Габором состоял в разделении исходного сигнала на последовательность временных окон. Далее выполняется ППФ каждого временного окна:

$$s(\omega, t) = \int_R s(u) g(t - u) e^{-i\omega u} du .$$

Таким образом, можно получить информацию о частотах, содержащихся лишь в пределах заданной временной области сигнала.

Соотношение между значением частоты и моментом времени, когда эта частота присутствовала в сигнале, определяется с точностью до длительности окна, что и ограничивает возможности метода.

3.5. РАЗЛОЖЕНИЕ В БИОРТОГОНАЛЬНОМ БАЗИСЕ

На основании взаимного расположения базисных функций в пространстве, среди существующих базисов выделяют ортогональные и неортогональные. Взаимное расположение базисных функций определяется по их скалярному произведению для ортогонального базиса $\{v_i\}$

$$\langle v_i, v_j \rangle = k_{ij} \delta(i - j), \text{ где } \delta(i - j) = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \quad (3.6)$$

при $k_{ij} = 1$ для любых i, j базис является ортонормированным (рис. 3.5, 3.6 а).

Несмотря на то, что ортогональные базисы имеют очевидные достоинства в приложении к задачам обработки сигнала, также используется ряд неортогональных базисов. Неортогональный или *биортогональный* базис (встречается также термин *дуальный* базис) представляет собой два набора базисных функций (векторов) $\{v_i\}$ и $\{u_j\}$, связанных отношением взаимной ортогональности. Скалярное произведение биортогональных базисных функций имеет вид:

$$\langle v_i, v_j \rangle = k_{ij} \delta(i - j), \text{ где } \delta(i - j) = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \quad (3.7)$$

при $k_{ij}=1$ для любых i, j базис является биортонормированным (Рис. 3.5, 3.6 б).

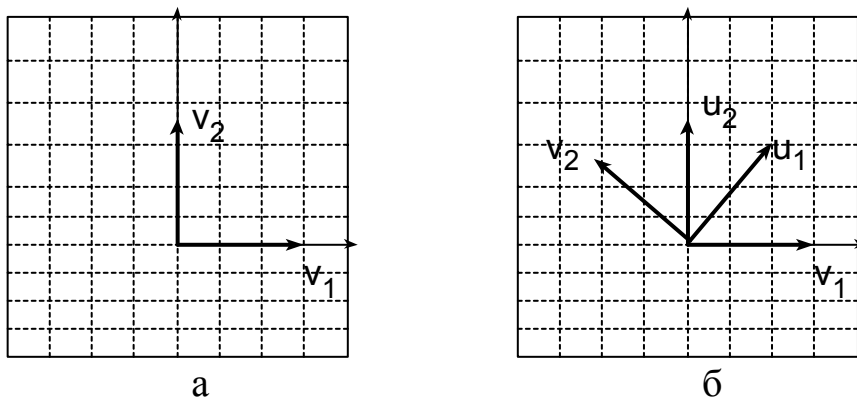


Рис. 3.5. а – ортогональный базис пространства $\mathbb{R}^2$,
б – биортогональный базис пространства $\mathbb{R}^2$.

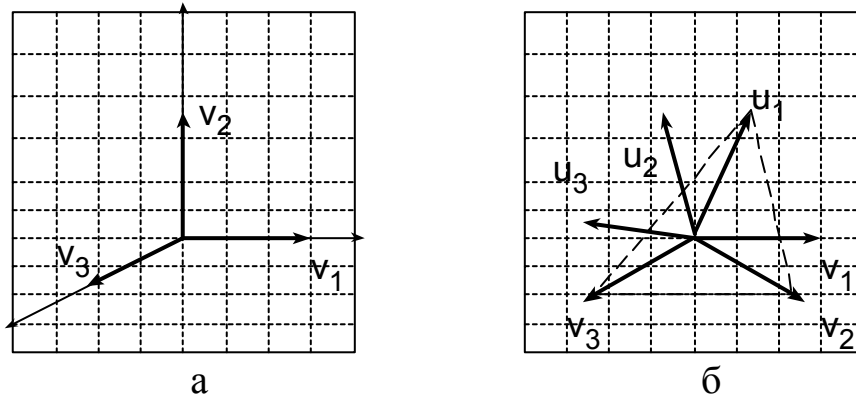


Рис. 3.6. а – ортогональный базис пространства $\mathbb{R}^3$,
б – биортогональный базис пространства $\mathbb{R}^3$

Для наглядности в трехмерном случае на рисунке 3.6, б показана ортогональность вектора u_1 плоскости, т.е. двумерному пространству, образованному векторами v_2, v_3 .

Для произвольного n -мерного пространства каждая компонента базиса v_i ортогональна $(n-1)$ -мерному подпространству $\{u_1, u_2, \dots, u_n \setminus u_j, j=i\}$.

Так как и $\{v_i\}$, и $\{u_j\}$ являются базисами некоторого пространства $\mathbb{R}^n$, то, согласно (3.6), f представима линейной комбинацией базиса $\{v_i\}$ и базиса $\{u_j\}$

$$f = \langle f, v_1 \rangle v_1 + \langle f, v_2 \rangle v_2 = \langle (\langle f, v_1 \rangle v_1 + \langle f, v_2 \rangle v_2), u_1 \rangle u_1 + \langle (\langle f, v_1 \rangle v_1 + \langle f, v_2 \rangle v_2), u_2 \rangle u_2.$$

Учитывая дистрибутивное свойство скалярного произведения, раскроем скобки:

$$f = \langle f, v_1 \rangle v_1 \cdot u_1 + \langle f, v_2 \rangle v_2 \cdot u_1 + \langle f, v_1 \rangle v_1 \cdot u_2 + \langle f, v_2 \rangle v_2 \cdot u_2,$$

согласно (3.4) исключим нулевые слагаемые:

$$f = \langle f, v_1 \rangle v_1 \cdot u_1 + \langle f, v_2 \rangle v_2 \cdot u_2 = \langle f, v_1 \rangle u_1 + \langle f, v_2 \rangle u_2. \quad (3.8)$$

Уравнение (3.8) иллюстрирует важнейшее свойство биортогональных базисов: сигнал, будучи разложен в *одном неортогональном базисе*, может быть восстановлен в *другом*, при условии, что эти базисы биортогональны.

3.6. ИСПОЛЬЗОВАНИЕ ПРЕОБРАЗОВАНИЯ ФУРЬЕ ДЛЯ АНАЛИЗА ПРОЦЕССОВ АНАЛОГОВОГО И ЦИФРОВОГО И ЦИФРО-АНАЛОГОВОГО ПРЕОБРАЗОВАНИЙ. ТЕОРЕМА ШЕННОНА-КОТЕЛЬНИКОВА-НАЙКВИСТА

Алгоритмическая блок-схема устройства дискретизации по времени и ее математическое описание приведены в раздее 1.3, рис. 1.5.

Устройство дискретизации сигнала состоит из устройства выделения сигнала, контролируемого сигналом-стробом $s_T(t)$, и устройства преобразования импульсов с бесконечно большой амплитудой и бесконечно короткой длительностью в последовательность цифровых импульсов с амплитудой, равной значению сигнала в моменты времени nT . Это устройство изображено на схеме в виде интегрирующего звена.

Дискретизирующий строб-сигнал $s_T(t)$ может быть представлен как сумма цифровых единичных импульсов с различной задержкой, т.е. $s_T(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT)$. Тогда дискретизированная по времени версия сигнала $x(t)$ может быть записана как

$$x_d(t) = x(t)s_T(t) = x(t) \sum_{n=-\infty}^{\infty} \delta(t - nT) = \sum_{n=-\infty}^{\infty} \delta(t - nT)x(nT).$$

Дискретизированный сигнал $x_d(t)$ и цифровая последовательность $x[n] = x(nT)$ разные вещи!

3.6.1. Описание процесса дискретизации сигнала в частотной области при помощи преобразования Фурье

Во временной области дискретизированный $x_d(t)$ и аналоговый сигнал $x(t)$ связаны между собой соотношением

$x_d(t) = x(t)s_T(t) = x(t) \sum_{n=-\infty}^{\infty} \delta(t - nT)$. Таким образом, в соответствии со

свойством преобразования Фурье $F\{x_1(t)x_2(t)\} = \frac{1}{2\pi} X_1(\Omega) * X_2(\Omega)$ мы можем записать следующее выражение для ПФ $X_d(j\Omega)$ дискретизированного сигнала $x_d(t)$:

$$X_d(j\Omega) = X(j\Omega) * F\left\{\sum_{n=-\infty}^{\infty} \delta(t - nT)\right\}.$$

В соответствии с [10], ПФ $S(j\Omega)$ последовательности дельта импульсов определяется в соответствии со следующим выражением

$$S(j\Omega) = F\left\{\sum_{n=-\infty}^{\infty} \delta(t - nT)\right\} = \frac{2\pi}{T} \sum_{k=-\infty}^{\infty} \delta\left(\Omega - \frac{2\pi}{T}k\right)$$

и, как мы видим, представляет собой последовательность дельта-импульсов вдоль частотной оси с интервалом в $\Omega_d = \frac{2\pi}{T}$. Как нетрудно заметить, $\Omega_d = 2\pi f_d$ и называется круговой частотой дискретизации ("круговой" добавляется из-за умножения на 2π).

Свёртка с последовательностью дельта-импульсов – это просто "копирование" одного сигнала по позициям дельта-импульсов, значит, ПФ дискретизированного сигнала:

$$X_d(j\Omega) = \frac{1}{T} \sum_{k=-\infty}^{\infty} X(j(\Omega - k \frac{2\pi}{T})) = \frac{1}{T} \sum_{k=-\infty}^{\infty} X(j(\Omega - k\Omega_d)).$$

Графически это может быть изображено в следующем виде (рис. 3.7).

Ω_N – максимальная частота, присутствующая в ПФ сигнала $x(t)$, носит название частоты Найквиста.

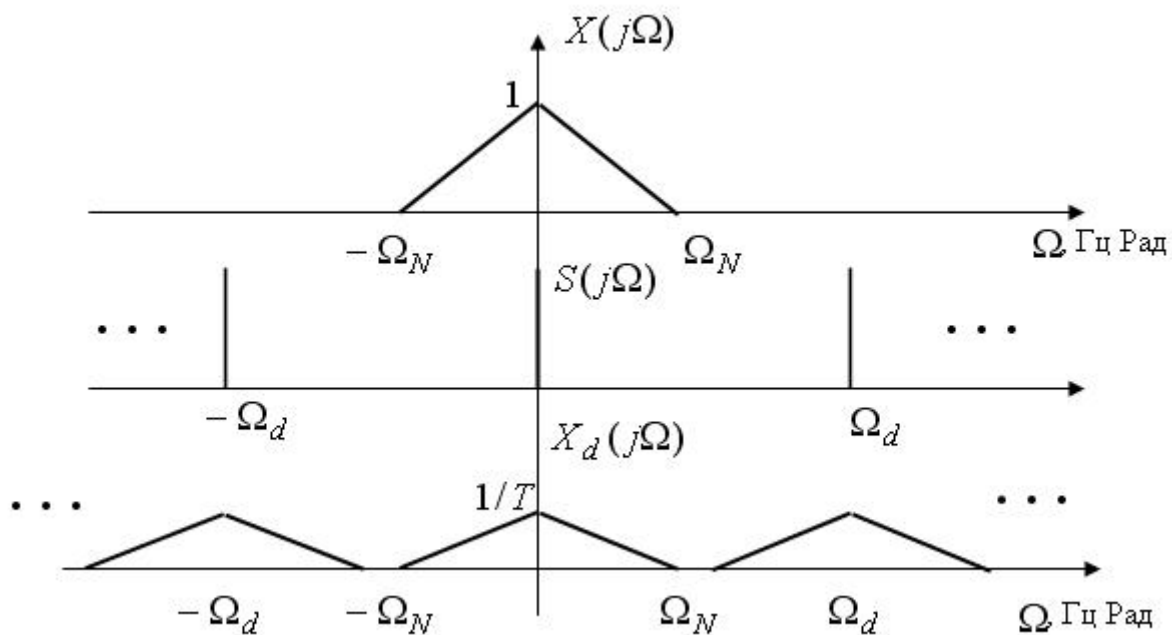


Рис. 3.7. Свертка с последовательностью дельта-импульсов

Таким образом, ПФ дискретизированного сигнала состоит из суперпозиции бесконечного числа сдвинутых (периодически повторяющихся) копий ПФ входного сигнала. Период повторения равен круговой частоте дискретизации $\Omega_d = 2\pi f_d = \frac{2\pi}{T}$, где T – интервал дискретизации.

3.6.2. Восстановление непрерывного сигнала

Преобразование Фурье сигнала и сам сигнал связаны между собой однозначным соответствием. Таким образом, если нам удастся при помощи некоторой операции получить из спектра $X_d(j\Omega)$ дискретизированного сигнала спектр $X(j\Omega)$ исходного сигнала, то мы можем считать задачу восстановления исходного сигнала решенной.

Рассмотрим идеальный фильтр нижних частот с передаточной характеристикой $H(j\Omega)$, изображённой на рис. 3.8. Обратите внимание на коэффициент усиления фильтра, равный T .

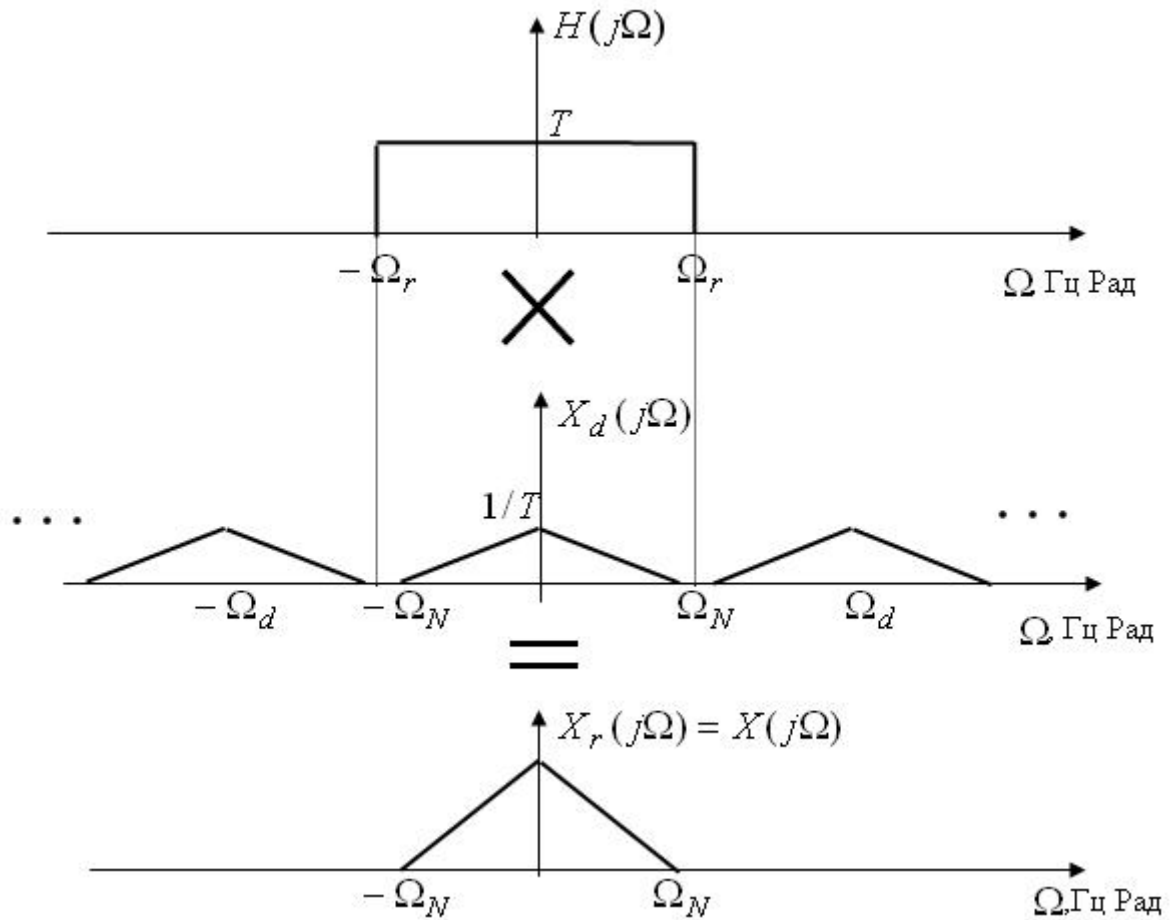


Рис. 3.8. Идеальный фильтр нижних частот

Если дискретизированный сигнал пропустить через такой фильтр, то ПФ результирующего сигнала $X_r(j\Omega)$ будет равно произведению ПФ дискретизированного сигнала и передаточной характеристики фильтра $H(j\Omega)$. Как нетрудно видеть, такая процедура приведёт к полному восстановлению исходного сигнала в случае, если $\Omega_d \geq 2\Omega_N$. Процесс восстановления исходного сигнала проиллюстрирован на рис. 3.8. Частота среза восстанавливающего фильтра Ω_r должна лежать в интервале $(\Omega_N, \Omega_d - \Omega_N)$. Удобный выбор – $\Omega_r = \frac{\Omega_d}{2} = \frac{\pi}{T}$.

Во временной области такое восстановление сигнала есть свёртка дискретизированного сигнала с импульсной характеристикой

идеального фильтра низких частот с частотой среза Ω_N . Другими словами,

$$h_r(t) = \frac{\sin(\Omega_r t)}{\Omega_r t} = \frac{\sin(\pi \cdot t/T)}{\pi \cdot t/T},$$

$$x_r(t) = \sum_{n=-\infty}^{\infty} x[n] \frac{\sin\left[\frac{\pi(t-nT)}{T}\right]}{\pi(t-nT)/T}.$$

Нетрудно видеть, что $h_r(t) = 0$ для $t = \pm T, \pm 2T, \dots, \pm nT$, а $h_r(0) = 1$. Таким образом, восстановленный сигнал равен дискретизированному сигналу в моменты времени дискретизации $t = nT$.

Решение задачи восстановления сигнала по дискретным отсчетам, т.е. интерполяции дискретного сигнала непрерывной функцией, позволяет получить разложение в ряд Шеннона-Котельникова в явной форме. Задача восстановления в частотной и временной области соответственно имеет вид [10]:

$$\Phi(f) = \Phi_{\text{дискр}}(f) \cdot \Pi(f, f_{\text{дискр}}) \Leftrightarrow \varphi(t) = \varphi_{\text{дискр}}(t) * \frac{\sin(\pi f t)}{\pi \cdot t}$$

Раскрывая операцию дискретной свертки, имеем:

$$\begin{aligned} \varphi(t) &= \sum_{n=-\infty}^{\infty} \varphi(nT_{\text{дискр}}) T_{\text{дискр}} \cdot \frac{\sin \pi f_{\text{дискр}}(t - nT_{\text{дискр}})}{\pi(t - nT_{\text{дискр}})} = \\ &= \sum_{n=-\infty}^{\infty} \varphi(nT_{\text{дискр}}) \cdot \frac{\sin(\pi f_{\text{дискр}} t - \pi n)}{\pi f_{\text{дискр}} t - \pi n} \end{aligned}$$

т.е. во временной области в моменты, отмеченные дельта-импульсами, ставятся функции Шеннона-Котельникова.

Базисные функции разложения по Шеннону-Котельникову $\frac{\sin(\pi f_{\text{дискр}} t - \pi n)}{\pi f_{\text{дискр}} t - \pi n}$ представляют собой наиболее известный случай,

когда разложение/восстановление сигнала ведется не с помощью кусочно-постоянных, кусочно-линейных или гармонических

функций. За пределами нескольких периодов дискретизации базисные функции можно считать нулевыми, и базис, соответственно, практически конечным, т.е. имеющим тенденцию к локализации [6].

Теоретическая возможность восстановления сигнала всего по двум отсчетам на период дискретизации без погрешности объясняется как следствие более высокой корреляции функций $\frac{\sin(\pi f_{\text{дискр}} t - \pi n)}{\pi f_{\text{дискр}} t - \pi n}$ с

интерполируемым сигналом, чем это наблюдается у перечисленных выше функций. Разложение сигнала в ряд Шеннона-Котельникова, наряду с преобразованием Фурье, является одним из популярных методов анализа сигналов.

Минимально допустимая частота дискретизации Ω_d сигнала с частотой Найквиста Ω_N равна $\Omega_d = 2\Omega_N$. В случае, когда $\Omega_d < 2\Omega_N$, происходит наложение спектров, и восстановление исходного сигнала не возможно.

Теорема Шеннона-Котельникова-Найквиста

Любой сигнал $x(t)$, спектр которого не содержит составляющих с частотами выше некоторого значения $\Omega_g = 2\pi f_g$, может быть без потерь информации представлен своими дискретными отсчетами $\{x(nT)\}$, взятыми с интервалом T , удовлетворяющим следующему неравенству $T = \frac{1}{f_g} = \frac{2\pi}{\omega_g}$.

3.6.3. Спектр цифровой последовательности

До сих пор мы рассматривали ПФ амплитудно-модулированной последовательности дельта-импульсов, получающейся из исходного сигнала посредством умножения на последовательность дельта-импульсов единичной амплитуды (см. раздел 1.3, рис. 1.5). Попробуем найти ПФ цифровой последовательности $x[n]$ на выходе устройства аналого-цифрового преобразования.

Применим преобразование Фурье к обеим частям следующего уравнения

$$x_d(t) = x(t)s_T(t) = x(t) \sum_{n=-\infty}^{\infty} \delta(t - nT) = \sum_{n=-\infty}^{\infty} \delta(t - nT)x(nT).$$

$$X_d(j\Omega) = \sum_{n=-\infty}^{\infty} x(nT)e^{-j\Omega Tn} = \sum_{n=-\infty}^{\infty} x[n]e^{-j\Omega Tn}. \text{ Так как } x[n] = x(nT).$$

Выражение, которое нас интересует (ПФ цифровой последовательности), выглядит как $X(e^{j\omega}) = \sum_{n=-\infty}^{\infty} x[n]e^{-j\omega n}$. Таким

образом, сопоставив два последних выражения и заметив, что для $\omega = \Omega T$ эти два выражения равны, мы можем записать

$X_d(j\Omega) = X(e^{j\omega})|_{\omega=\Omega T} = X(e^{j\Omega T})$. Теперь на основании ранее найденного выражения для $X_d(j\Omega)$ мы можем записать

$$X(e^{j\Omega T}) = X_d(j\Omega) = \frac{1}{T} \sum_{k=-\infty}^{\infty} X(j(\omega - k\Omega_d)).$$

Переходя к безразмерной частоте $\omega = \Omega T$, мы получим

$$X(e^{j\omega}) = X_d\left(j\frac{\omega}{T}\right) = \frac{1}{T} \sum_{k=-\infty}^{\infty} X\left(j\left(\frac{\omega}{T} - k\frac{2\pi}{T}\right)\right).$$

Однако для практического применения и построения устройств обработки сигнала, основанных на данном принципе, является препятствием необходимость получения полной информации о сигнале.

3.7. Z-ПРЕОБРАЗОВАНИЕ

3.7.1. Ограничения на использование преобразования Фурье для анализа систем и сигналов

ПФ существует далеко не для всех последовательностей. Достаточным условием существования ПФ последовательности является её абсолютная суммируемость. Иными словами, у последовательности $x[n]$ существует ПФ, если $\sum_{n=-\infty}^{\infty} |x[n]| < \infty$.

Для любых систем с конечной импульсной характеристикой (КИХ), равно как и для сигналов с конечным числом ненулевых

отсчётов, ПФ существует просто в силу ограниченности числа элементов в сумме преобразования Фурье. Однако, среди систем с БИХ преобразование Фурье существует лишь для небольшого множества таковых с импульсной характеристикой, удовлетворяющей условию абсолютной суммируемости. Таким образом, возникает необходимость использования аппарата Z-преобразования.

3.7.2. Затухающие комплексные экспоненты – собственные функции линейных систем

Проверим утверждение в заголовке

$$\begin{aligned} \sum_{k=-\infty}^{\infty} h[k]x[n-k] \Big|_{x[n]=r^n e^{j\omega n}} &= \sum_{k=-\infty}^{\infty} h[k]r^{n-k} e^{j\omega(n-k)} = \\ &= r^n e^{j\omega n} \sum_{k=-\infty}^{\infty} h[k]r^{-k} e^{-j\omega k} = r^n e^{j\omega n} H(re^{-j\omega}) \end{aligned}$$

Таким образом, мы доказали, что затухающие комплексные экспоненты являются собственными функциями ЛС. Соответствующие собственные числа вычисляются в соответствии с выражением $H(re^{j\omega}) = \sum_{k=-\infty}^{\infty} h[k](re^{j\omega})^k$.

Сравнивая с выражением для Фурье преобразования $H(e^{j\omega}) = \sum_{k=-\infty}^{\infty} h[k]e^{-j\omega k}$, мы видим, что теперь собственные числа зависят от двух переменных r, ω или, проще говоря, являются функциями комплексного числа $z = re^{-j\omega}$, определённого на всей комплексной плоскости. Заменяя $z = re^{j\omega}$, мы получим более компактное выражение для нашего преобразования (Z-преобразование):

$$H(z) = \sum_{k=-\infty}^{\infty} h[k]z^{-k}. \quad (3.9)$$

Соответственно, так же как и Фурье преобразование, Z-преобразование (3.9) применимо не только к импульсным характеристикам, но и к сигналам.

Из дальнейшего сравнения следует, что Фурье преобразование получается из Z-преобразования при $r=1$, т.е. на единичной окружности. Также мы можем видеть, что Z-преобразование последовательности $h[n]$ равно Фурье преобразованию последовательности $h[n]r^{-n}$. Выбор r позволяет обеспечить сходимость суммы Z-преобразования.

3.7.3. Z-преобразование

Z—преобразование сигнала $x[n]$ выражается следующим образом:

$$X(z) = \sum_{k=-\infty}^{\infty} x[n]z^{-n}. \quad (3.10)$$

Также как и ПФ, Z—преобразование существует не для всех последовательностей $x[n]$ и не для всех значений r .

В соответствии с нашим наблюдением о сходстве выражений ПФ и Z-преобразования и условия существования ПФ мы заключаем, что Z-преобразование последовательности $x[n]$ существует, для таких

r , если $\sum_{n=-\infty}^{\infty} |x[n]r^{-n}| < \infty$. Это расширяет набор последовательностей,

для которых существует Z-преобразование, по сравнению с ПФ.

Последовательность $x[n] = u[n]$ не является абсолютно суммируемой и поэтому её ПФ не существует. С другой стороны, последовательность $u[n]r^{-n}$ абсолютно суммируема при $r > 1$ и, таким образом, обуславливает существование Z-преобразования исходной последовательности $x[n]$, равного

$$X(z) = Z\{x[n]\} = Z\{u[n]\} = \sum_{n=-\infty}^{\infty} u[n]z^{-n} = \sum_{n=0}^{\infty} z^{-n} = \frac{1}{1-z^{-1}},$$

для $r > 1$. Поскольку $r = |z|$, то принято записывать область сходимости как $|z| > 1$. На рис. 3.9 изображена область сходимости z -преобразования $x[n] = u[n]$ в виде заштрихованной области z -плоскости.

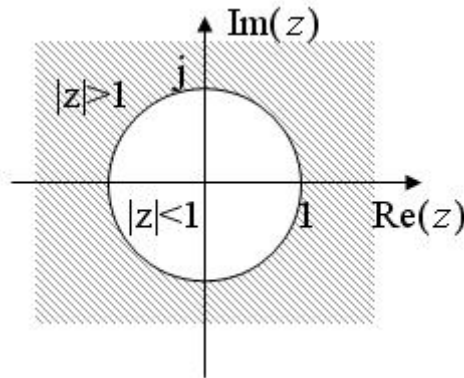


Рис. 3.9. Область сходимости z -преобразования

В отличие от ПФ, для однозначного соответствия между исходной последовательностью и Z -преобразованием необходимо указать область сходимости. Например, последовательность $y[n] = -u[-n-1]$ имеет ту же формулу Z -преобразования, что и последовательность $x[n]$, однако область сходимости – иная, $|z| < 1$.

Пример 1

Вычислим z -преобразование последовательности $x[n] = a^n u[n]$.

$$X(z) = Z\{x[n]\} = Z\{a^n u[n]\} = \sum_{n=-\infty}^{\infty} a^n u[n] z^{-n} = \sum_{n=0}^{\infty} (az^{-1})^n.$$

Для получения конечной формулы, нам необходимо указать область сходимости (рис. 3.10).

Последнее выражение – бесконечная сумма геометрической последовательности с комплексными членами. Она сходится к

$\frac{1}{1 - az^{-1}}$ в случае, если $|az^{-1}| = |a||z|^{-1} < 1$ или, иными словами, когда $|z| > |a|$. Таким образом, наш окончательный ответ

$$X(z) = Z\{a^n u[n]\} = \frac{1}{1 - az^{-1}}, \quad |z| > |a|.$$

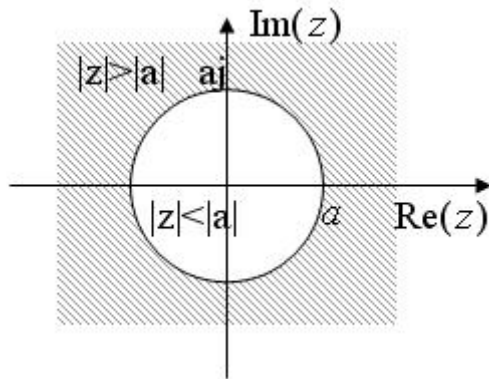


Рис. 3.10. Область сходимости z-преобразования для примера 1

Пример 2

Вычислим z-преобразование последовательности

$$x[n] = -a^n u[-n-1].$$

$$\begin{aligned} X(z) &= Z\{x[n]\} = Z\{-a^n u[-n-1]\} = - \sum_{n=-\infty}^{\infty} a^n u[-n-1] z^{-n} = - \sum_{n=-\infty}^{-1} (az^{-1})^n = \\ &= - \sum_{n=1}^{\infty} (az^{-1})^{-n} = 1 - \sum_{n=0}^{\infty} (a^{-1}z)^n. \end{aligned}$$

Для получения конечной формулы, нам необходимо указать область сходимости. Последнее выражение – бесконечная сумма геометрической последовательности с комплексными членами. Она сходится к $1 - \frac{1}{1 - a^{-1}z} = \frac{1}{1 - az^{-1}}$ в случае, если $|a^{-1}z| = |a|^{-1}|z| < 1$ или, иными словами, когда $|z| < |a|$. Таким образом, наш окончательный ответ $X(z) = Z\{-a^n u[-n-1]\} = \frac{1}{1 - az^{-1}}, |z| < |a|$.

Достаточно очевидно (докажите сами), что Z-преобразование линейно.

Пример 3

Воспользовавшись свойством линейности, найдём Z-преобразование следующей суммы

$$Z\left\{\left(\frac{1}{2}\right)^n u[n] + \left(-\frac{1}{3}\right)^n u[n]\right\} = Z\left\{\left(\frac{1}{2}\right)^n u[n]\right\} + Z\left\{\left(-\frac{1}{3}\right)^n u[n]\right\}.$$

Теперь воспользуемся результатами предыдущих примеров, заменив a на $1/2$ и на $1/3$. Получим

$$Z\left\{\left(\frac{1}{2}\right)^n u[n]\right\} + Z\left\{\left(-\frac{1}{3}\right)^n u[n]\right\} = \frac{1}{1 - \frac{1}{2}z^{-1}} + \frac{1}{1 + \frac{1}{3}z^{-1}}.$$

Область сходимости – пересечение двух областей сходимости (рис. 3.11), $|z| > \frac{1}{2}$ и $|z| > \frac{1}{3}$, что в итоге соответствует $|z| > 1/2$.

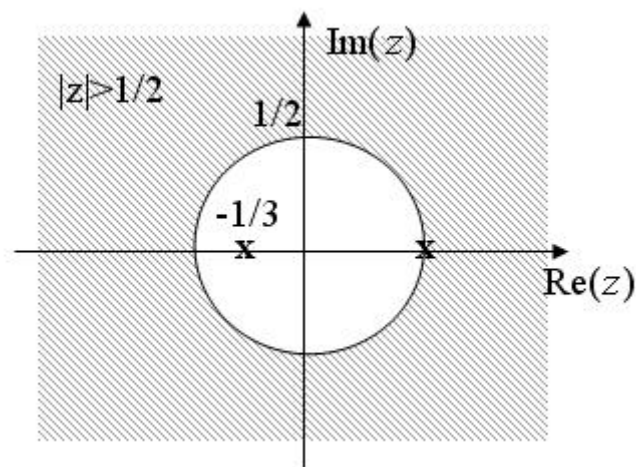


Рис. 3.11. Область сходимости z -преобразования для примера 3

Крестиками на рисунке обозначены полюса комплексной функции комплексного аргумента $\frac{1}{1 - \frac{1}{2}z^{-1}} + \frac{1}{1 + \frac{1}{3}z^{-1}}$. Область

сходимости не должна включать в себя полюса, так как, по определению, в полюсе функция стремится к бесконечности.

Пример 4

Вычислим Z -преобразование следующей последовательности:

$$x[n] = \begin{cases} a^n, & 0 \leq n \leq N-1, \\ 0, & \dots \end{cases} \quad X(z) = \sum_{n=0}^{N-1} a^n z^{-n} = \sum_{n=0}^{N-1} (az^{-1})^{-n}, \quad \text{используя}$$

формулу для суммы конечной геометрической прогрессии с показателем az^{-1} , получим следующий ответ

$$X(z) = \frac{1 - (az^{-1})^N}{1 - az^{-1}} = \frac{1}{z^{N-1}} \frac{z^N - a^N}{z - a}, \text{ область сходимости } |z| > 0$$

из-за полюса $N - 1$ -го порядка в нуле!

Диаграмма нулей ($N - 1$ ноль в $z_n = ae^{j\frac{2\pi k}{N}}, k > 0$) и полюсов (1 полюс в $z_0 = 0$) для $N = 8$ приведена на рис. 3.14.

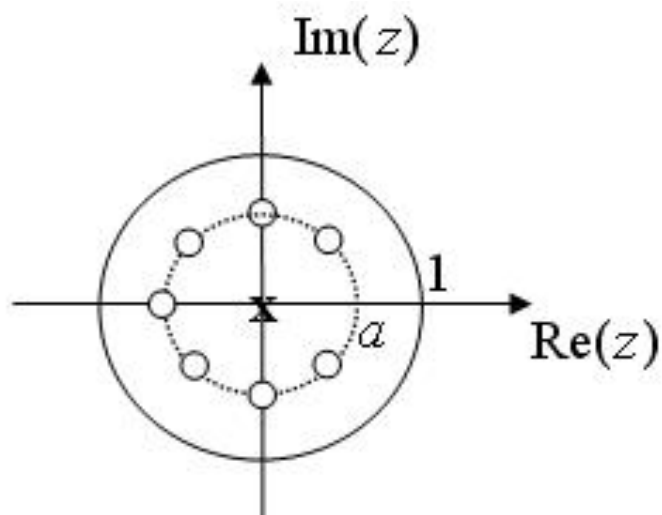


Рис. 3.12. Диаграмма нулей и полюсов для примера 4.

Т а б л и ц а 3.2

Пары Z-преобразований

Последовательность $x[n]$	Z-преобразование $X(z)$	Область сходимости
$\delta[n]$	1	$\forall z$
$u[n]$	$\frac{1}{1 - z^{-1}}$	$ z > 1$
$-u[-n - 1]$	$\frac{1}{1 - z^{-1}}$	$ z < 1$
$\delta[n - m]$	z^{-m}	$\forall z$, за исключением $z = 0$ (если $m > 0$) или $z = \infty$ (если $m < 0$)

$a^n u[n]$	$\frac{1}{1 - az^{-1}}$	$ z > a $
$-a^n u[-n-1]$	$\frac{1}{1 - az^{-1}}$	$ z < a $
$na^n u[n]$	$\frac{az^{-1}}{(1 - az^{-1})^2}$	$ z > a $
$-na^n u[-n-1]$	$\frac{az^{-1}}{(1 - az^{-1})^2}$	$ z < a $
$\cos(\omega \cdot n) \cdot u[n]$	$\frac{1 - \cos(\omega)z^{-1}}{1 - 2\cos(\omega)z^{-1} + z^{-2}}$	$ z > 1$
$\sin(\omega \cdot n) \cdot u[n]$	$\frac{\sin(\omega)z^{-1}}{1 - 2\cos(\omega)z^{-1} + z^{-2}}$	$ z > 1$
$(r^n \sin(\omega n)) \cdot u[n]$	$\frac{r \sin(\omega)z^{-1}}{1 - 2r \cos(\omega)z^{-1} + r^2 z^{-2}}$	$ z > r$
$(r^n \cos(\omega n)) \cdot u[n]$	$\frac{1 - r \cos(\omega)z^{-1}}{1 - 2r \cos(\omega)z^{-1} + r^2 z^{-2}}$	$ z > r$
$\begin{cases} a^n, 0 \leq n \leq N-1, \\ 0, n \notin [0, N-1] \end{cases}$	$\frac{1 - a^N z^{-N}}{1 - az^{-1}}$	$ z > 0$

3.7.4. Обратное Z-преобразование

1. Используя табл. 3.2

$$a^n u[n] \xleftrightarrow{Z} \frac{1}{1 - az^{-1}}, \quad |z| > |a|.$$

2. Используя деление и разложение на простые дроби

$$X(z) = \frac{b_0 \prod_{k=1}^M (1 - c_k z^{-1})}{a_0 \prod_{k=1}^N (1 - d_k z^{-1})} = \sum_{r=0}^{M-N} B_r z^{-r} + \sum_{k=1}^N \frac{A_k}{1 - d_k z^{-1}}$$

Коэффициенты $A_k = (1 - d_k z^{-1})X(z) \Big|_{z=d_k}$, а элемент суммы

$\sum_{r=0}^{M-N} B_r z^{-r}$ определяется при помощи деления полиномов.

3.7.5. Свойства Z-преобразования

1. Линейность $ax[n] + by[n] \xleftrightarrow{Z} aX(z) + bX(z)$,
2. Сдвиг времени: $x[n-d] \xleftrightarrow{Z} z^{-d} X(z)$, из/в области сходимости исключается/включается $z = 0$ или $z = \infty$, в зависимости от знака d .
3. Умножение на экспоненту $z_0^n x[n] \xleftrightarrow{Z} X(z/z_0)$, граница области сходимости $|z_0| R_x$.
4. Обращение времени $x[-n] \xleftrightarrow{Z} X(1/z)$, граница области сходимости $1/R_x$.
5. Теорема о начальном значении $x[n] = 0 \forall n < 0 \Rightarrow x[0] = \lim_{z \rightarrow \infty} X(z)$.
6. Свёртка двух последовательностей $x[n] * y[n] \xleftrightarrow{Z} X(z)Y(z)$, области сходимости – пересечение области сходимости последовательностей $x[n]$ и $y[n]$.
7. Дифференцирование $nx[n] \leftrightarrow -z \frac{d}{dz} X(z)$, из/в области сходимости исключается/включается, $z = 0$ или $z = \infty$.

3.8. АНАЛИЗ ЛИНЕЙНЫХ СИСТЕМ

3.8.1. Свойства Z-преобразования односторонних и двусторонних последовательностей

$x[n]$ называется *правосторонней* последовательностью, если $x[n] = 0 \quad \forall n < N$. Рассмотрим Z-преобразование такой последовательности

$$\begin{aligned} X_{>}(z) &= \sum_{n=-\infty}^{\infty} x[n]z^{-n} = \sum_{n=N}^{\infty} x[n]z^{-n} = \\ &= x[N]z^{-N} + x[N+1]z^{-(N+1)} + \dots + x[N+n]z^{-N-n}. \end{aligned}$$

Из этого выражения мы делаем вывод, что для правосторонних последовательностей

$$\lim_{z \rightarrow \infty} (z^N X_{>}(z)) = \lim_{z \rightarrow \infty} (x[N] + x[N+1]z^{-1} + \dots + x[N+n]z^{-n}) = x[N], \quad \text{и}$$

таким образом, когда $N \geq 0$, $z = \infty$ не является полюсом для $X_{>}(z)$.

Пример 1

$x[n] = a^{n+5}u[n+5]$ Для этой последовательности $x[n] = 0, \forall n < -5$ и таким образом, $N = -5$. Z-преобразование такой

последовательности $X(z) = \frac{z^5}{1-az^{-1}}, |z| > a$. Как видно из выражения

для Z-преобразования этой последовательности $z = \infty$ является полюсом, так как $X(z) = \infty, z \rightarrow \infty$. Однако,

$$\lim_{z \rightarrow \infty} (z^{-5} X(z)) = \lim_{z \rightarrow \infty} \left(\frac{1}{1-az^{-1}} \right) = 1 = x[-5]!$$

Пример 2

Известно, что Z-преобразование правосторонней последовательности $X(z) = \frac{0,5z^{-2} + 3z^{-3}}{1-0,1z^{-1}}$. Чему равно минимальное n , для которого $x[n] \neq 0$? Чему равна $x[n]$ для этого n ?

Представим $X(z)$ в форме $X(z) = z^{-2} \frac{0,5 + 3z^{-1}}{1 - 0,1z^{-1}}$.

Проанализировав это выражение, нетрудно заключить, что минимальное N , для которого $\lim(z^N X(z)) = \text{const} < \infty, \neq 0$ это $N = 2$.

Проверим это

$$\lim_{z \rightarrow \infty} \left(z^2 \left[z^{-2} \frac{0,5 + 3z^{-1}}{1 - 0,1z^{-1}} \right] \right) = \lim_{z \rightarrow \infty} \left(\frac{0,5 + 3z^{-1}}{1 - 0,1z^{-1}} \right) = 0,5 \neq \infty.$$

Соответственно, минимальное значение n для которого $x[n] \neq 0$ – это $n = 2$ и $x[2] = 0,5$.

Область сходимости Z-преобразования правосторонней последовательности простирается в бесконечность от полюса с максимальным радиусом.

Пример 3

Точка $z = \infty$ может быть включена или исключена из области сходимости, в зависимости от того, по какую сторону от $n = 0$ находится первый ненулевой элемент последовательности.

$x[n]$ называется *левосторонней* последовательностью, если $x[n] = 0 \quad \forall n > N$. Рассмотрим Z-преобразование такой последовательности

$$X_{<}(z) = \sum_{n=-\infty}^{\infty} x[n]z^{-n} = \sum_{n=-\infty}^N x[n]z^{-n} = \dots x[n]z^{-n} + x[N-1]z^{-N+1} + x[N]z^{-N}.$$

Из этого выражения мы делаем вывод, что для левосторонних последовательностей,

$\lim_{z \rightarrow 0} (z^N X_{<}(z)) = \lim_{z \rightarrow \infty} (x[n]z^{-n+N} + x[N-1]z + x[N]) = x[N]$, и таким образом, когда $N \leq 0$, $z = 0$ не является полюсом для $X_{<}(z)$.

По сравнению с аналогичными заключениями для правосторонней последовательности, в случае левосторонней последовательности $z = \infty$ заменяется на $z = 0$, что находится в полном соответствии со свойством об обращении времени

$x[-n] \xleftrightarrow{Z} X(1/z)$, ведь левосторонняя последовательность может быть получена из правосторонней путём обращения времени.

Соответственно, область сходимости Z-преобразования левосторонней последовательности простирается от нуля до полюса с минимальным радиусом. Точка $z = 0$ может быть включена или исключена из области сходимости, в зависимости от того, по какую сторону от $n = 0$ находится последний ненулевой элемент последовательности.

Пример 4

Определение области сходимости для правосторонней и левосторонней последовательностей.

Рассмотрим последовательность $x[n] = a^n u[n]$, Z-преобразование этой последовательности $X(z) = 1 + az^{-1} + a^2 z^{-2} + \dots + (az^{-1})^n$ есть сумма геометрической последовательности с показателем az^{-1} . Она сходится лишь в случае, когда $|az^{-1}| < 1 \Leftrightarrow |z| > |a|$ к выражению $X(z) = \frac{1}{1 - az^{-1}}$, т.е. область сходимости простирается в бесконечность от круга радиусом $r = |a|$.

С другой стороны, для левосторонней последовательности $x[n] = -a^n u[-n - 1]$, Z-преобразование есть следующая сумма $X(z) = -(a^{-1}z + a^{-2}z^2 + \dots + a^{-n}z^n)$, т.е. сумма геометрической последовательности с показателем $a^{-1}z$. Эта сумма сходится лишь в случае, когда $|a^{-1}z| < 1 \Leftrightarrow |z| < |a|$ к выражению $X(z) = \frac{1}{1 - az^{-1}}$, т.е. область сходимости – внутренний круг, ограниченный окружностью радиусом $r = |a|$.

3.8.2. Линейные системы, описываемые линейным разностным уравнением

Многие линейные системы цифровой обработки сигналов (ЦОС) могут быть описаны линейным разностным уравнением, связывающим выход системы $y[n]$ со входом $x[n]$ следующего вида

$$\sum_{k=0}^N a_k y[n-k] = \sum_{k=0}^M b_k x[n-k]. \quad (3.11)$$

Помимо нахождения отклика $y[n]$ для любых $x[n]$ для систем, заданных уравнением (3.11) при помощи прямых итераций, анализ свойств таких систем лучше проводить при помощи Z -преобразования. Используя свойство линейности Z -преобразования, подсчитаем Z -преобразование от правой и левой части уравнения

$$Y(z) \sum_{k=0}^N a_k z^{-k} = X(z) \sum_{k=0}^M b_k z^{-k}.$$

Мы использовали тот факт, что оператор z^{-1} , в соответствии со свойствами Z -преобразования осуществляет задержку сигнала на один отсчёт. Так как выход линейной системы связан с входом в Z -области как $Y(z) = H(z)X(z)$, то передаточная характеристика системы определяется как

$$H(z) = \frac{Y(z)}{X(z)} = \frac{\sum_{k=0}^M b_k z^{-k}}{\sum_{k=0}^N a_k z^{-k}}.$$

При анализе систем в Z -области передаточную характеристику $H(z)$ принято называть системной функцией. Соответственно, импульсная характеристика такой системы вычисляется как обратное Z -преобразование от $H(z)$, т.е. $h[n] = Z^{-1}\{H(z)\}$.

Таким образом уравнение обработки записывается как

$$Y(z) = H(z)X(z) = \frac{\sum_{k=0}^M b_k z^{-k}}{\sum_{k=0}^N a_k z^{-k}} X(z).$$

В зависимости от порядка расположения множителей $X(z)$, $\sum_{k=0}^M b_k z^{-k}$ и $\sum_{k=0}^N a_k z^{-k}$ возможны несколько различных реализаций линейных систем. Здесь мы приведём каноническую форму (рис. 3.13), использующую минимальное количество элементов задержки равное $\max(M, N)$. Последовательность множителей в уравнении обработки приведена в следующем выражении (3.12).

$$Y(z) = \left(\frac{1}{\sum_{k=0}^N a_k z^{-k}} X(z) \right) \sum_{k=0}^M b_k z^{-k}. \quad (3.12)$$

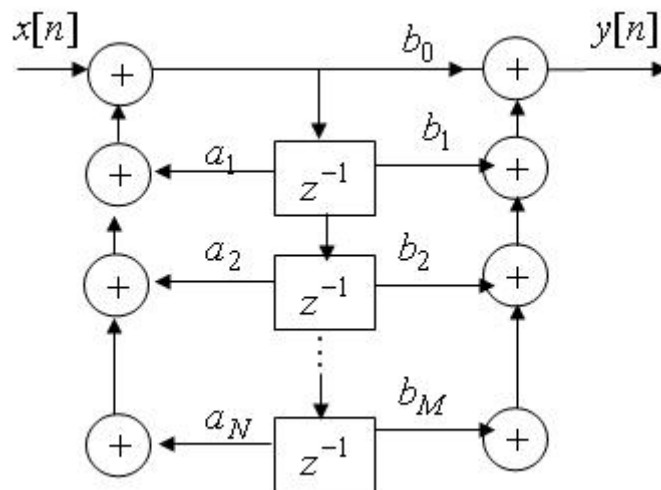


Рис. 3.13. Каноническая форма реализации линейной системы

На рис. 3.13, который соответствует канонической форме реализации линейных систем, мы предположили что $M = N$. Если это не так, то просто некоторые из коэффициентов принимаются равными нулю. Также, без потери общности мы предположили, что $a_0 = 1$. В

противном случае, скорректированные коэффициенты могут быть получены делением всех коэффициентов на a_0 .

При анализе линейных систем, системная функция очень часто представляется в виде

$$H(z) = \frac{\sum_{k=0}^M b_k z^{-k}}{\sum_{k=0}^N a_k z^{-k}} = \frac{b_0 \prod_{k=1}^M (1 - c_k z^{-1})}{a_0 \prod_{k=1}^N (1 - d_k z^{-1})}.$$

Каждый множитель $1 - c_k z^{-1}$ в числителе привносит один "ноль" в точке $z = c_k$ и один «полюс» в точке $z = 0$. Соответственно, каждый множитель $1 - d_k z^{-1}$ в знаменателе привносит один "полюс" в точке $z = d_k$ и один "ноль" в точке $z = 0$.

3.8.3. Устойчивость и физическая реализуемость

Переход от разностного уравнения к уравнению обработки был осуществлён, опираясь на свойство линейности и инвариантности во времени. Свойства устойчивости и физической реализуемости не были использованы при этом переходе. Также, из разностного уравнения мы не можем получить область сходимости. Очевидно, что выбранная область сходимости будет влиять на форму соответствующей импульсной характеристики $h[n]$ системы и, как следствие, на устойчивость и физическую реализуемость систем.

Для того, чтобы система была физически реализуемой, её импульсная характеристика должна быть правосторонней последовательностью. Соответственно, область сходимости должна лежать по правую сторону от полюса с самым большим радиусом. Для устойчивости, необходимо, чтобы область сходимости включала область, внутреннюю по отношению к единичной окружности.

3.8.4. Простейшее звено с одним "нулём"

Передаточная функция системы $H(z) = 1 - az^{-1} = \frac{z - a}{z}$ —

$$H(z)|_{z=e^{j\omega}} = \frac{e^{j\omega} - re^{j\theta}}{e^{j\omega}} = \frac{\mathbf{v}_1 - \mathbf{v}_2}{\mathbf{v}_1}$$

получена путём замены $a = re^{j\theta}$ для

подчёркивания факта, что a — комплексное и может быть изображено в виде вектора. Рассмотрим диаграмму нулей и полюсов этого простейшего звена (рис. 3.14)

$$\arg(1/1 - az^{-1}) = -\arg(1 - re^{j\theta}e^{-j\omega}) = -a \tan\left(\frac{r \sin(\omega - \theta)}{1 - r \cos(\omega - \theta)}\right)$$

$$\text{АЧХ} - \text{модуль передаточной функции } |H(e^{j\omega})| = \frac{|\mathbf{v}_1 - \mathbf{v}_2|}{|\mathbf{v}_1|} = |\mathbf{v}_3|.$$

$$\text{ФЧХ, соответственно } \text{Arg}(H(e^{j\omega})) = \text{Arg}(\mathbf{v}_3) - \text{Arg}(\mathbf{v}_1).$$

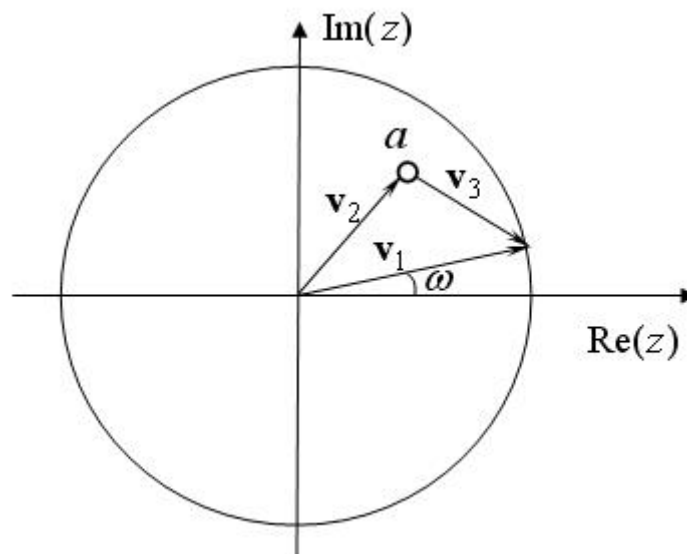


Рис. 3.14. Диаграмма нулей и полюсов

Вектор $\mathbf{v}_1$ — полюс-вектор, т.е. вектор, направленный из полюса в точку на единичной окружности, где мы хотим вычислить значение АЧХ и ФЧХ. Вектор $\mathbf{v}_3$ — нуль-вектор, т.е. вектор, направленный из

полюса в точку на единичной окружности, где мы хотим вычислить значение АЧХ и ФЧХ.

В случае элементарного звена, имеющего полюс в точке $z = a$, всё остаётся точно так же, только числитель и знаменатель меняются местами.

Таким образом, АЧХ равна отношению произведения длин нуль-векторов к произведению длин полюсов-векторов. ФЧХ равна сумме углов нуль-векторов за вычетом углов полюс-векторов.

3.8.5. Все пропускающие системы

Системы с АЧХ, равной единице на всем частотном интервале, называются все пропускающими системами (ВПС). Они формируют класс систем, имеющий большую теоретическую и практическую важность. Простейшая ВПС имеет следующую системную функцию

$$H_{vp}(z) = \frac{z^{-1} - a^*}{1 - az^{-1}}. \quad \text{Соответственно, передаточная характеристика}$$

может быть записана в виде $H_{vp}(e^{j\omega}) = e^{-j\omega} \frac{1 - a^* e^{j\omega}}{1 - a e^{-j\omega}}$ и тогда

$|H_{vp}(e^{j\omega})| = 1$, так как эта дробь есть отношение двух комплексно-сопряжённых чисел.

Ноль в $z = \frac{1}{a^*}$, полюс в $z = a$. Если записать $a = re^{j\theta}$, то легко видеть, что полюс и ноль находятся по разные стороны от единичной окружности на одном радиус-векторе $e^{j\theta}$ (рис. 3.15).

В результате записать следующую формулу системной функции ВПС с действительной импульсной характеристикой

$$H_{vp}(z) = \prod_{k=1}^{M_r} \frac{z^{-1} - d_k^*}{1 - d_k z^{-1}} \prod_{k=1}^{M_c} \frac{(z^{-1} - e_k)(z^{-1} - e_k^*)}{(1 - e_k z^{-1})(1 - e_k^* z^{-1})}.$$

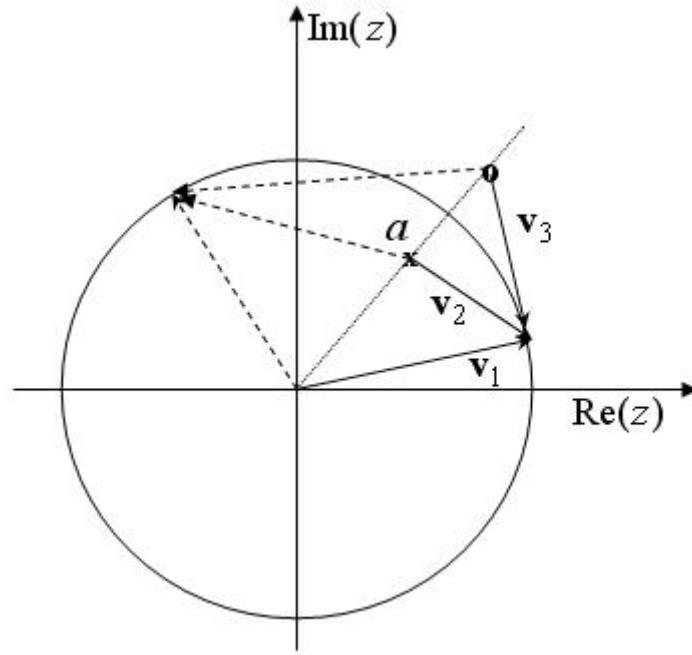


Рис. 3.15. Диаграмма нулей и полюсов

$H_{vp}(z)$ состоит из звеньев первого и второго порядков. Звенья первого порядка соответствуют действительным нулям и полюсам и идентичны звену, рассмотренному в начале параграфа. Звенья второго порядка, $\frac{(z^{-1} - e_k)(z^{-1} - e_k^*)}{(1 - e_k z^{-1})(1 - e_k^* z^{-1})}$, соответствуют комплексным нулям и полюсам и имеют АЧХ, равную единице на всём частотном диапазоне, ФЧХ может быть выражена как

$$\begin{aligned} \text{Arg} \left(\frac{(e^{-j\omega} - re^{-j\theta})(e^{-j\omega} - re^{j\theta})}{(1 - re^{-j\theta}e^{-j\omega})(1 - re^{j\theta}e^{-j\omega})} \right) = \\ = -2\omega - 2 \arctan \left(\frac{r \sin(\omega - \theta)}{1 - r \cos(\omega - \theta)} \right) - 2 \arctan \left(\frac{r \sin(\omega + \theta)}{1 - r \cos(\omega + \theta)} \right). \end{aligned}$$

Для получения этого выражения мы представили $e_k = re^{j\theta}$.

Можно показать, что групповая задержка ВПС, вычисляемая как производная по частоте от ФЧХ, всегда положительна. Из этого следует, что ВПС, каскадно соединённая с любой другой системой, увеличит общую групповую задержку.

3.8.6. Связь амплитудно-частотной и фазо-частотной характеристик

Существует класс линейных систем, для которых знание АЧХ полностью определяет ФЧХ, а знание ФЧХ определяет АЧХ с точностью до коэффициента. Другими словами, ФЧХ определяет форму АЧХ. Для того чтобы понять, как такое может быть, рассмотрим выражение для рациональной дроби

$$C(z) = H(z)H^*(1/z^*) = \left(\frac{b_0}{a_0}\right)^2 \frac{\prod_{k=1}^M (1 - c_k z^{-1})(1 - c_k^* z)}{\prod_{k=1}^N (1 - d_k z^{-1})(1 - d_k^* z)}. \quad (3.13)$$

Это в некотором смысле искусственный приём, но обратите внимание на тот факт, что $C(z)|_{z=e^{j\omega}} = |H(e^{j\omega})|^2$. Рациональная дробь $C(z)$ так же, как и дробь $H(z)$, характеризуется нулями и полюсами.

Если нам дано выражение для АЧХ, то можно считать, что нам дана дробь $C(z)$, которую мы можем разложить на множители и представить в виде (3.13). Половина полюсов и нулей (рис. 3.16) этой дроби соответствуют системной функции $H(z)$, вторая половина появляется как результат вычисления модуля, т.е. умножения на $H^*(1/z^*)$.

Обратите внимание, что $H^*(1/z^*)$ имеет нули полюса, получающиеся отражением через единичную окружность соответствующих нулей и полюсов $H(z)$. Результатом этого процесса является диаграмма нулей (рис. 3.16) с парами $re^{j\theta}$ и $\frac{1}{r}e^{j\theta}$, пример которой изображён на рис. 3.15. Каждая такая пара обозначена одной цифрой.

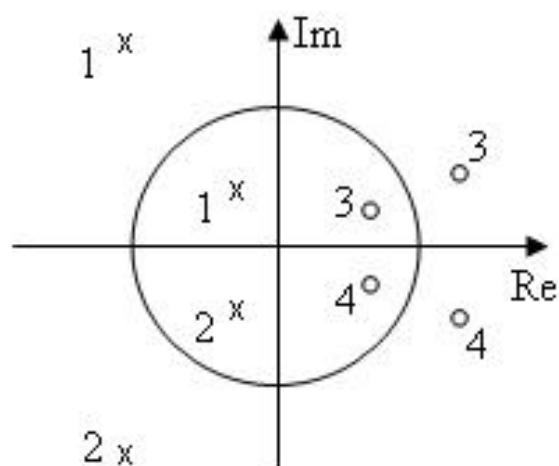
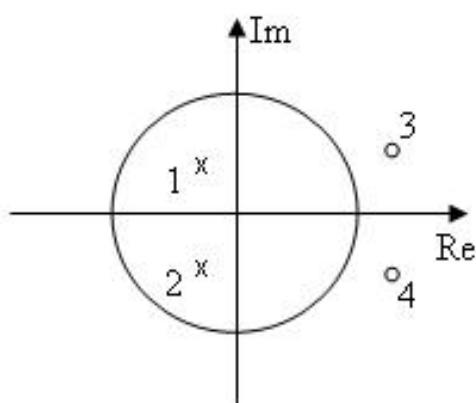


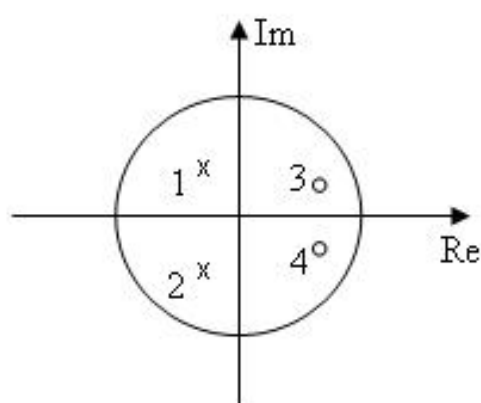
Рис. 3.16. Диаграмма нулей и полюсов

Если, к примеру, мы знаем, что система $H(z)$, соответствующая данному нам выражению $C(z)$, является устойчивой системой, то для $H(z)$ мы можем однозначно выбрать полюса, находящиеся внутри единичной окружности!

В соответствии с примером на рис. 3.16 у нас остается два варианта в выборе нулей из пар 3 и 4. Мы можем выбрать нули, находящиеся за единичной окружностью, как это показано на рис. 3.17, а.



а



б

Рис. 3.17. Возможные конфигурации нулей и полюсов $H(z)$ для устойчивой системы с действительными коэффициентами

Мы также можем выбрать пару комплексно-сопряжённых нулей внутри единичной окружности, в соответствии с 3.19, б. Обратите

внимание, что если бы у нас не было ограничения на действительность импульсной характеристики, то появились бы две дополнительные возможности, не ограниченные требованием комплексно-сопряжённости нулей.

Таким образом, у нас есть лишь две альтернативы выбора нулей при условии, что $H(z)$ – устойчивая система с действительными коэффициентами. Эти две возможности изображены на рисунке 3.17, а, б. Как следствие, две альтернативы выбора нулей определяют две различные системные функции, имеющие идентичные АЧХ. Этим двум системным функциям соответствуют две различных ФЧХ. Таким образом, в данном примере требование к устойчивости системы и действительности импульсной характеристики позволяют лишь две возможных кривых ФЧХ для систем с заданной АЧХ.

У системы, соответствующей рисунку 3.17, б, существует устойчивая обратная система, так как все нули находятся внутри единичной окружности. И значит, если мы потребуем, чтобы для системы с заданной АЧХ существовала устойчивая физически реализуемая обратная система, мы можем иметь лишь единственную ФЧХ, соответствующую выбору нулей и полюсов на рисунке 3.17, б.

3.8.7. Системы с минимальной фазовой задержкой

Система, у которой все нули и полюса находятся внутри единичной окружности, называется системой с минимальной фазовой задержкой (МФС).

Для системы с минимальной фазой:

1. Существует устойчивая и физически реализуемая обратная система.
2. Групповая задержка минимальна из всех ЛС с идентичной АЧХ.
3. АЧХ минимально фазовой системы однозначно определяет её ФЧХ.
4. ФЧХ минимально фазовой системы определяет АЧХ за исключением постоянного множителя.

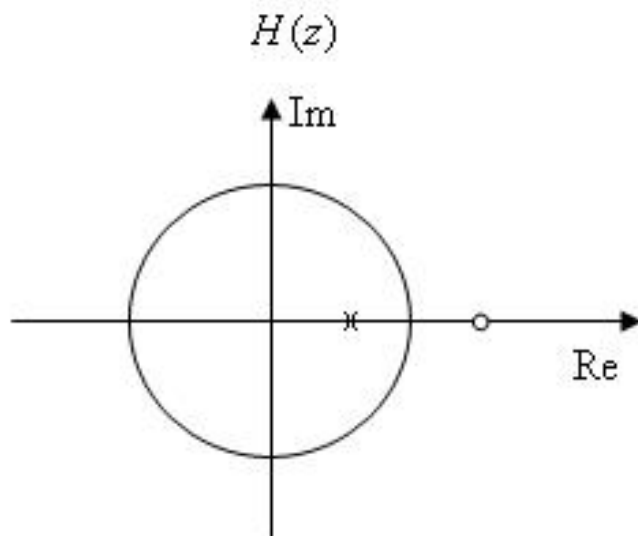
5. Сумма квадратов импульсной характеристики растёт быстрее, чем эта же величина, соответствующая любой другой системе с идентичной АЧХ. Иными словами, энергия импульсной характеристики МФС сосредоточена в наименьшем числе первых отсчетов.

Минимально-фазовая система может быть получена из любой устойчивой системы путём "отражения" нулей, находящихся за пределами единичной окружности внутрь. Т.е., если у некоторой системы $H(z)$ есть ноль $c_k = r_k e^{j\theta_k}$ с $r_k > 1$, т.е. за пределами единичной окружности, то возможно, сохранив АЧХ, перейти к минимально фазовой системе путём замены указанного нуля на

$$\tilde{c}_k = \frac{1}{r_k} e^{j\theta_k}.$$

Пример

$$H(z) = \frac{1 - 1,5z^{-1}}{1 - 0,6z^{-1}}$$



Система не является минимально-фазовой, так как есть ноль за пределами единичной окружности. Графики АЧХ и ФЧХ (не главное значение) приведены на рис. 3.18.

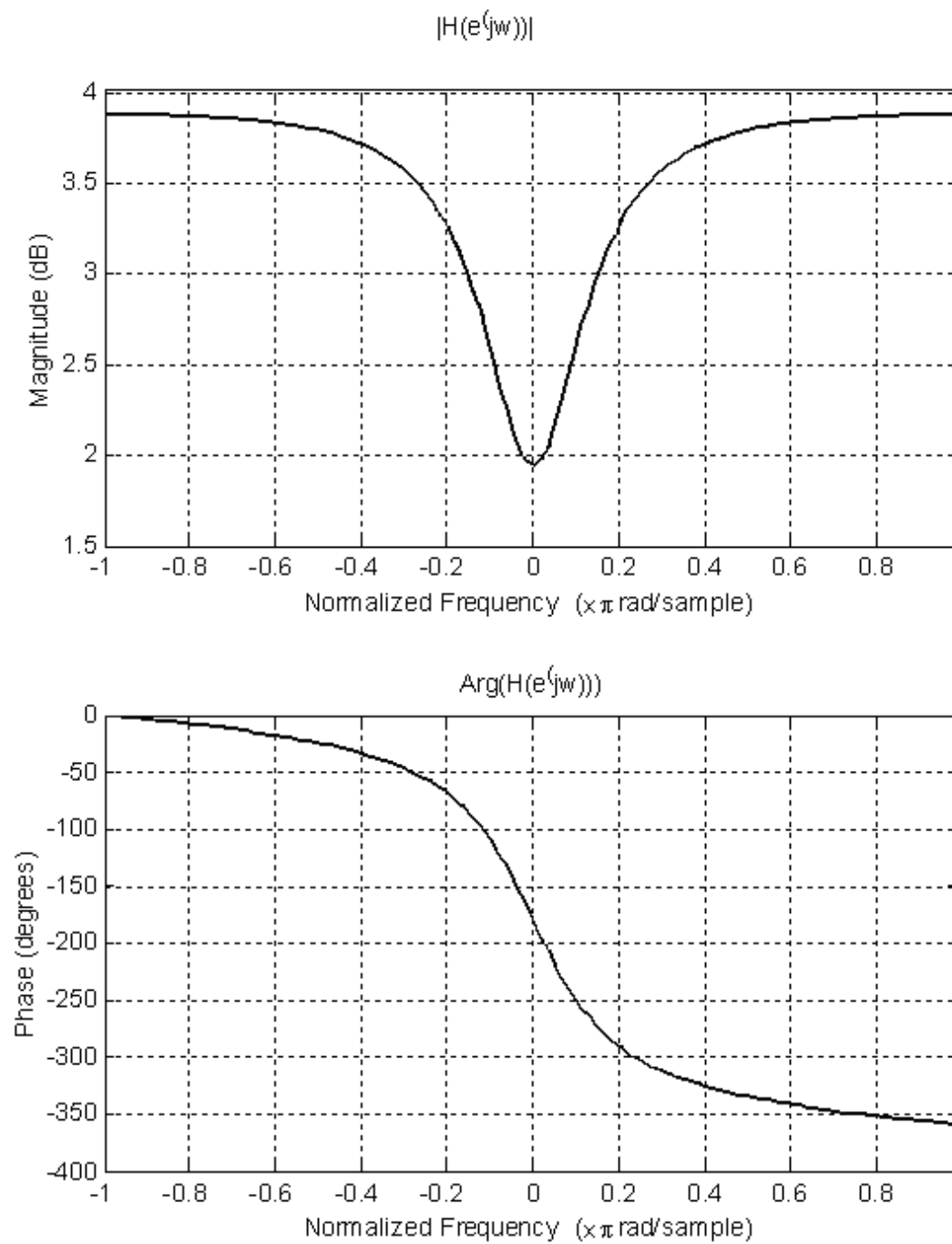


Рис. 3.18. АЧХ и ФЧХ для не минимально фазовой системы

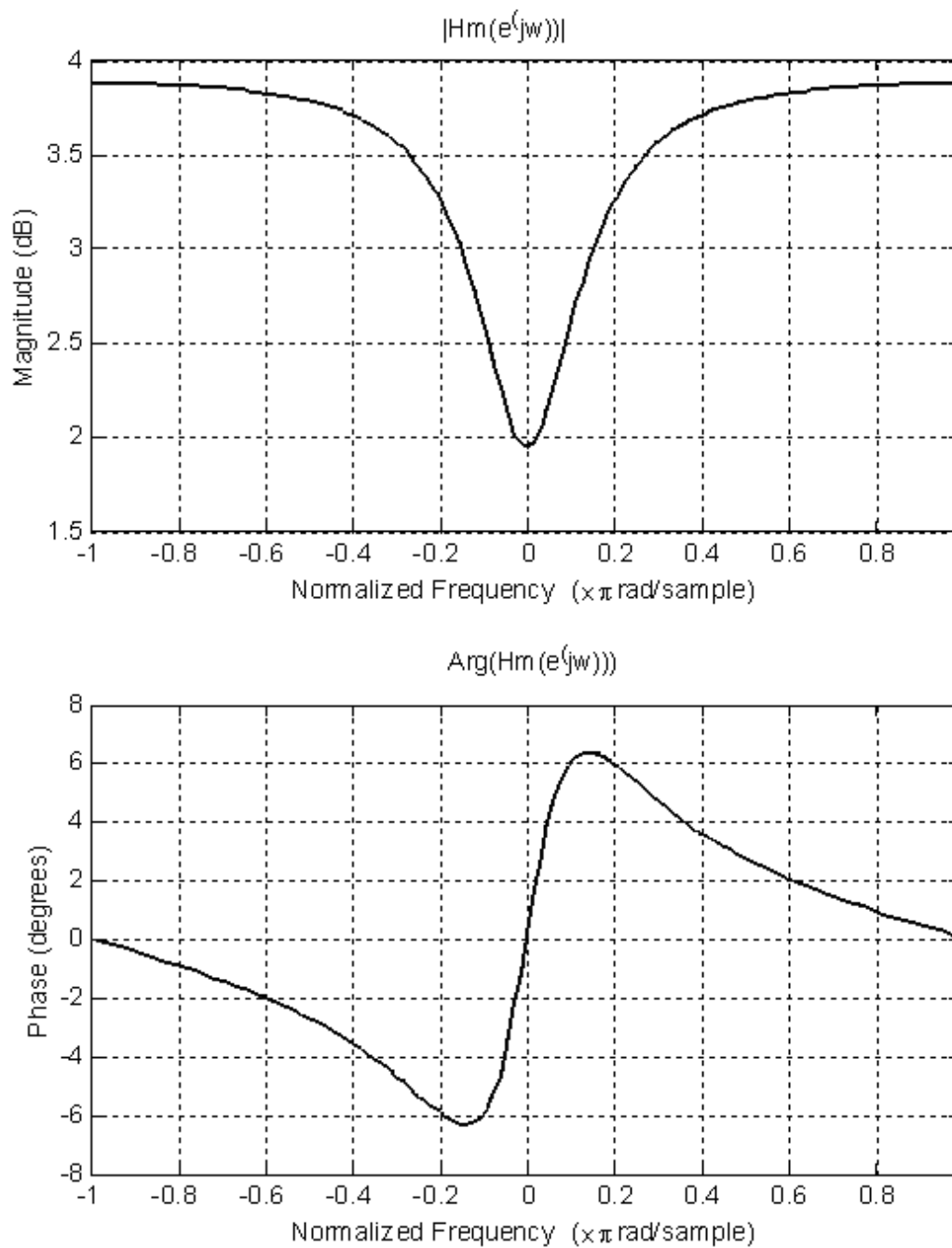
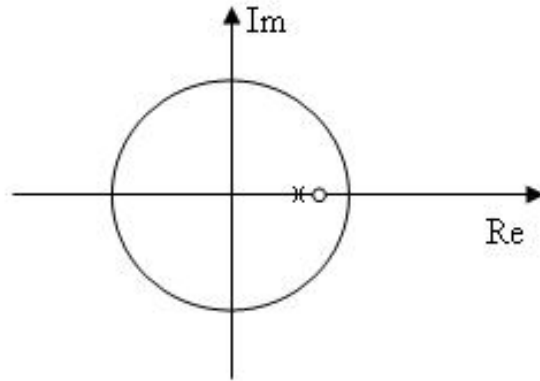


Рис. 3.19. АЧХ и ФЧХ для минимально фазовой системы

Минимально фазовая система получается из $H(z)$ путём отражения $c_1 = 1,5$ в $\tilde{c}_1 = \frac{1}{1,5}$ и умножением на 1,5 для сохранения коэффициента усиления, т.е.

$$H_m(z) = \frac{1 - 1,5z^{-1}}{1 - 0,6z^{-1}}$$



АЧХ и ФЧХ этой МФС изображены на рис. 3.19. Обратите внимание на идентичность АЧХ и разность ФЧХ. Обратите внимание на маленькие значения фазы для МФС. Мы видим, что практически вся энергия ИХ МФС сосредоточена в первом отсчёте.

Графики импульсных характеристик двух систем приведены на рис. 3.20 сверху. Нижняя часть демонстрирует факт быстреего "выброса энергии" МФС посредством подсчёта суммы квадратов отсчётов ИХ до текущего значения абсциссы на графике.

Любая система, заданная при помощи $H(z)$ в виде рациональной дроби, может быть представлена как последовательное соединение всепропускающей и минимальнофазовой систем, т.е. $H(z) = H_{\min}(z)H_{vp}(z)$. В случае нашего примера

$$H(z) = \frac{1 - 1,5z^{-1}}{1 - 0,6z^{-1}} = \underbrace{\frac{1 - \frac{1}{1,5}z^{-1}}{1 - 0,6z^{-1}}}_{H_{\min}(z)} \underbrace{\frac{1 - 1,5z^{-1}}{1 - \frac{1}{1,5}z^{-1}}}_{H_{vp}(z)} = H_{\min}(z)H_{vp}(z).$$

3.9. СОВРЕМЕННЫЕ МЕТОДЫ ЧАСТОТНО-ВРЕМЕННОГО АНАЛИЗА

Задача улучшения частотно-временного разрешения, т.е. локализации базиса в частотно-временной области является крайне нетривиальной задачей. Решение проблемы было предложено в работах одной из основоположников теории вейвлет-преобразования И. Добеши [6]. Появление техники вейвлет-анализа произошло

благодаря развитию средств частотно-временного анализа, в основе которых лежит преобразование Фурье и положения теории информации и кодирования. Совершенствование классического метода разложения сигнала в ряд Котельникова (вейвлет-разложение Шеннона-Котельникова) путем создания локализованного в частотно-временной области базиса предоставило мощное средство исследования природы сигнала [6].

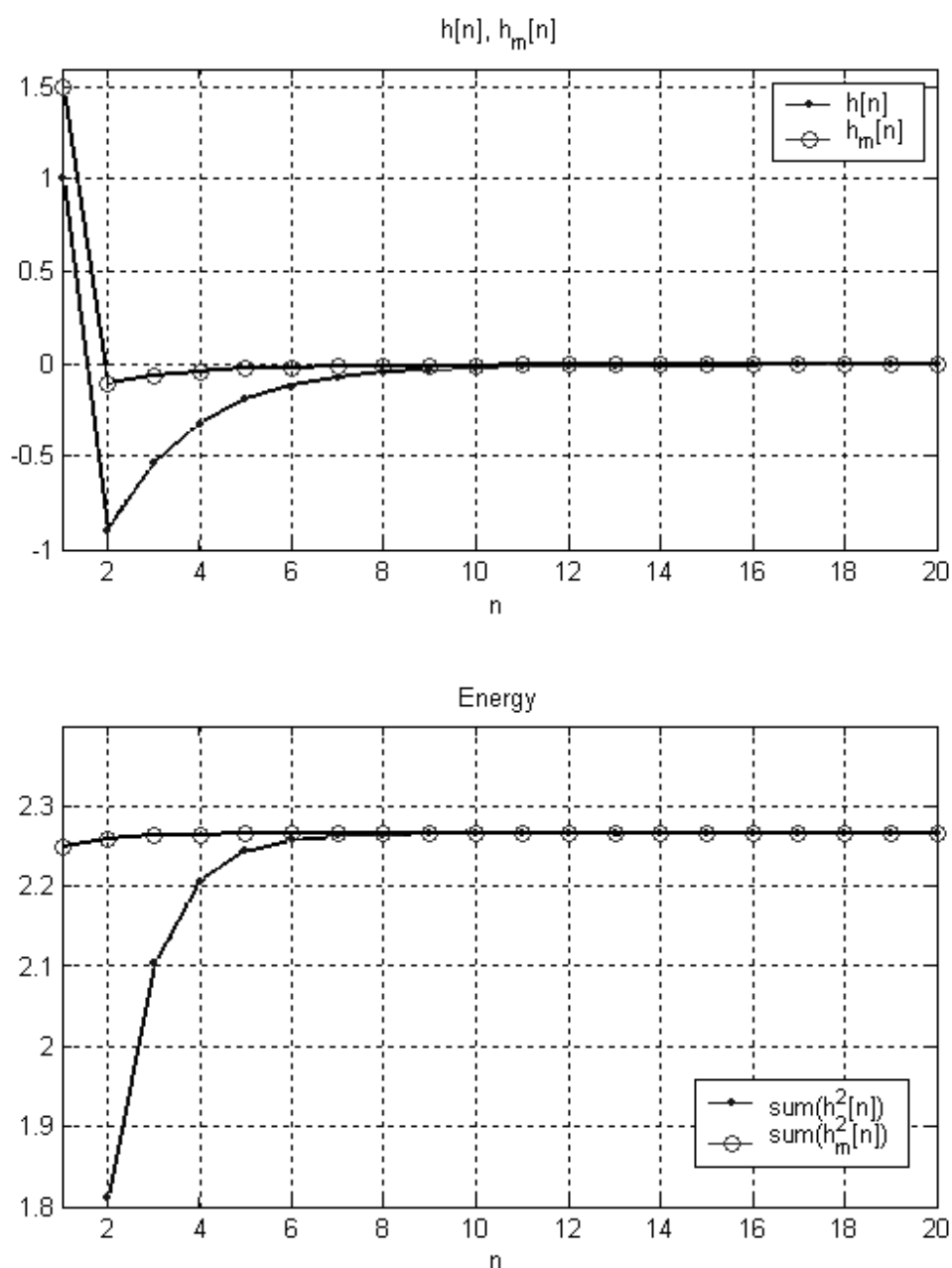


Рис. 3.20. Графики импульсных характеристик

Построение семейств вейвлет-базисов, имеющих высокую локализацию в частотно-временной области, привело к появлению нового раздела частотно-временного анализа. Наибольший интерес для инженерных приложений имеет создание дискретных базисов и быстрого алгоритма дискретного вейвлет-преобразования [6] (см. раздел 7).

3.9.1. Разложение в вейвлет-базисе

Название метода происходит от слова "*wavelet*" – небольшая волна, всплеск, и связано с видом базисных функций во временной области. Основным достоинством вейвлет-преобразования является одновременное представление сигнала во временной и частотной области, т.е. возможность выполнения локализованного спектрального анализа. Основопологающим свойством базисных функций вейвлет-анализа служит краткая протяженность по времени и частоте. Данное свойство представляет собой феномен с позиции классической теории сигналов, однако, с физической точки зрения это вполне объяснимо. В основе квантовой механики лежит понятие волновой функции или волнового пакета $\psi(\omega, t, x)$ – суперпозиции ряда монохроматических волн (рис. 2.21). И можно подобрать такие значения частот волн в составе пакета, что их суперпозиция будет иметь вид, представленный на рис. 3.22. Волновой пакет характеризуется вариацией частот монохроматических волн в его составе $\Delta\omega$ и имеет интервал локализации Δx .

Это соотношение носит название условия нормировки.

Аналогично введем базисную функцию $\psi(t)$, общий вид которой показан на рис. 3.24, и построим семейство базисных функций:

$$\frac{1}{\sqrt{a}} \cdot \frac{t-b}{\psi(a)} \quad a > 1, \quad b > 0, \quad (3.15)$$

где множитель $\frac{1}{\sqrt{a}}$ обеспечивает выполнение условия нормировки (3.14), b в аргументе ψ -функции выполняет сдвиг вдоль оси t ,

масштабный множитель a в знаменателе выполняет растяжение/сжатие базисной функции.

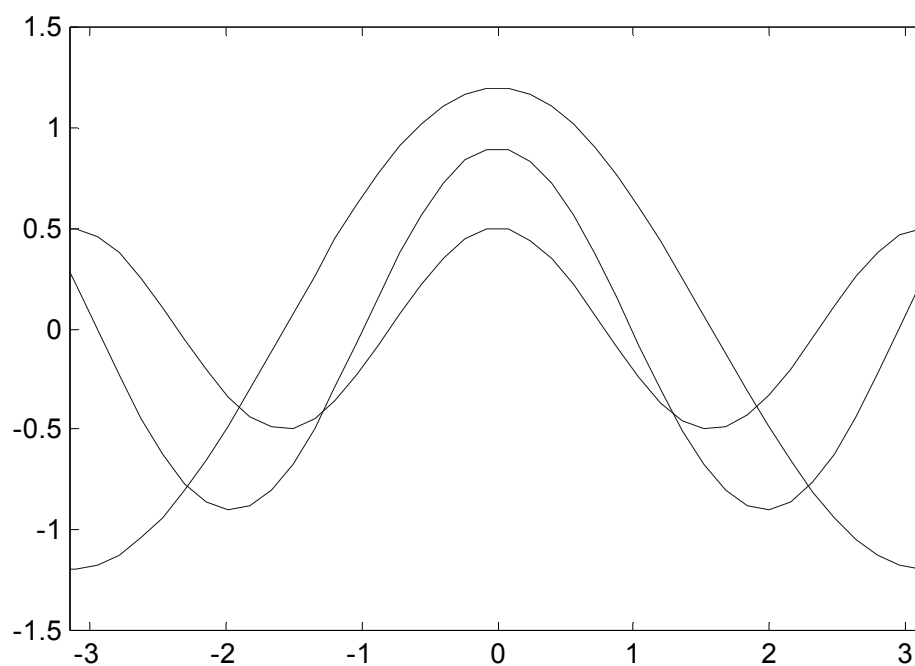


Рис. 3.21. Монохроматические волны в составе волнового пакета

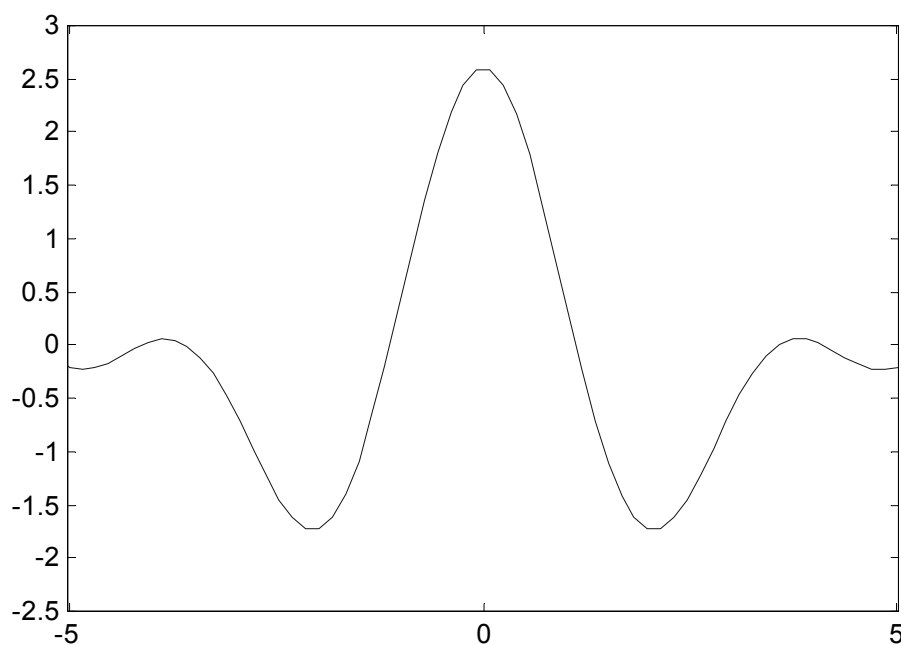


Рис. 3.22. Волновая ψ -функция

Локальность базисной функции в частотно-временной области запишем в виде [6]:

$$|\psi(t)| \leq \frac{c_1}{(1+|t|)^{1+\varepsilon}} \quad |\Psi(\omega)| \leq \frac{c_2}{(1+|\omega|)^{1+\varepsilon}}, \quad \forall \varepsilon \geq 0,$$

где:

$$\Psi(\omega) = \int_{-\infty}^{\infty} \psi(t) e^{-j\omega t} dt. \quad (3.16)$$

Тогда для любой функции $f(t)$ из Гильбертова пространства $L_2(\mathbf{R})$, т.е. бесконечномерного пространства функций с конечной энергией:

$$L_2(R) = \left\{ \|f\|^2 = \int_{-\infty}^{\infty} |f(t)|^2 dt < \infty \right\}, \quad (3.17)$$

на котором определено скалярное произведение:

$$\langle f, g \rangle = \int f(t) g^*(t) dt \quad (3.18)$$

существует прямое вейвлет-преобразование [1]:

$$\begin{aligned} W(a, b) &= \left\langle f(t), \psi\left(\frac{t-b}{a}\right) \right\rangle = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \psi^*\left(\frac{t-b}{a}\right) dt = \\ &= \sqrt{a} \int_{-\infty}^{\infty} F(\omega) \Psi^*(a\omega) e^{jb\omega} d\omega. \end{aligned} \quad (3.19)$$

Обратное вейвлет-преобразование имеет вид:

$$f(t) = W^{-1}(a, b) = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_0^{\infty} W(a, b) \psi\left(\frac{t-b}{a}\right) \frac{da db}{a^2}, \quad (3.20)$$

где:

$$C_\psi = \int_0^{\infty} \frac{|\Psi(\omega)|^2}{\omega} d\omega < \infty. \quad (3.21)$$

Соотношение (3.21) является условием существования обратного преобразования. Запишем:

$$\psi\left(\frac{t-b}{a}\right) = \sqrt{a} \int_{-\infty}^{\infty} \Psi(a\tau) e^{j\tau(t-b)} d\tau. \quad (3.22)$$

Подставим (3.19) и (3.22) в (3.20):

$$\begin{aligned} f(t) &= W^{-1}(a, b) = \\ &= \int_{-\infty}^{\infty} F(\tau) \left\{ \frac{a}{C_{\psi}} \int_0^{\infty} \Psi^*(a\omega) \Psi(a\tau) \int_{-\infty}^{\infty} e^{j(\omega-\tau)b} \frac{dad b}{a^2} \right\} e^{j\tau \cdot t} d\tau. \end{aligned} \quad (3.23)$$

Чтобы выполнялось (3.23) необходимо, чтобы выражение в {} равнялось единице:

$$\frac{a}{C_{\psi}} \int_0^{\infty} \Psi^*(a\omega) \Psi(a\tau) \int_{-\infty}^{\infty} e^{j(\omega-\tau)b} \frac{dad b}{a^2} = 1.$$

Учитывая, что

$$\int_{-\infty}^{\infty} e^{j(\omega-\tau)b} db = \delta(\omega - \tau), \quad (3.24)$$

необходимо, чтобы при $\omega=\tau$

$$\int_0^{\infty} \frac{|\Psi(a\omega)|^2}{a} da < \infty = C_{\psi}. \quad (3.25)$$

Выражение (3.25) очевидно тождественно (3.21) и отражает условие обратимости преобразования. На основании теоремы Парсеваля конечность энергии базисной функции $\psi(t)$ позволяет отнести ее также к пространству $L_2(\mathbf{R})$.

Для выполнения условия (3.21) необходимо, чтобы $\Psi(0)=0$, т.е. спектр вейвлет-функции был ограничен снизу. Кратность K нуля $\Psi(\omega)$ при $\omega=0$, т.е.

$$\left[\left(\frac{d}{d\omega} \right)^k \Psi(0) \right] = 0, \quad k = 0, K \quad (3.26)$$

носит название регулярности вейвлет-функции и связана с соответствующей характеристикой во временной области:

$$\int t^k \psi(t) dt, \quad k = 0, K, \quad (3.27)$$

называемой количеством нулевых моментов вейвлет-функции.

Регулярность и количество нулевых моментов являются основными характеристиками вейвлет-функции, определяющими

степень ее локализации в частотной области. Регулярность, обеспечивая требуемую гладкость кривой, ограничивает спектр вейвлет-функции со стороны высоких частот. Количество нулевых моментов связано с пересечением кривой оси t (число переходов через ноль), и, следовательно, ограничивает спектр со стороны низких частот. Учитывая это, спектр произвольной вейвлет-функции, отвечающей условию (3.21) и имеющей нуль K -той кратности при $\omega=0$, представляет собой полосу частот $\omega \pm \Delta\omega$, где $\omega + \Delta\omega > \omega - \Delta\omega > 0$.

3.9.2. Вейвлет-преобразование как средство частотно-временного анализа

Непрерывное вейвлет-преобразование можно записать следующим образом:

$$d_{j,k}(\text{шкала}, \text{сдвиг}) = \int_{-\infty}^{\infty} f(t) \psi(t) \psi(a, \text{сдвиг}) dt, \quad (3.28)$$

где $\psi(t)$ – базисная вейвлет-функция; $\psi(\text{шкала}, \text{сдвиг}, t)$ – базис, образованный шкалированием и сдвигом $\psi(t)$; $d_{j,k}(\text{шкала}, \text{сдвиг})$ – набор вейвлет-коэффициентов, которые являются функциями от $\psi(\text{шкала}, \text{сдвиг}, t)$. Параметр t определяет текущее время, относительно которого производится сдвиг базисной функции.

При умножении вейвлет-коэффициентов $d_{j,k}$ на $\psi(t)$ со сдвигом b и шкальным множителем a получают составляющие исходного сигнала. Под шкалированием функции понимается ее растяжение или сжатие по оси времени t . Рассмотрим пример шкалирования гармонической функции (рис. 3.23).

Как было определено выше, вейвлет-преобразование есть результат интегрирования произведений сигнала $f(t)$ на вейвлет-функцию $\psi(\alpha, k, t)$. В процессе вычислений получается набор коэффициентов, которые являются коэффициентами вхождения $\psi(\alpha, k, t)$ в исследуемый сигнал.

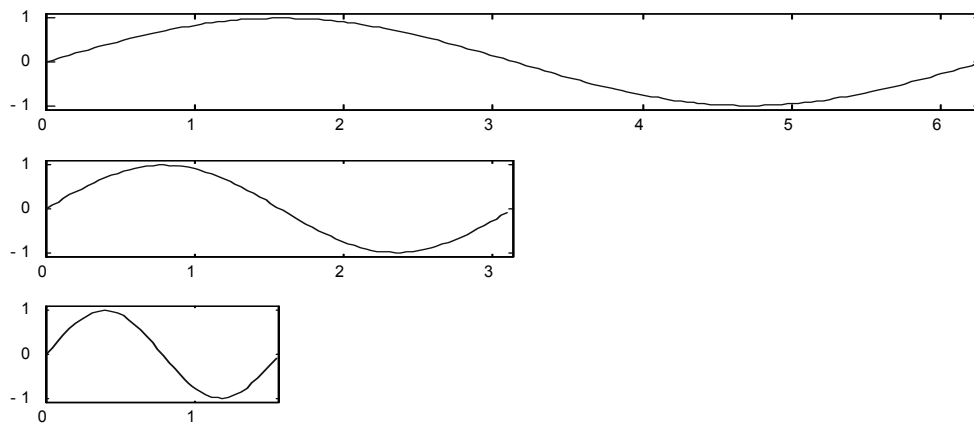


Рис. 3.23. Шкалирование гармонической функции $\sin(\alpha t)$:
соответственно, $j = 0$; $j = 1$; $j = 2$

Аналогично выполняется шкалирование вейвлет-функции $\psi(t)$ (рис. 3.24).

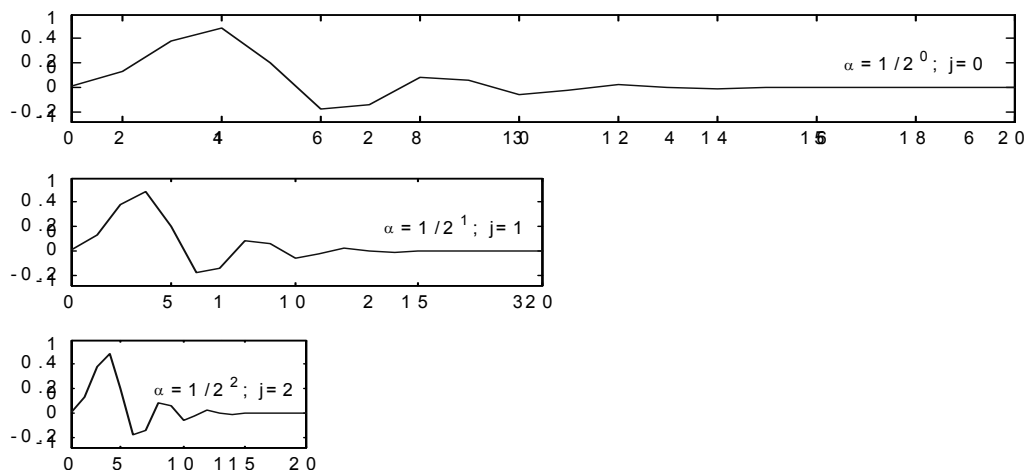


Рис. 3.24 Шкалирование вейвлет-функции $\psi(\alpha, t)$:
соответственно $j = 0$; $j = 1$; $j = 2$

Анализ составляющих сигнала $f(t)$ может быть рассмотрен иначе: как построение последовательных аппроксимаций $f(t)$ некоторой аппроксимационной функцией $\varphi(a, k, t)$. Определим на нулевом шаге аппроксимации $\varphi(0, k, t)$ как саму функцию $f(t)$. Далее растягивая аппроксимирующую функцию и вычисляя взаимную корреляцию с $f(t)$, получим более сглаженные образы функции $f(t)$ в подпространствах $\{V_j \in L_2(\mathbf{R}), j=0, J\}$, где $V_0=L_2(\mathbf{R})$ (рис. 3.25).

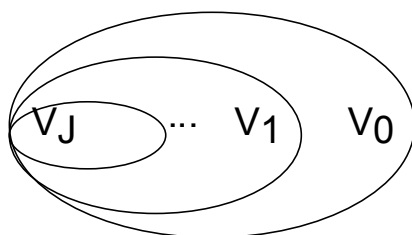


Рис. 3.25. Вложенные подпространства аппроксимации V_j

Такой подход к исследованию функций на основе серий последовательных аппроксимаций называется кратно масштабным анализом функций [6]. Очевидно, что при построении все более грубых аппроксимаций для $f(t)$ теряется часть информации о первоначальной функции, т.е. ее особенности (детали). Подпространство деталей служит ортогональным дополнением к аппроксимационному подпространству (рис. 3.26), таким образом:

$$V_j = V_{j+1} \oplus W_{j+1}. \quad (3.29)$$

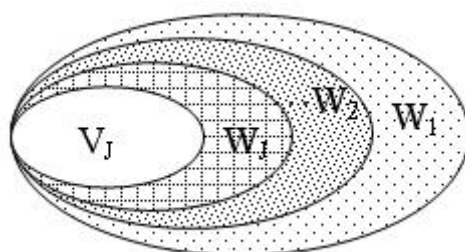


Рис. 3.26. Подпространства-дополнения W_j

Воздействие аппроксимационной функции аналогично низкочастотной фильтрации сигнала $f(t)$, тогда как детали выделяются как при полосно-пропускающей фильтрации. Так как спектр вейвлет-функции соответствует частотной характеристике некоторого полосно-пропускающего фильтра, то для кратно масштабного частотно-временного анализа сигналов необходимо поставить в соответствие вейвлет-функции $\psi(t)$ парную ей функцию с низкочастотным спектром $\varphi(t)$. Данная функция называется масштабирующей функцией. Взаимное соответствие вида масштабирующей и вейвлет-функций показано на рис. 3.27.

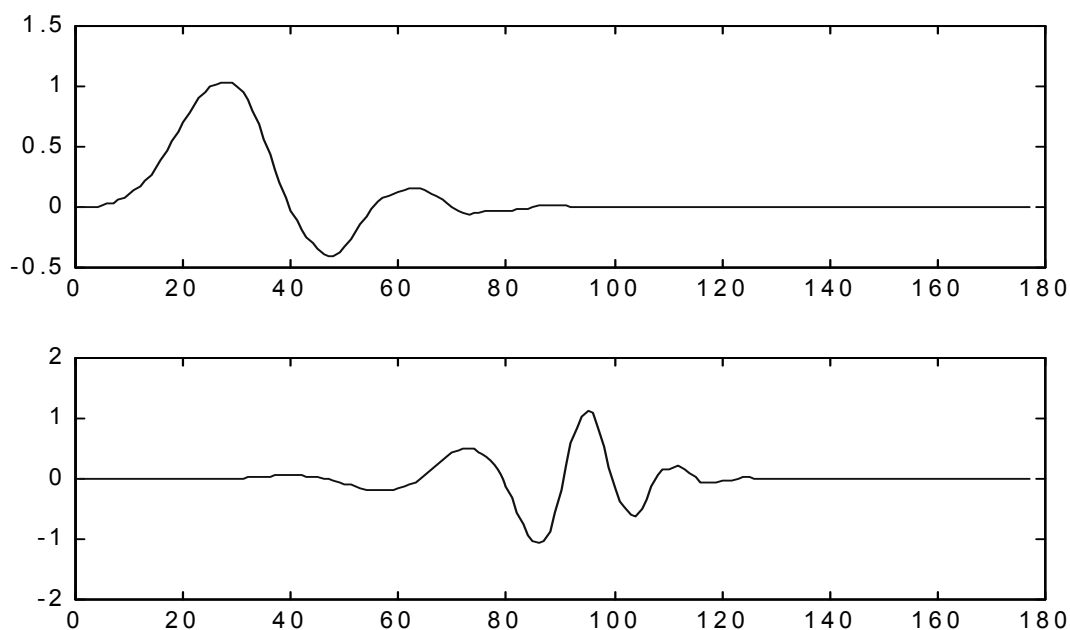


Рис. 3.27. Базисные функции вейвлет-анализа: масштабирующая функция $\varphi(t)$ (вверху), вейвлет-функция $\psi(t)$ (внизу)

3.9.3. Обобщенный алгоритм частотно-временного вейвлет-анализа

Обобщенный алгоритм частотно-временного вейвлет-анализа включает в себя следующие этапы:

1. Выбрать базис вейвлет-анализа $\{\varphi(t), \psi(t)\}$, исходя из предполагаемой природы сигнала $f(t)$ (см. табл. 3.3).
2. Вычислить коэффициенты на всем протяжении $f(t)$ с помощью сдвига $\varphi(t-k)$, $\psi(t-k)$.
3. Обнулить сдвиг и выполнить шкалирование.
4. Выполнить очередной проход сигнала с помощью сдвига и повторить для всех значений шкального множителя α .

По окончании вейвлет-преобразования (ВП) имеется набор коэффициентов $C(\alpha, k)$, каждый из которых является результатом разложения $f(t)$ по шкалам и сдвигам базисных функций. Рассмотрим подробнее соотношение между понятиями шкала и частота и покажем, каким образом в процессе вейвлет-преобразования производится частотный анализ сигнала, сопровождающийся

локализацией во времени всех частотных особенностей сигнала. С уменьшением шкального множителя α , значит, при сжатии вейвлет-функции выявляются его составляющие высокой частоты. При значениях шкального множителя $\alpha \approx 1.0$, т.е. при растянутой функции, имеющей большую протяженность, в процессе вычисления коэффициентов базисная функция сравнивается с большей областью сигнала $f(t)$, а значения коэффициентов описывают вхождение в сигнал его низкочастотных составляющих. Значения параметра k служат при этом метками во временной области.

Процесс получения вейвлет-коэффициентов исследуемого сигнала $f(t)$ и их дальнейший анализ позволяет не только исследовать реальные сигналы произвольной формы, но и выполнять более эффективно сжатие (компрессию) сигнала или изображения удаление шума (фильтрацию) [5].

Переход от время-частотного преобразования к время-шкальному обуславливается также и тем, что усреднение выбросов сигнала в процессе суммирования по разным шкалам делает вейвлет-преобразование менее чувствительным к шуму.

3.9.4. Базисные функции вейвлет-преобразования

Все базисные вейвлет функции обычно разделяют на несколько основных групп (см. табл. 3.3). Подобно любому базису разложения базис ВП оценивается по следующим критериям: ортогональность, полнота (представление любой кусочно-непрерывной дифференцируемой $f(t)$ в виде линейной комбинации базиса), возможность получения этого представления с помощью КИХ-фильтров (компактный носитель) и существование обратного преобразования (синтеза). В применении к задачам анализа сигналов интересной особенностью является возможность аппаратной реализуемости ВП в данном базисе.

Как видно из табл. 3.3 наиболее легко реализуемыми являются следующие семейства вейвлет-функций: Haar (H), Daubechies (DB),

Symlet (sym), Coiflet (coif), а также семейство биортогональных вейвлет-функций. Вид этих функций показан на рис. 3.30.

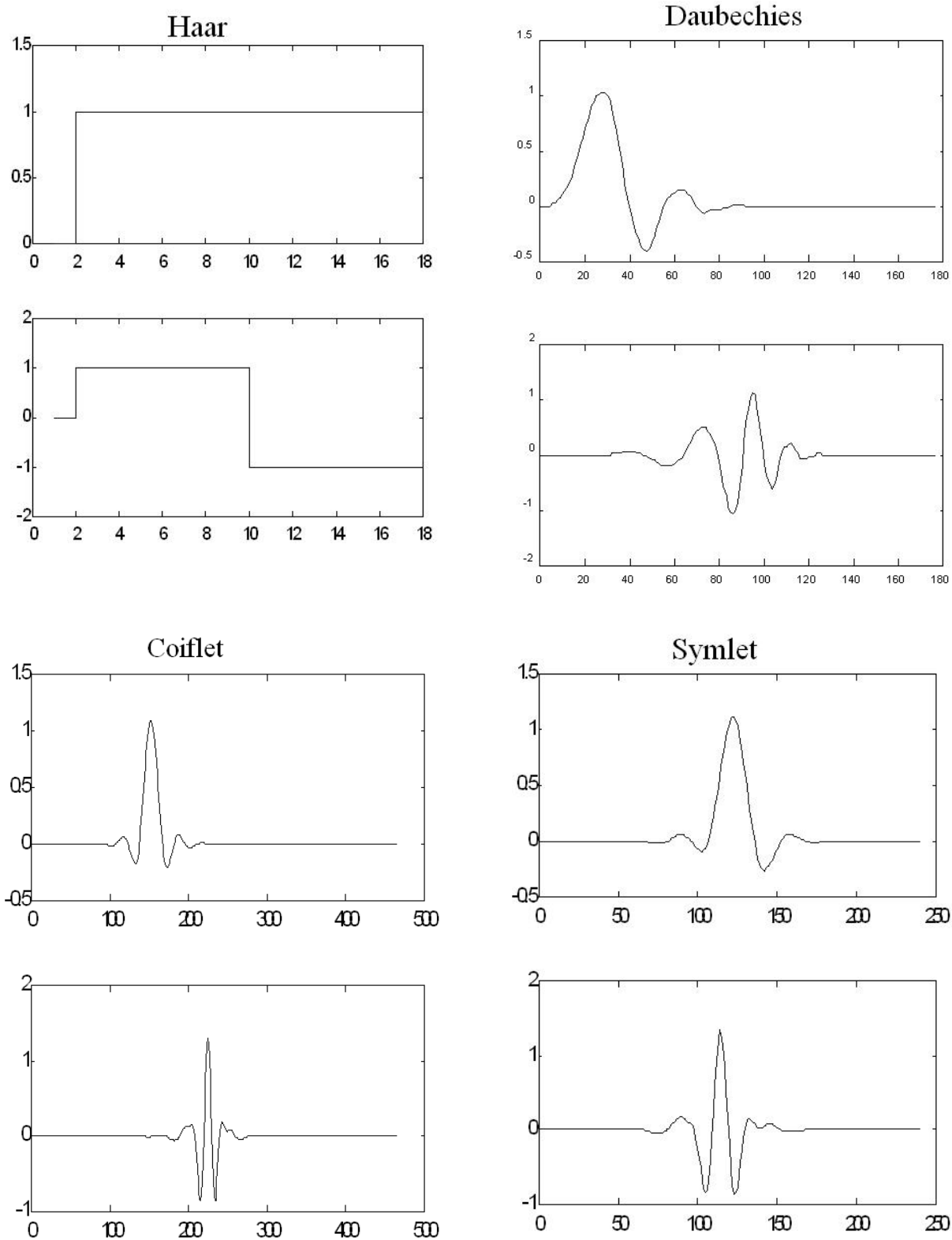


Рис. 3.30. Вид базисных функций

Т а б л и ц а 3.3

Свойства базисных функций вейвлет-анализа

Группа вейвлет - функций	Вейвлет-функции грубой аппроксимации		Бесконечно дифференцируемые	Ортогональные функции с компактным носителем.				Биортогональные
Тип вейвлет – функции	Morlet	Mexican Hat	Meyer	Haar	Daubechies	Coiflet	Symlet	B-spline
Ортогональность	—	—	+	+	+	+	+	—
Компактный носитель	—	—	—	+	+	+	+	+
Наличие обратного преобразования	—	—	+	+	+	+	+	+
Аппаратная реализуемость	—	—	реализуется с помощью БИХ-фильтров	реализуются с помощью КИХ-фильтров				
Прочие свойства	симметрия, явное математическое описание		симметрия, бесконечная дифференцируемость	симметрия, явное мат. описание				

4. ОПЕРАЦИИ ПРЕОБРАЗОВАНИЯ СИГНАЛОВ

На разных этапах процессов получения и обработки информации материальное представление сигналов в узлах информационной системы, а также их математическое описание, могут изменяться путем соответствующего преобразования сигналов.

4.1. ФИЛЬТРАЦИЯ

Фильтром называется система, пропускающая компоненты сигнала из заданного диапазона частот [4].

4.1.1. Проектирование фильтров

Различают следующие виды фильтров:

а. *Фильтры низких частот (ФНЧ)*, пропускающие низкочастотные компоненты сигнала и задерживающие высокочастотные;

б. *Фильтры высоких частот (ФВЧ)*, пропускающие высокочастотные компоненты сигнала и задерживающие низкочастотные;

в. *Фильтры средних частот (ФСЧ)* или полосно-пропускающие фильтры, пропускающие лишь компоненты сигнала из заданной полосы частот;

г. *Режекторные фильтры*, пропускающие все компоненты сигнала за исключением компонент из заданной полосы частот.

д. *Гребённые фильтры* – представляют собой набор полосно-пропускающих фильтров.

На рис. 4.1 приведены примеры АЧХ идеальных фильтров разных типов. Мы видим, что коэффициент усиления фильтра является функцией частоты. В *полосе пропускания* – коэффициент усиления $|H(j\Omega)| \gg 0$, а в *полосе задержания* $|H(j\Omega)| \ll 1$. У реальных фильтров, кроме этих полос, существуют *полосы перехода*,

расположенные между полосой пропускания и полосой задержания (рис. 4.2).

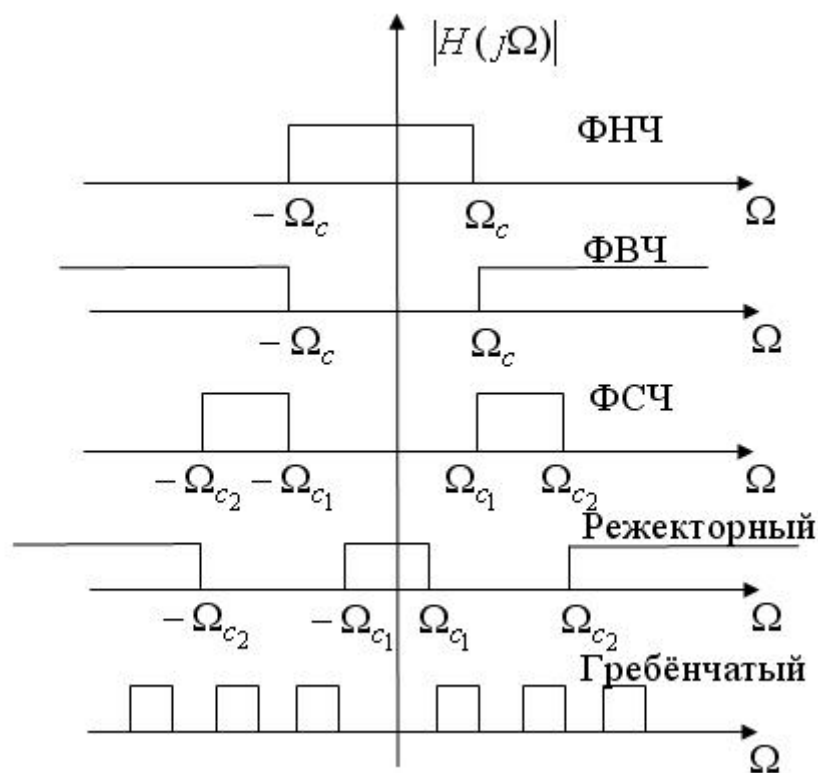


Рис. 4.1. Примеры АЧХ идеальных фильтров

Многие методы проектирования фильтров основаны на разработке ФНЧ и последующей модификации коэффициентов для получения фильтров других типов.

1. Параметры, определяемые при проектировании фильтра

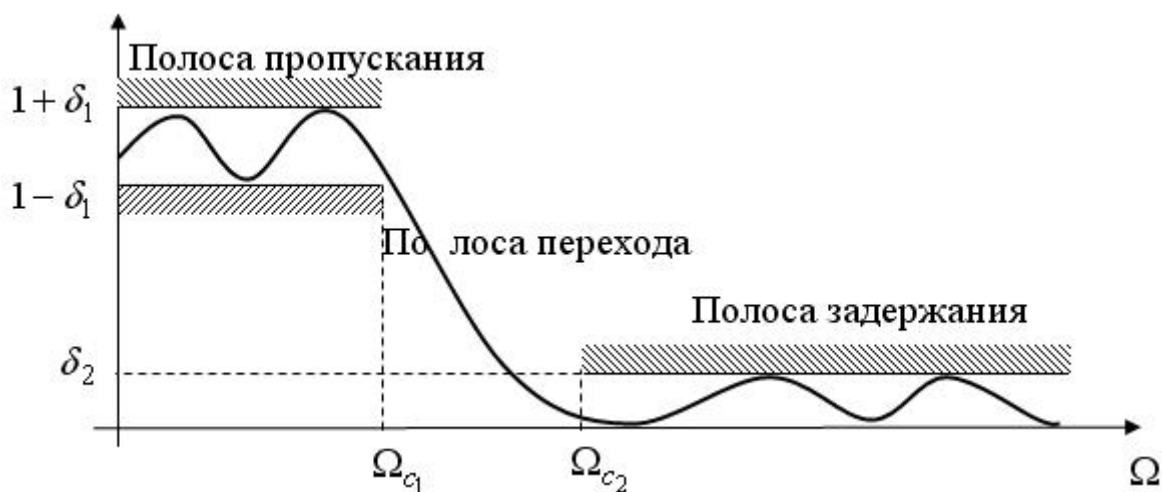


Рис. 4.2. Пример АЧХ реального фильтра нижних частот

Набор следующих параметров, характеризующих форму АЧХ фильтра низких частот, обычно берётся за основу при разработке фильтра:

Ω_{c1} — конец полосы пропускания;

Ω_{c2} — начало полосы задержания;

δ_1 — допустимые флуктуации в полосе пропускания;

δ_2 — допустимые флуктуации в полосе задержания.

Обратите внимание, что частоты заданы в области реальных, Ω , а не цифровых частот, ω . Наша задача — построить схему обработки аналогового сигнала, осуществляющую цифровую фильтрацию сигнала в соответствии с вышеприведёнными характеристиками в соответствии с общей схемой цифровой обработки (рис. 4.3).

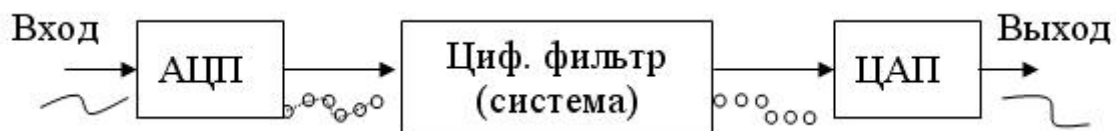


Рис. 4.3. Общая схема цифровой обработки

Для выполнения этой задачи нам необходимо перевести характеристики фильтра из области аналоговых частот в область цифровых частот. *Формально процесс проектирования цифрового фильтра начинается с задания характеристик в области цифровых частот.*

2. Перенос характеристик фильтра в область цифровых частот

Цифровая система с передаточной функцией $H(e^{j\omega})$, построенная в соответствии со схемой, приведённой на рис. 4.3, в случае, если на её вход поступает сигнал с ограниченным спектром, дискретизированный с периодом T , удовлетворяющим теореме Шеннона-Котельникова-Найквиста, ведёт себя идентично аналоговой системе с передаточной характеристикой

$$H_{eff}(j\Omega) = \begin{cases} H(e^{j\Omega T}), & |\Omega| < \pi / T, \\ 0, & |\Omega| > \pi / T. \end{cases}$$

В таком случае, используя соотношение $\omega = \Omega T$ между цифровой и аналоговой частотой, нетрудно привести следующее соответствие характеристик фильтров (табл. 4.1), заданных в области аналоговых и цифровых частот.

Т а б л и ц а 4.1

Соответствие характеристик фильтров

Параметр	Аналоговый	Цифровой
Конец полосы пропускания (ПП)	Ω_{c_1}	$\omega_{c_1} = \Omega_{c_1} T$
Начало полосы задержания (ПЗ)	Ω_{c_2}	$\omega_{c_2} = \Omega_{c_2} T$
Допустимые флуктуации в ПП	δ_1	δ_1
Допустимые флуктуации в ПЗ	δ_2	δ_2

Пример

Следующие характеристики заданы в области низких частот

$\Omega_{c_1} = 2\pi 2000$; $\Omega_{c_2} = 2\pi 3000$; $\delta_1 \leq 0,01$; $\delta_2 \leq 0,001$. Необходимо перевести эти характеристики в область цифровых частот при условии, что интервал дискретизации (величина, обратная частоте дискретизации) $T = 10^{-4}$ секунды.

$$\omega_{c_1} = \Omega_{c_1} T = 2\pi 2000 \times 10^{-4} = 0,4\pi,$$

$$\omega_{c_2} = \Omega_{c_2} T = 2\pi 3000 \times 10^{-4} = 0,6\pi.$$

Величины флуктуаций в переводе не нуждаются.

3. Принцип инвариантности импульсной характеристики

Предположим, что у нас есть аналоговая линейная система с ИХ $h_a(t)$ и мы хотим построить цифровую систему, в соответствии со схемой на рис. 4.3, осуществляющую преобразование сигнала, идентичное аналоговой системе. В цифровую область мы можем перейти лишь после выбора периода дискретизации T , что, в свою очередь, накладывает ограничение на частотный диапазон сигналов,

для которых наша цифровая система будет работать корректно. Выбрав T , продискретизируем $h_a(t)$ для получения ИХ цифровой системы, т.е. $h[n] = h_a(nT)$. В таком случае, в соответствии с теоремой о преобразовании спектра дискретизированного сигнала, передаточная функция системы с ИХ $h[n]$ есть

$$H(e^{j\omega}) = \frac{1}{T} \sum_{k=-\infty}^{\infty} H_a(j\frac{\omega}{T} - 2\pi \frac{k}{T}), \quad \text{где } H_a(j\Omega) \text{ — передаточная}$$

функция соответствующей аналоговой системы. Так как мы строим цифровую систему для обработки аналогового сигнала (рис. 4.3) и выбираем период дискретизации, позволяющий корректно обрабатывать аналоговый сигнал, избегая эффектов наложения в частотной области, то нас интересует лишь первый период передаточной функции цифровой системы, а именно

$$H_1(e^{j\omega}) = \frac{1}{T} H_a\left(j\frac{\omega}{T}\right). \quad \text{Обратите внимание, что в результате}$$

дискретизации появился дополнительный коэффициент $\frac{1}{T}$, влияющий на общий коэффициент усиления нашей цифровой системы. Соответственно, для того чтобы скомпенсировать этот эффект, при получении $h[n]$ мы должны умножить последнюю на T .

Итак, для того чтобы построить цифровую систему с ИХ $h[n]$, осуществляющую обработку сигнала, идентичную линейной системе с ИХ $h_a(t)$, нам необходимо

1. Выбрать период дискретизации, так чтобы избежать эффекта наложения спектров.

2. Принять в качестве ИХ цифровой системы $h[n] = Th_a(nT)$.

4. Использование принципа инвариантности ИХ при проектировании цифровых фильтров. Метод усечения ИХ

Для того чтобы реализовать фильтр с импульсной характеристикой $h[n]$, отсчёты которой мы получили путём

дискретизации ИХ аналогового фильтра, нам необходимо ограничить число отсчётов в ИХ цифрового фильтра.

Цифровой фильтр с усечённой ИХ будет иметь частотные характеристики, отличные от таковых исходного аналогового фильтра. Рассмотрим эффект усечения ИХ на примере идеального фильтра низких частот с частотой среза Ω_c .

ИХ аналогового идеального фильтра низких частот может быть записана как $h_a(t) = \frac{\sin(\Omega_c t)}{\pi t}$. В соответствии с принципом инвариантности, при периоде дискретизации T , ИХ идеального цифрового фильтра будет равна

$$h[n] = T \frac{\sin(\Omega_c T n)}{\pi T n} = \frac{\sin(\Omega_c T n)}{\pi n} = \frac{\sin(\omega_c n)}{\pi n}.$$

Усечение ИХ эквивалентно умножению на прямоугольное окно

$$w[n] = \begin{cases} 1, & -N \leq n \leq N \\ 0, & n < -N, n > N \end{cases}$$

$$\hat{h}[n] = h[n]w[n],$$

где $\hat{h}[n]$ – усечённая ИХ.

Соответственно, в частотной области этот процесс может быть записан как свёртка

$$\hat{H}(e^{j\omega}) = \frac{1}{2\pi} H(e^{j\omega}) W(e^{j\omega}),$$

где $W(e^{j\omega}) = \frac{\sin[\omega(2N+1)/2]}{\sin(\omega/2)}$ – Фурье преобразование $w[n]$.

Вследствие свёртки идеальная передаточная функция фильтра искажается и превращается в передаточную функцию $\hat{H}(e^{j\omega})$, обозначенную как $H_1(j\omega)$ на рис. 4.4. Искажения идеальной передаточной функции могут быть охарактеризованы при помощи ширины переходной зоны U и высоты первого "бокового лепестка" Q (рис. 4.4). В идеале мы хотим уменьшить U и Q .

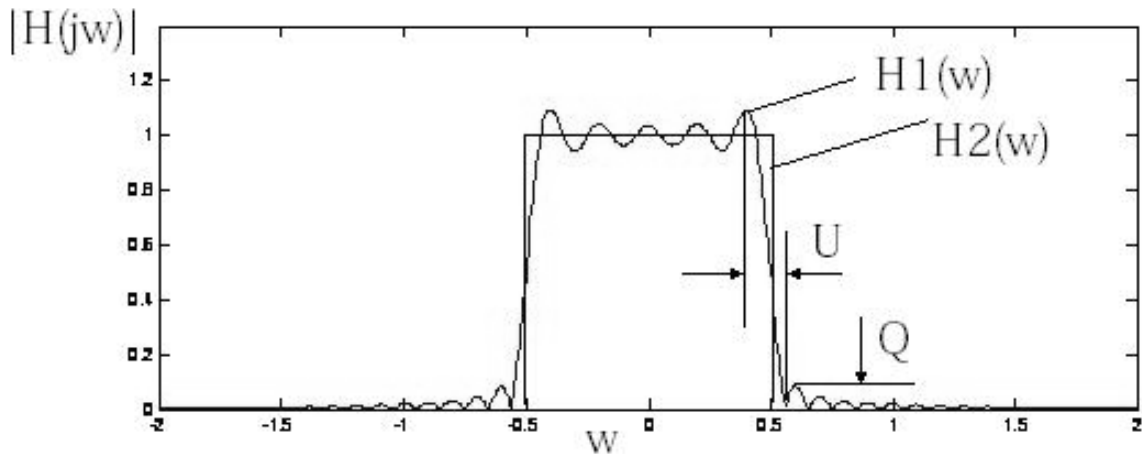


Рис. 4.4. Передаточная функция фильтра

Выбор формы окна $w[n]$, усекающего ИХ, влияет на результирующие U и Q . Помимо прямоугольного окна, существует большое количество различных окон, названных в честь предложивших их исследователей: окна Ханна, Хэмминга, Блэкмена, Бартлетта. К сожалению, окна, которое одновременно уменьшает и U , и Q , не существует. Каждое из перечисленных окон позволяет найти посредством экспериментов желаемый компромисс между шириной переходной полосы и высотой бокового лепестка. Подбор окна, результирующего в фильтр, удовлетворяющий желаемым характеристикам $(\omega_{c1}, \omega_{c2}, \delta_1, \delta_2)$, требует большого числа экспериментов, избежать которого позволяет использование окна Кайзера, для которого существуют эмпирические формулы расчёта характеристик окна, позволяющих достичь желаемых $\omega_{c1}, \omega_{c2}, \delta_1, \delta_2$.

К настоящему времени к методам аппроксимации для фильтров с линейной фазовой характеристикой следует отнести метод усечения идеальной импульсной характеристики оконной функцией (метод "окна") и методы оптимизации на основе среднеквадратического и минимаксного критерия оптимизации. Побуждающим моментом развития метода "окна" явился аномальный эффект Гиббса, сущность которого сводится к неустранимости максимальной ошибки аппроксимации с ростом порядка фильтра при критерии минимума

среднеквадратической ошибки аппроксимации идеальной кусочно-постоянной амплитудно-частотной характеристики (АЧХ). Наиболее гибким является параметрическое окно, найденное Кайзером. В то же время, при прочих равных условиях, постановка задачи на основе минимаксного критерия позволяет добиться лучших результатов. Более того, на основе минимаксного критерия оптимизации могут быть решены и задачи аппроксимации заданных АЧХ в классе КИХ-фильтров (финитных сигналов) с нелинейной и произвольной фазовой характеристикой, а также БИХ-фильтров без использования аналоговых фильтров-прототипов. Заслуживает внимания и задачи минимаксной аппроксимации в классе КИХ-фильтров для разложения сигналов, представленных в дискретной форме. Разработка цифровых фильтров на основе минимаксного критерия оптимизации изложена в пособии [3].

5. Метод Кайзера

Окно Кайзера выражается при помощи модифицированных функций Бесселя нулевого порядка, а именно

$$w[n] = \begin{cases} \frac{I_0\left[\beta\left(1 - [n/N]^2\right)^{1/2}\right]}{I_0(\beta)}, & -N \leq n \leq N, \\ 0, & n < -N, n > N. \end{cases}$$

В этом выражении $I_0()$ – модифицированная функция Бесселя нулевого порядка, β и N – параметры окна, вычисляемые по следующим эмпирическим формулам:

$$\begin{aligned} \Delta\omega &= \omega_{c_2} - \omega_{c_1}, \quad \delta = \min(\delta_1, \delta_2), \quad A = -20 \log_{10} \delta, \\ \beta &= \begin{cases} 0,1102(A - 8,7) & A > 50, \\ 0,5842(A - 21)^{0.4} + 0,07886(A - 21), & 21 \leq A \leq 50, \\ 0,0 & A < 21. \end{cases} \\ N &= \frac{0,5(A - 8)}{2,285\Delta\omega}. \end{aligned}$$

4.1.2. Исследование частотных характеристик банка квадратурно-зеркальных фильтров

Рассмотрим применение изложенных выше положений при анализе поведения структуры фильтров, называемой банком квадратурно-зеркальных фильтров. Как будет видно далее здесь потребуется использование связи импульсных характеристик фильтров-прототипов нижних частот и полосно-пропускающих фильтров, теоремы Шеннона-Котельникова, преобразования спектров при дискретизации и интерполяции сигнала.

1. АЧХ пары квадратурных фильтров

На рис. 4.5 приведены АЧХ двух фильтров низких и высоких частот соответственно, связанные следующим соотношением:

$$H_1(e^{j\omega}) = 1 - H_0(e^{j\omega}). \quad (4.1)$$

Очевидно, что $H_1(e^{j\omega})$ смещена относительно $H_0(e^{j\omega})$ на $\pi/2$, вследствие чего пара фильтров называется квадратурно-зеркальными.

Соотношение для импульсных характеристик, имеет вид:

$$h_1(n) = (-1)^n h_0(n). \quad (4.2)$$

Переходя к обозначению $z=e^{j\omega}$, запишем (4.1) в виде:

$$H_1(z) = H_0(-z). \quad (4.3)$$

Для того чтобы восстановить входной сигнал $X(z)$, необходимо объединить его низкочастотные и высокочастотные компоненты (рис. 4.6). Действие блоков прореживания $\downarrow 2$ и интерполяции $\uparrow 2$ на спектр сигнала $X(z)$ будет для наглядности проанализировано на однокаскадной структуре, хотя очевидно, что в этом случае изменения частоты работы фильтров не требуется.

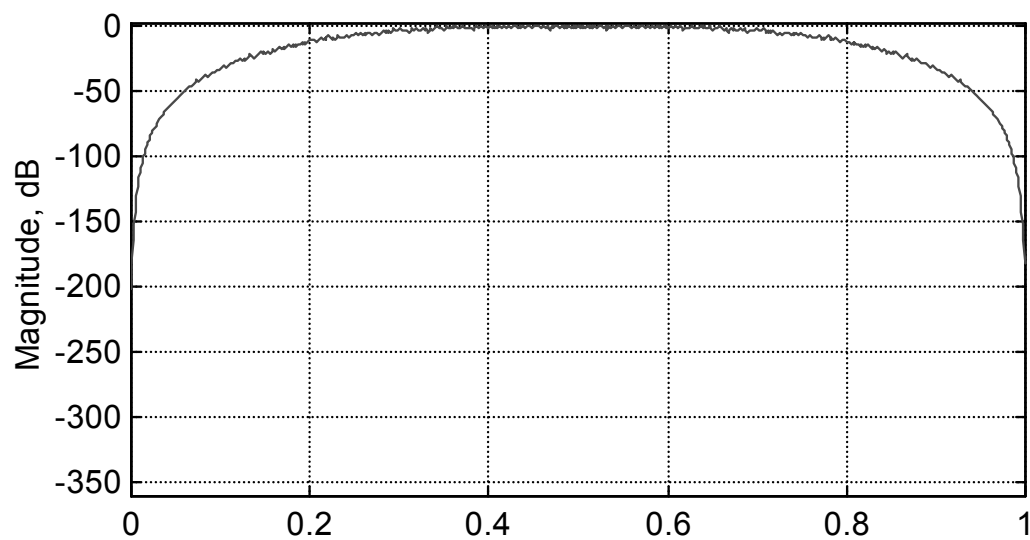
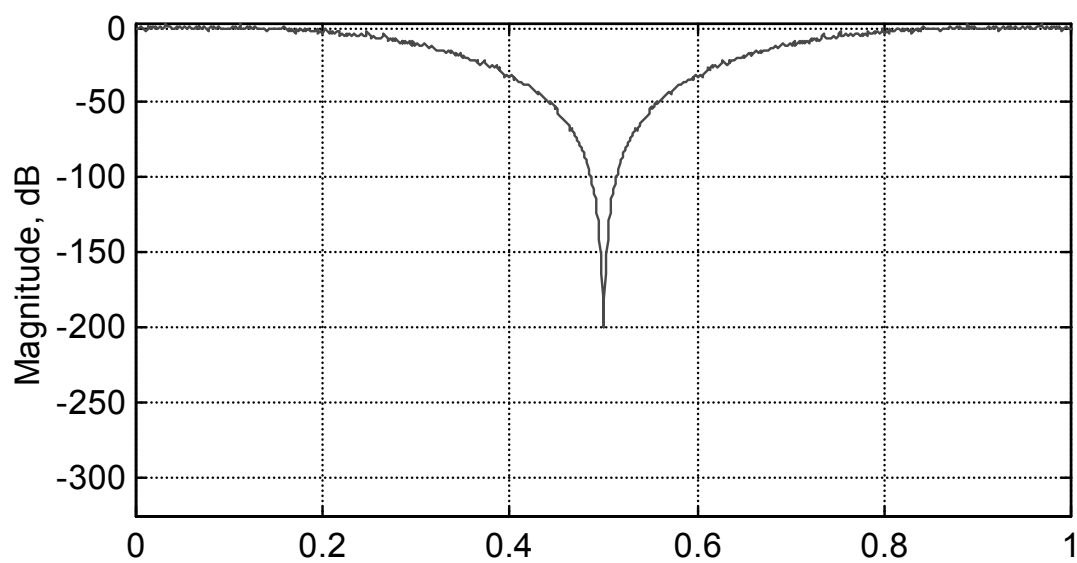


Рис. 4.5. Фрагменты АЧХ квадратурно-зеркальных фильтров

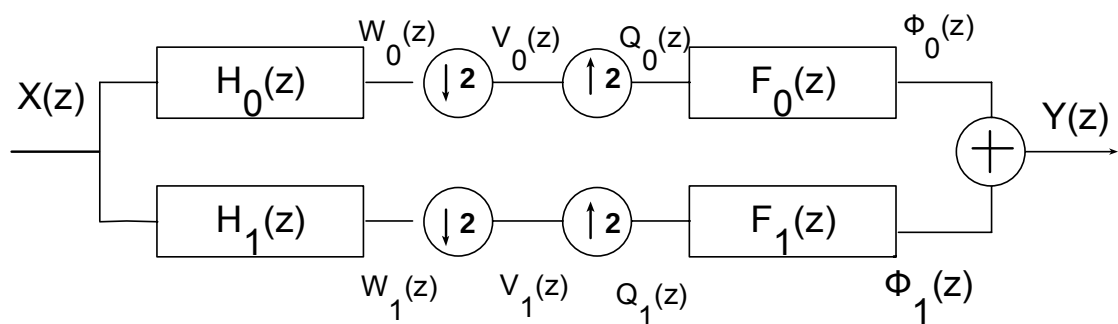


Рис. 4.6 Банк фильтров. Однокаскадная структура

2. Передаточная функция банка фильтров

Рассмотрим подробно стадии изменения сигнала банком фильтров и выведем передаточную функцию этой системы:

$$Y(z) = X(z)T(z). \quad (4.4)$$

Пусть входной сигнал $X(z)$ является равномерно распределенным случайным шумовым сигналом, отсчеты которого взяты с частотой f_s . По горизонтальной оси графиков отложим частоту $f = \omega/(2\pi f_s)$ и приведем АЧХ сигнала.

Вид АЧХ высокочастотной ветви совпадает с низкочастотной с точностью до сдвига на $f_N = f_s/2$, где f_N – частота Найквиста.

1. Разделение сигнала в частотной области (рис. 4.7).

$$W_0(z) = X(z)H_0(z)$$

$$W_1(z) = X(z)H_1(z)$$

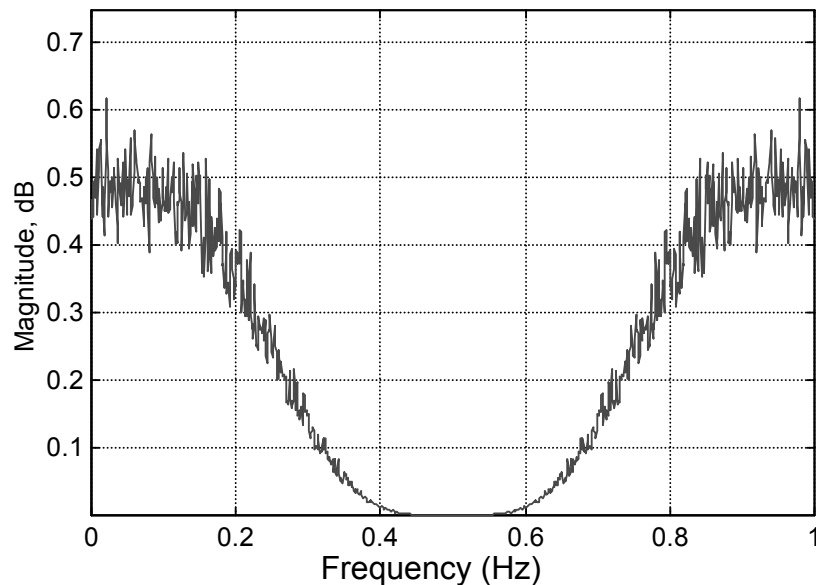


Рис. 4.7. Сигнал на выходе фильтра $H_0(z)$

2. Прореживание.

$$V_0(z) = \downarrow 2W_0(z) = \frac{1}{2}[X(z)H_0(z) + X(-z)H_0(-z)]$$

$$V_1(z) = \downarrow 2W_1(z) = \frac{1}{2}[X(z)H_1(z) + X(-z)H_1(-z)]$$

Снижение частоты дискретизации вдвое приводит к появлению первого дублирующего спектра на f_N (рис. 4.8).

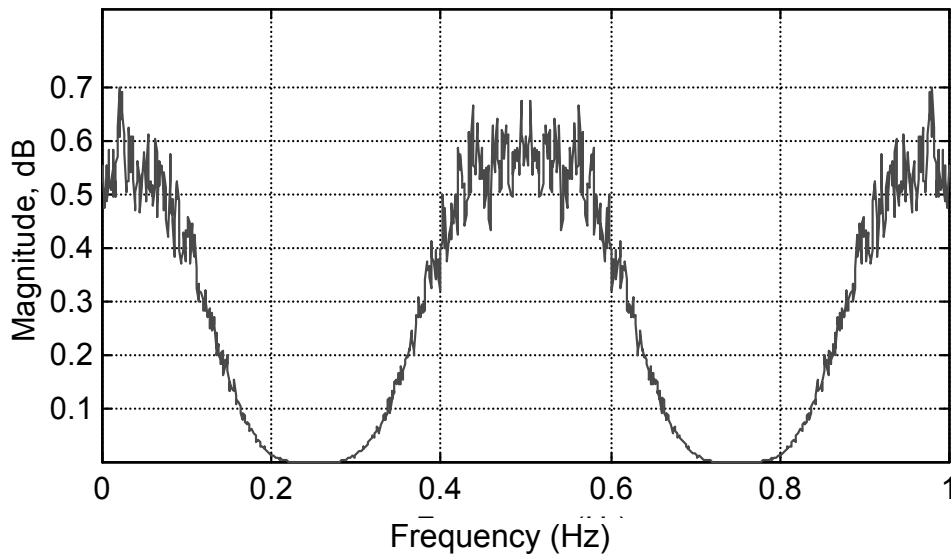


Рис. 4.8. Спектр сигнала после прореживания: период сигнала 0:0.5

3. Интерполяция.

Операция интерполяции в данной задаче состоит в добавлении нулевых отсчетов между отсчетами сигнала:

$$\begin{aligned}
 Q_0(z) &= \uparrow 2 V_0(z) = \uparrow 2 \downarrow 2 \left[\frac{1}{2} [X(z)H_0(z) + X(-z)H_0(-z)] \right] = \\
 &= \frac{1}{2} [X(z)H_0(z) + X(-z)H_0(-z)], \\
 Q_1(z) &= \uparrow 2 V_1(z) = \uparrow 2 \downarrow 2 \left[\frac{1}{2} [X(z)H_1(z) + X(-z)H_1(-z)] \right] = \\
 &= \frac{1}{2} [X(z)H_1(z) + X(-z)H_1(-z)].
 \end{aligned}$$

Значение f_s при этом возрастает до начального значения, однако общий вид спектра сигнала не изменяется, т.к. никакой информации в сигнала не добавляется. Изменение касается только "периода" сигнала в частотной области (рис. 4.9).

Для подавления лишнего дублирующего спектра внутри периода применяется интерполирующий фильтр.

4. Интерполирующая фильтрация (Рис. 4.10)

$$\begin{aligned}
 \Phi_0(z) &= F_0(z)Q_0(z) = F_0(z) \frac{1}{2} [X(z)H_0(z) + X(-z)H_0(-z)] \\
 \Phi_1(z) &= F_1(z)Q_1(z) = F_1(z) \frac{1}{2} [X(z)H_1(z) + X(-z)H_1(-z)]
 \end{aligned}$$

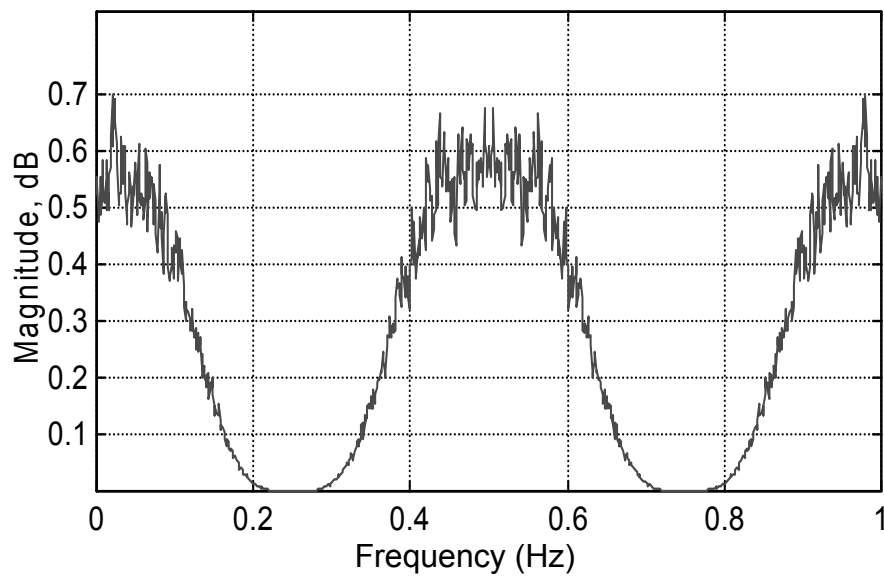


Рис. 4.9. Спектр сигнала после интерполяции: период сигнала 0:1.0

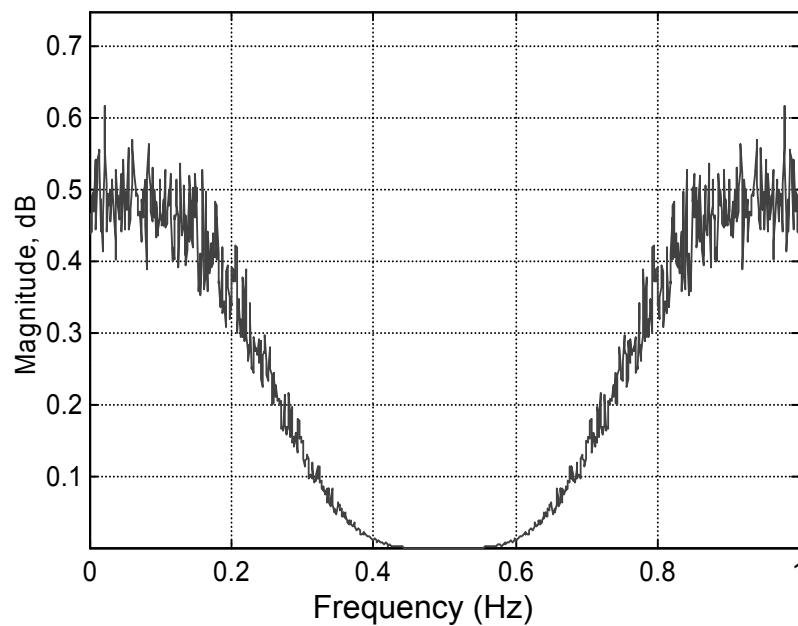


Рис. 4.10. Спектр сигнала после интерполирующего фильтра

5. Восстановление сигнала выполняется в блоке сложения:

$$\begin{aligned}
 Y(z) &= F_0(z) \frac{1}{2} [X(z)H_0(z) + X(-z)H_0(-z)] + \\
 &+ F_1(z) \frac{1}{2} [X(z)H_1(z) + X(-z)H_1(-z)] = \\
 &= \frac{1}{2} [F_0(z)[X(z)H_0(z) + X(-z)H_0(-z)] + \\
 &+ F_1(z)[X(z)H_1(z) + X(-z)H_1(-z)]] .
 \end{aligned}
 \tag{4.5}$$

3. Условия идеального преобразования сигнала банком

Восстановление будет выполнено идеально при выполнении двух условий:

Условие 1:

Банк фильтров обеспечивает только задержку сигнала.

$$H_0(z)F_0(z) + H_1(z)F_1(z) = 2z^{-n} \quad \text{или}$$

$$H_0(z)F_0(z) = H_1(z)F_1(z) = z^{-n}$$

Данное условие может быть выполнено следующим образом.

Представим процесс фильтрации во временной области в виде матрицы $A(z)$ вида:

$$\begin{array}{cccccccc} a(0) & a(1)z^{-1} & a(2)z^{-2} & a(1)z^{-1} & 0 & 0 & 0 & 0... \\ ...0 & a(0) & a(1)z^{-1} & a(2)z^{-2} & a(1)z^{-1} & 0 & 0 & 0... \\ ...0 & 0 & a(0) & a(1)z^{-1} & a(2)z^{-2} & a(1)z^{-1} & 0 & 0... \\ ...0 & 0 & 0 & a(0) & a(1)z^{-1} & a(2)z^{-2} & a(1)z^{-1} & 0... \end{array}$$

Для того, чтобы восстановленный сигнал не имел искажений, т.е.

$$H_0(z)F_0(z) = H_1(z)F_1(z) = 1,$$

необходимо, чтобы при

$$H_i(z) = A(z) \Rightarrow F_i(z) = A^{-1}(z).$$

Однако, для матрицы, записанной по степеням z^{-1} , обратная матрица содержит положительные степени z , что делает интерполирующий фильтр $F_i(z)$ некаузальным. Это устраняется путем преобразования обратной матрицы к виду по степеням z^{-1} и вызывает появление множителя z^{-n} , т.е. задержки (см. Пример). Во временной области это соответствует переносу центрального отсчета импульсной характеристики фильтра с $n = 0$ на $k = N/2$ и нумерации отсчетов не по $n = -N/2..N/2$, а по $k = 0..N$, где N — порядок фильтра.

Пример:

$$\begin{bmatrix} 1 & z^{-1} \\ z^{-1} & -z^{-2} \end{bmatrix} * \begin{bmatrix} 1/2 & 1/2 z \\ 1/2 & -1/2 z^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Обратную матрицу преобразуем к виду:

$$\begin{bmatrix} 1/2 & 1/2 z \\ 1/2 & -1/2 z^2 \end{bmatrix} = z^2 * \begin{bmatrix} 1/2 z^{-2} & 1/2 z^{-1} \\ 1/2 z^{-1} & -1/2 \end{bmatrix}.$$

В общем виде можно записать:

$$A(z) * A^{-1}(z) * z^2 = E \quad \text{или}$$

$$A(z) * A^{-1}(z) = E * z^2,$$

где матрица $A^{-1}(z)$ является уже каузальной.

Таким образом, передаточную функцию каждой ветви системы можно получить равной единице с точностью до задержки, что обеспечивает выполнение условия 1.

Условие 2:

Необходимо устранить последствия наложения спектров из-за изменения частоты дискретизации путем адекватного выбора характеристик интерполирующих фильтров.

Согласно (4.5) для выполнения этого условия необходимо устранить компоненты $T(z)$, соответствующие лишним дублирующим спектрам, т.е. зависящие от $-z$. Это достигается установлением следующего соотношения между фильтрами, образующими банк:

$$H_0(-z)F_0(z) + H_1(-z)F_1(z) = 0. \quad (4.6)$$

Условие 2 выполняется, если положить

$$F_0(z) = H_0(-z) \quad \text{и} \quad F_1(z) = -H_0(-z). \quad (4.7)$$

(4.5) и (4.7) позволяют свести методику синтеза банка фильтров только к нахождению $H_0(z)$, или АЧХ фильтра-прототипа.

Учитывая соотношение (4.7), можно переписать (4.5) в виде:

$$\begin{aligned} Y(z) &= 1/2 [H_0(z)F_0(z) + H_1(z)F_1(z)]X(z) = \\ &= \left[1/\sqrt{2} H_0(z) 1/\sqrt{2} F_0(z) + 1/\sqrt{2} H_1(z) 1/\sqrt{2} F_1(z) \right] X(z) \end{aligned} \quad (4.8)$$

или

$$Y(z) = 1/2 [T_0(z) + T_0(-z)]X(z), \quad (4.9)$$

где:

$$T_0(z) = H_0(z)F_0(z) = H_0(z)H_0(-z) = |H_0(z)|^2. \quad (4.10)$$

Пример синтеза импульсных характеристик фильтров в составе банка

Синтез ИХ банка фильтров состоит в определении квадрата передаточной функции $T_0(z) = |H_0(z)|^2$, ее факторизации для получения $H_0(z)$ и определении $F_0(z) = H_0(-z)$, $H_1(z) = F_0(-z)$ и $F_1(z) = -H_0(-z)$.

Так как передаточная функция всей системы должна быть все пропускающей, то не ограничивая общности, можно положить:

$$|H_0(z)|^2 + |H_0(-z)|^2 = 1 \text{ или} \\ |H_0(\omega)|^2 + |H_0(\omega + \pi)|^2 = 1. \quad (4.11)$$

Вид (4.11) позволяет предположить, что решение $H_0(\omega)$ существует в виде тригонометрического полинома от ω . С другой стороны, очевидно, что для качественного субполосного разделения каналов $H_0(\omega)$ необходим резкий перепад между полосами пропускания и заграждения субполосных фильтров. То есть частотная характеристика $H_0(\omega)$ фильтра нижних частот должна иметь ноль некоторой кратности N на частоте $\omega = \pi$. Чем выше порядок кратности нуля АЧХ, тем она будет более близка к прямоугольной. Это достигается выбором $H_0(\omega) = (1 + e^{j\omega})^N \Rightarrow H_0(z) = (1 + z)^N$ – биномиальный полином от z , имеющий корень кратности N на $z = -1$.

Для того чтобы объединить эти соображения, рассмотрим следствие формулы Эйлера для комплексной экспоненты:

$$\left| \frac{1 + e^{j\omega}}{2} \right|^2 = \cos^2\left(\frac{\omega}{2}\right) \\ \text{т. к. } \operatorname{Re}\left[\frac{1 + e^{j\omega}}{2}\right] = \frac{1 + \cos \omega}{2} \quad \operatorname{Im}\left[\frac{1 + e^{j\omega}}{2}\right] = \frac{1 + \sin \omega}{2}, \text{ то} \\ \left| \frac{1 + e^{j\omega}}{2} \right|^2 = \frac{(1 + \cos \omega)^2 + (1 + \sin \omega)^2}{4} = \frac{2 + 2\cos \omega}{4} = \frac{1 + \cos \omega}{2} = \cos^2\left(\frac{\omega}{2}\right),$$

С учетом свойств тригонометрических функций и (4.11) это выражение может быть переписано в виде:

$$\left[\cos^2\left(\frac{\omega}{2}\right)\right]^N P\left[\sin^2\left(\frac{\omega}{2}\right)\right] + \left[\sin^2\left(\frac{\omega}{2}\right)\right]^N P\left[\cos^2\left(\frac{\omega}{2}\right)\right] = 1 \text{ или}$$

$$(1-y)^N P(y) + y^N P(1-y) = 1, \text{ где } y = \cos^2\left(\frac{\omega}{2}\right), \quad y \in [0,1].$$

Для нахождения $P(y)$ в явном виде потребуется результат теоремы Безу.

Теорема Безу:

Для двух полиномов p_1, p_2 степени соответственно n_1, n_2 при условии, что они не имеют общих корней, существуют единственные полиномы q_1, q_2 степени соответственно $n_1 - 1, n_2 - 2$ такие, что $p_1(x)q_1(x) + p_2(x)q_2(x) = 1$.

Согласно теореме Безу (4.11) преобразуем к виду

$$(1-y)^N q_2(1-y) + y^N q_1(1-y) = 1$$

в силу единственности полиномов $q_2(1-y) = q_1(y)$, тогда

$$(1-y)^N q_1(1-y) + y^N q_1(1-y) = 1,$$

откуда:

$$q_1(y) = (1-y)^{-N} \left[1 - y^N q_1(1-y) \right] = (1-y)^{-N} - (1-y)^{-N} y^N q_1(1-y).$$

Используя разложение в ряд Тейлора для $(1-y)^{-N}$, запишем:

$$q_1(y) = \sum C_k^{N+k-1} y^k + o(y^N).$$

Т.к. степень $q_1(y)$ равна $N-1$, то ограничимся:

$$q_1(y) = \sum C_k^{N+k-1} y^k, \text{ где } k = 0 \dots N-1,$$

откуда:

$$P(y) = \sum C_k^{N+k-1} y^k, \text{ где } k = 0 \dots N-1 \text{ — биномиальный полином.}$$

Этот результат позволяет в явном виде записать $T_0[\cos(\omega)] = |H_0(\omega)|^2$. Для получения $H_0(\omega)$ выполняется факторизация полинома $T_0[\cos(\omega)]$ на основании леммы Рица.

Лемма Рица:

Для полинома вида $A(\omega) = \sum a_m \cos(m\omega)$, $a_m \in R, m = 0 \dots M$ существует полином $B(\omega) = \sum b_m e^{jm\omega}$, $b_m \in R, m = 0 \dots M$ такой, что $|B(\omega)|^2 = A(\omega)$

Покажем расчет фильтра-прототипа для $N=2$.

Имеем:

$$T[\cos \omega] = y^N P(1-y) = \left[\cos^2\left(\frac{\omega}{2}\right) \right]^N P\left[\sin^2\left(\frac{\omega}{2}\right)\right]$$
$$H_0(\omega) = \left[\frac{(1 + e^{j\omega})}{2} \right]^N R(e^{j\omega}), \text{ где } |R(e^{j\omega})|^2 = P(1-y) = P\left[\sin^2\left(\frac{\omega}{2}\right)\right],$$

тогда:

$$P(y) = 1 + 2y$$

$$P(1-y) = 1 + 2 - 2y = 3 - 2y, \quad \text{т.к. } y = \cos^2\left(\frac{\omega}{2}\right)$$

$$P(1-y) = 3 - 2\cos^2\left(\frac{\omega}{2}\right) = 3 - (1 + \cos \omega) = 2 \cos \omega = \frac{4 - e^{j\omega} + e^{j\omega}}{2} =$$
$$= \frac{(-t^2 + 4t + 1)}{2}, \quad \text{где } t = e^{j\omega}.$$

$P(1-y)$ имеет два корня $t_{1,2} = 2 \pm \sqrt{3}$, тогда

$$P(y) = \frac{1}{2} \left[(e^{j\omega} - 2 - \sqrt{3})(e^{j\omega} - 2 + \sqrt{3}) \right],$$

для $R(e^{j\omega})$ выберем корень $t_{1,2} = 2 - \sqrt{3}$, лежащий в пределах единичной окружности (построим минимально фазовый фильтр):

$$R(e^{j\omega}) = (e^{j\omega} - 2 + \sqrt{3}) = \left[(1 + \sqrt{3})e^{j\omega} + 1 - \sqrt{3} \right],$$

тогда $H_0(\omega)$ примет вид:

$$H_0(\omega) = \left[\frac{1}{2} (1 + e^{j\omega}) \right]^2 \left[(1 + \sqrt{3})e^{j\omega} + 1 - \sqrt{3} \right].$$

Перейдем к $z = e^{j\omega}$, тогда

$$H_0(z) = \frac{1}{4}(1+z)^2[(1+\sqrt{3})z + 1 - \sqrt{3}] =$$

$$= \frac{1+\sqrt{3}}{4}z^3 + \frac{3+\sqrt{3}}{4}z^2 + \frac{3-\sqrt{3}}{4}z + \frac{1-\sqrt{3}}{4}.$$

После нормировки умножением на $1/\sqrt{2}$ и округления получим коэффициенты фильтра $h_0(n)$ в окончательной форме $h_0(n) = \{0,48 \quad 0,84 \quad 0,22 \quad -0,13\}$, по которым определяются коэффициенты трех оставшихся фильтров банка.

Приведенная процедура расчета теоретически может быть выполнена для произвольно больших четных значений N . Длина полученных фильтров составляет $2N$.

5. Оценка погрешностей банка квадратурных фильтров

Ранее нами были установлены условия устранения искажений сигнала в канале. Передаточная функция системы имеет вид:

$$Y(z) = 1/2[T_0(z) + T_0(-z)]X(z), \text{ где}$$

$$T_0(z) = H_0(z)F_0(z) = H_0(z)H_0(-z) = |H_0(z)|^2.$$

Перепишем это соотношение в виде:

$$|A(z)| = |H_0(z)|^2 + |H_0(-z)|^2 = 1$$

или

$$|A(e^{j\omega})| = |H_0(e^{j\omega})|^2 + |H_0(e^{j\omega+\pi})|^2 = 1.$$

Введем количественную характеристику погрешности восстановления сигналов банком фильтров в виде:

$$e_p(\omega) = |A(e^{j\omega})| - 1 = \left[|H_0(e^{j\omega})|^2 + |H_0(e^{j\omega+\pi})|^2 \right] - 1,$$

тогда погрешность всей системы имеет вид:

$$E(\omega) = E_p(\omega) + \alpha \cdot E_s(\omega),$$

где $E_p(\omega) = \int_0^\pi [e(\omega)]^2 d\omega$ — энергетическая характеристика искажений

сигнала, α – весовой множитель, $E_s(\omega = \int_{\omega_S}^{\pi} |H_0(e^{j\omega})|^2 d\omega$ – мощность сигнала в полосе заграждения.

4.2. МОДУЛЯЦИЯ

4.2.1. Модуляция аналоговых сигналов

Сигналы от любых источников информации передаются по линиям связи к приемникам, например, в измерительно-вычислительные системы регистрации и обработки данных. Как правило, информационные сигналы являются низкочастотными и ограниченными по ширине спектра, тогда как методы передачи сигналов рассчитаны на работу с высокочастотным сигналом [5]. При этом важным вопросом является частотное разделение каналов передачи информации с целью эффективного использования каналообразующего оборудования и выделенного для передачи частотного диапазона. Перенос спектра сигналов из низкочастотной области на заданную частоту, т.е. в выделенную для их передачи область высоких частот выполняется операцией *модуляции*.

Обозначим низкочастотный сигнал, подлежащий передаче по какому-либо каналу связи, $s(t)$ (рис. 4.11). В канале связи для передачи данного сигнала выделяется определенный диапазон высоких частот и формируется вспомогательный периодический высокочастотный сигнал $u(t) = f(t; a_1, a_2, \dots, a_m)$. Совокупность параметров a_i определяет форму вспомогательного сигнала. Значения параметров a_i в отсутствие модуляции являются величинами постоянными. Если на один из этих параметров перенести сигнал $s(t)$, т.е. сделать его значение пропорционально зависимым от значения $s(t)$ во времени (или по любой другой независимой переменной), то форма сигнала $u(t)$ приобретает новое свойство. Она служит для переноса информации, содержащейся в сигнале $s(t)$. Сигнал $u(t)$ называется несущим сигналом, *несущим колебанием* или просто *несущей* (carrier), а физический процесс переноса информации на

параметры несущего сигнала – его *модуляцией*. Исходный информационный сигнал $s(t)$ называют *модулирующим*, результат модуляции – *модулированным сигналом*. Обратную операцию выделения модулирующего сигнала из модулированного колебания называют *демодуляцией* или *детектированием*.

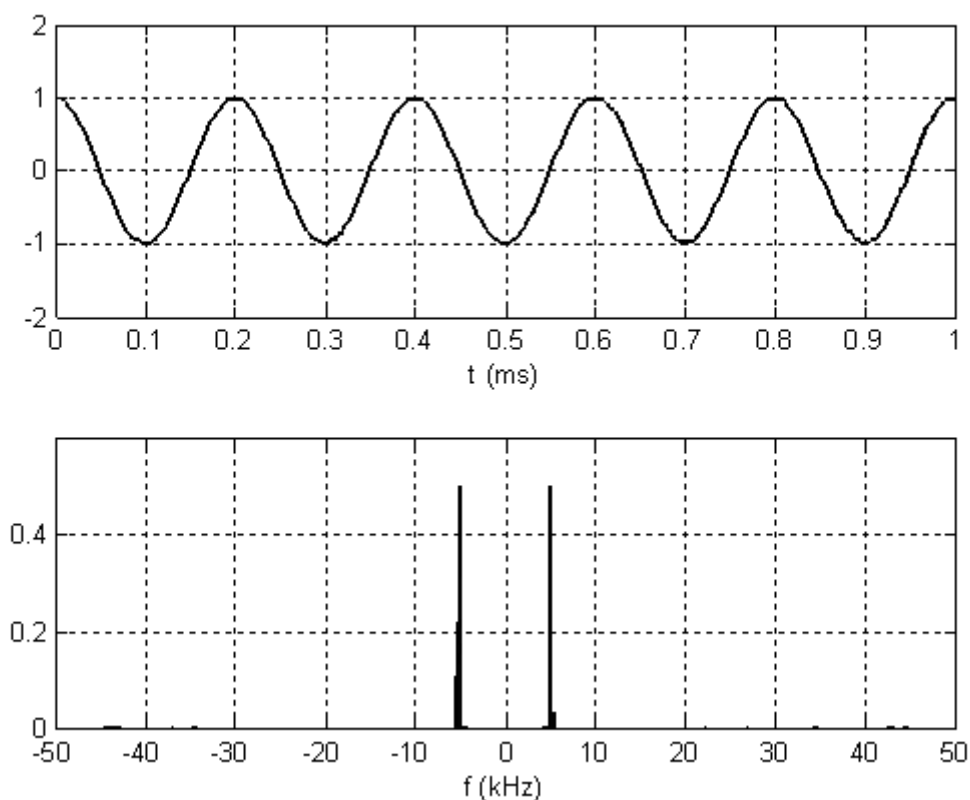


Рис 4.11. Сигнал $s(t) = \cos(\Omega t)$ ($\Omega = 2\pi \cdot 5$ кГц)

Наиболее распространенной формой несущих сигналов являются гармонические колебания:

$$u(t) = U \cos(\omega \cdot t + \varphi),$$

которые имеют три свободных параметра: U , ω и φ . В зависимости от того, на какой из данных параметров переносится информация, различают *амплитудную (АМ)*, *частотную (ЧМ)* или *фазовую (ФМ) модуляцию* несущего сигнала [5]. Частотная и фазовая модуляция тесно взаимосвязаны, поскольку изменяют аргумент функции косинуса, и их обычно объединяют под общим названием – *угловая модуляция*. При одновременной модуляции амплитуды и фазы

несущей возникает *квадратурная* модуляция, основным свойством которой является достижение более высокой скорости передачи по сравнению с методами модуляции одного параметра несущей.

При использовании в качестве несущих сигналов периодических последовательностей импульсов (например, прямоугольных) свободными параметрами модуляции могут быть амплитуда, длительность, частота следования и фаза (положение импульса относительно тактовой точки) импульсов. Таким образом, существует четыре основных вида импульсной модуляции: АИМ, ЧИМ и ФИМ.

В качестве несущих сигналов можно использовать не только периодические колебания, но и стационарные случайные процессы. В качестве модулируемых параметров случайных сигналов используются моменты случайных процессов. Так, например, модуляция второго момента случайных последовательностей (модуляция по мощности) представляет собой аналогию амплитудной модуляции [5].

1. Амплитудная модуляция/демодуляция

Амплитудная модуляция (*amplitude modulation*) исторически была первым видом модуляции, освоенным на практике. В настоящее время АМ применяется в основном только для радиовещания на сравнительно низких частотах (не выше коротких волн) и для передачи изображения в телевизионном вещании. Это вызвано низким КПД использования энергии модулированных сигналов.

При АМ выполняется перенос информации $s(t) \Rightarrow U(t)$ при постоянных значениях параметров несущей частоты ω и φ . АМ – сигнал представляет собой произведение информационной огибающей $U(t)$ и гармонического колебания ее заполнения с более высокими частотами:

$$U(t) = U_m [1 + Ms(t)], \quad (4.12)$$

где $U(t) = U_m [1 + Ms(t)]$, U_m – постоянная амплитуда несущего колебания при отсутствии входного (модулирующего) сигнала $s(t)$, M – глубина АМ.

В простейшем случае, если модулирующий сигнал (рис. 4.11) представлен одночастотным гармоническим колебанием с амплитудой S_o , то коэффициент модуляции равен отношению амплитуд модулирующего и несущего колебания $M=S_o/U_m$. Значение M должно находиться в пределах от 0 до 1 для всех гармоник модулирующего сигнала. При значении $M<1$ форма огибающей несущего колебания полностью повторяет форму модулирующего сигнала $s(t)$ (рис. 4.12).

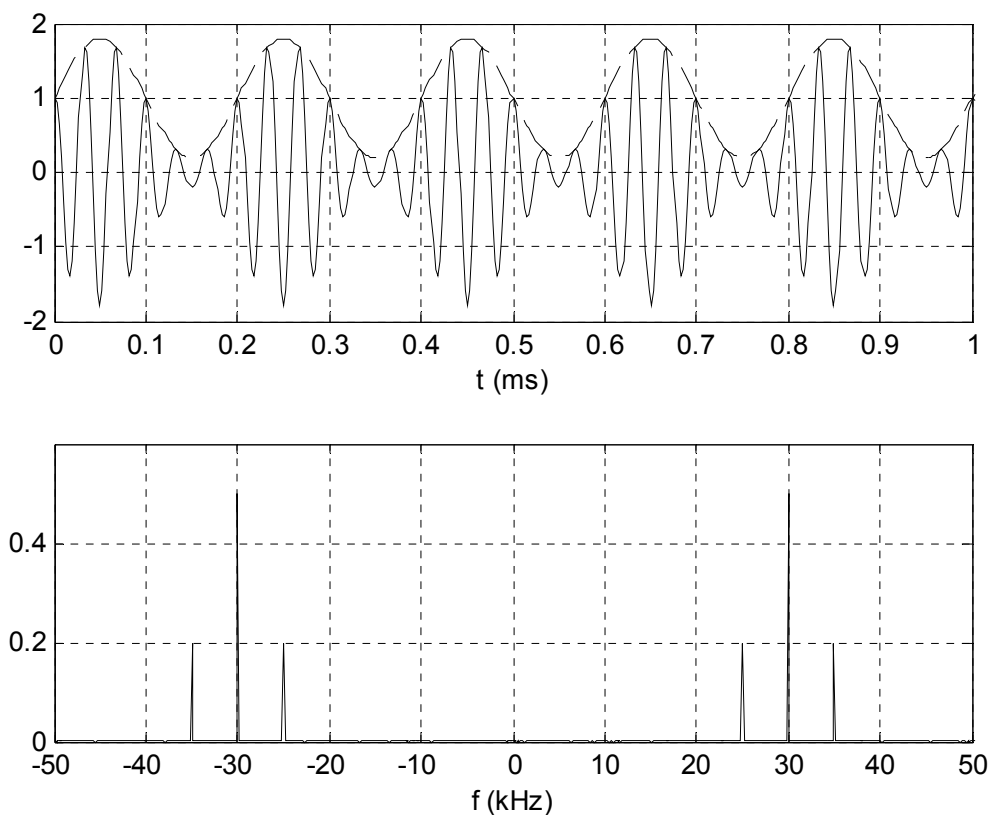


Рис 4.12. Амплитудная модуляция (АМ) ($\omega_0 = 30$ кГц, $\Omega = 5$ кГц)

Простейшая форма модулированного сигнала создается при *однотональной* амплитудной модуляции – модуляции несущего сигнала гармоническим колебанием с одной частотой Ω :

$$u(t) = U_m [1 + M \cos(\Omega t)] \cos(\omega_0 \cdot t) \quad (4.13)$$

Значения начальной фазы углов примем равными нулю. Поскольку $\cos(x) \cdot \cos(y) = (1/2)[\cos(x+y) + \cos(x-y)]$, из выражения (4.13) получаем:

$$u(t) = U_m \cos(\omega_0 t) + \frac{U_m M}{2} \cos[(\omega_0 + \Omega)t] + \frac{U_m M}{2} \cos[(\omega_0 - \Omega)t] \quad (4.14)$$

Соотношение (4.14) называется *основной теоремой модуляции*: модулирующее колебание с частотой Ω перемещается в область частоты ω_0 и расщепляется на два колебания, симметричные относительно частоты ω_0 , с частотами соответственно $(\omega_0 + \Omega)$ – верхняя боковая частота, и $(\omega_0 - \Omega)$ – нижняя боковая частота. Физическая ширина спектра модулированного сигнала в два раза больше ширины спектра сигнала модуляции.

Коэффициент полезного действия данного типа модуляции определяется отношением мощности боковых частот к общей средней мощности модулированного сигнала

$$\eta_{AM} = \frac{\frac{U_m^2 M^2}{4}}{P_U} = \frac{M^2}{M^2 + 2}. \quad (4.15)$$

Отсюда следует, что при $M=1$ КПД амплитудной модуляции составляет только 33%, а на практике обычно меньше 20%.

Многотональный модулирующий сигнал имеет произвольный спектральный состав:

$$s(t) = \sum_{n=1}^N a_n \cos(\Omega_n \cdot t + \Phi_n), \quad (4.16)$$

где значения амплитуд a_n и начальных фаз Φ_n упорядоченной возрастающей последовательности гармоник Ω_n произвольны. Подставляя (4.16) в (4.12), получим обобщенное уравнение амплитудно-модулированного сигнала и его физического спектра:

$$u(t) = U_m \left[1 + \sum_{n=1}^N M_n \cos(\Omega_n \cdot t + \Phi_n) \right] \cos(\omega_0 \cdot t) \quad (4.17)$$

$$\begin{aligned}
u(t) = & U_m \cos(\omega_0 \cdot t + \varphi_0) + \\
& + \frac{U_m}{2} \sum_{n=1}^N M_n \cos[(\omega_0 + \Omega_n) \cdot t + \varphi_0 + \Phi_n] + \\
& + \frac{U_m}{2} \sum_{n=1}^N M_n \cos[(\omega_0 + \Omega_n) \cdot t + \varphi_0 - \Phi_n]
\end{aligned} \tag{4.18}$$

2. Балансная амплитудная модуляция

Как следует из вышеприведенных данных, основная доля мощности АМ-сигнала приходится на несущую частоту. При балансной модуляции (или АМ с подавлением несущей частоты (АМ-ПН)) производится перемножение двух сигналов – модулирующего и несущего, при котором происходит подавление несущего колебания, соответственно, КПД модуляции становится равным 100%. Так, для однотонового сигнала (без учета начальных фаз колебаний) при $U(t) = M \cdot \cos(\Omega t)$ имеем:

$$\begin{aligned}
u(t) = & U_m M \cos(\Omega t) \cos(\omega_0 \cdot t) = \\
= & \frac{U_m M}{2} \{ \cos[(\omega_0 + \Omega) \cdot t] + \cos[(\omega_0 - \Omega) \cdot t] \},
\end{aligned} \tag{4.19}$$

т.е. два одинаковых по амплитуде гармонических сигнала с верхней и нижней боковыми частотами. Пример сигнала с балансной модуляцией приведен на рис. 4.13 сверху. Амплитудный спектр сигнала подобен приведенному на рис. 4.13 внизу с отсутствующей несущей частотой ω_0 .

Аналогично, многотональный балансно-модулированный сигнал имеет две симметричные относительно частоты ω_0 группы верхних и нижних боковых колебаний:

$$\begin{aligned}
u(t) = & \frac{U_m M}{2} \left(\sum_{n=1}^N M_n \cos[(\omega_0 + \Omega_n) \cdot t + \Phi_n] + \right. \\
& \left. + \sum_{n=1}^N M_n \cos[(\omega_0 + \Omega_n) \cdot t - \Phi_n] \right).
\end{aligned} \tag{4.20}$$

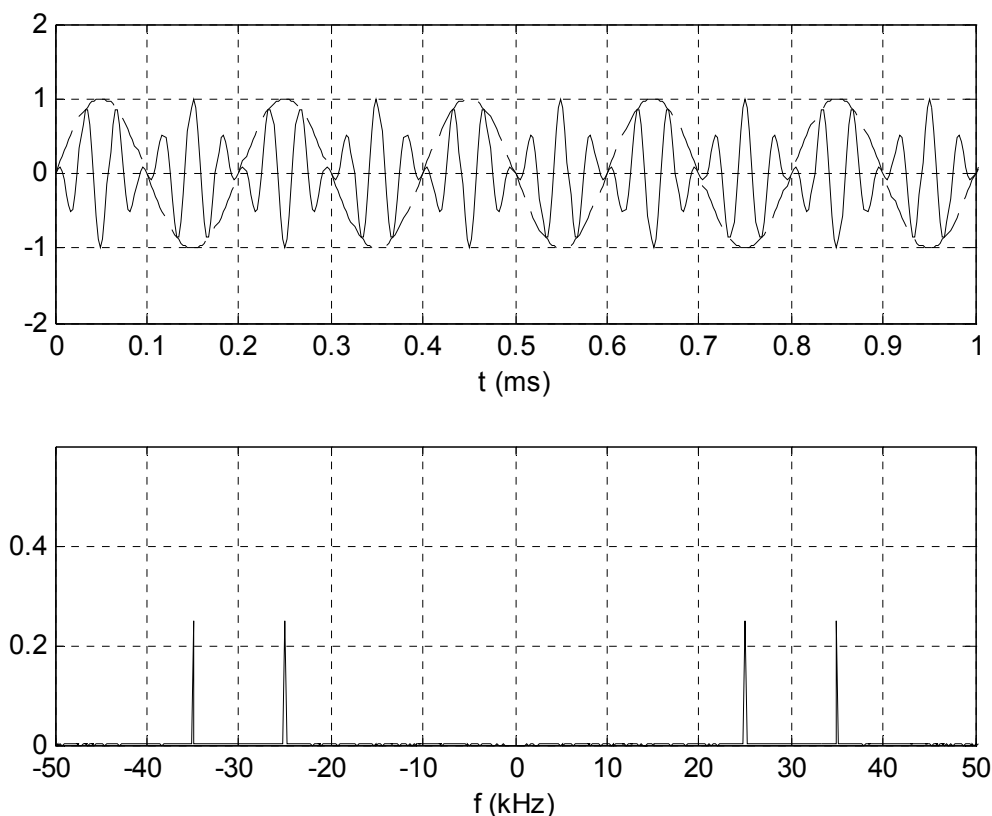


Рис 4.13. Балансная амплитудная (DSB-SC) модуляция
 $(\omega_0 = 30 \text{ кГц}, \Omega = 5 \text{ кГц})$

Физическая сущность подавления несущей частоты заключается в следующем. При переходе огибающей биений $U(t)$ через нуль фаза несущей частоты скачком изменяется на 180° , поскольку функция косинуса огибающей имеет разные знаки слева и справа от нуля. При этом в достаточно высокочастотной системе (с малыми потерями энергии), настроенной на частоту ω_0 , колебания, возбужденные одним периодом биений, будут гаситься последующим периодом.

Однако балансная модуляция не получила широкого распространения в связи с трудностями, возникающими при демодуляции сигналов. Синхронное детектирование, как и для АМ, позволяет выполнять демодуляцию без каких-либо проблем, но при условии известной частоты несущей сигнала и точной фазовой синхронизации опорной частоты с несущей. Но во входном сигнале демодулятора при АМ-ПН несущая частота отсутствует, и

автоматическая синхронизация становится невозможной. Для снятия этой трудности обычно применяют неполное подавление несущей и оставляют в модулированном сигнале определенный "остаток" несущей (пилот-сигнал), который и используется для фазочастотной автосинхронизации при демодуляции.

3. Однополосная амплитудная модуляция

При идентичности информации в группах верхних и нижних боковых частот нет необходимости в их одновременной передаче. Одна из них перед подачей сигнала в канал связи может быть удалена, чем достигается двукратное сокращение полосы занимаемых сигналом частот. Уравнение сигнала с одной боковой полосой (ОБП – сигнал, *single side band* – SSB) может быть получено непосредственно из (4.18). Для верхней или нижней боковой полосы имеем:

$$u(t) = U_m \cos(\omega_0 t + \varphi_0) + \frac{U_m}{2} \sum_{n=1}^N M_n \cos[(\omega_0 \pm \Omega_n) \cdot t \pm \varphi_0 \pm \Phi_n] \quad (4.21)$$

Внешняя форма ОБП-сигнала сходна с обычным АМ-сигналом, но ее огибающая имеет в два раза меньшую амплитуду по сравнению с АМ при $M = 1$ (рис. 4.14).

Для демодуляции ОБП-сигнала может использоваться как двухполупериодное, так и синхронное детектирование, со всеми особенностями, присущими этим методам. Результаты демодуляции отличаются от демодуляции АМ – сигналов только в два раза меньшей амплитудой выходных сигналов.

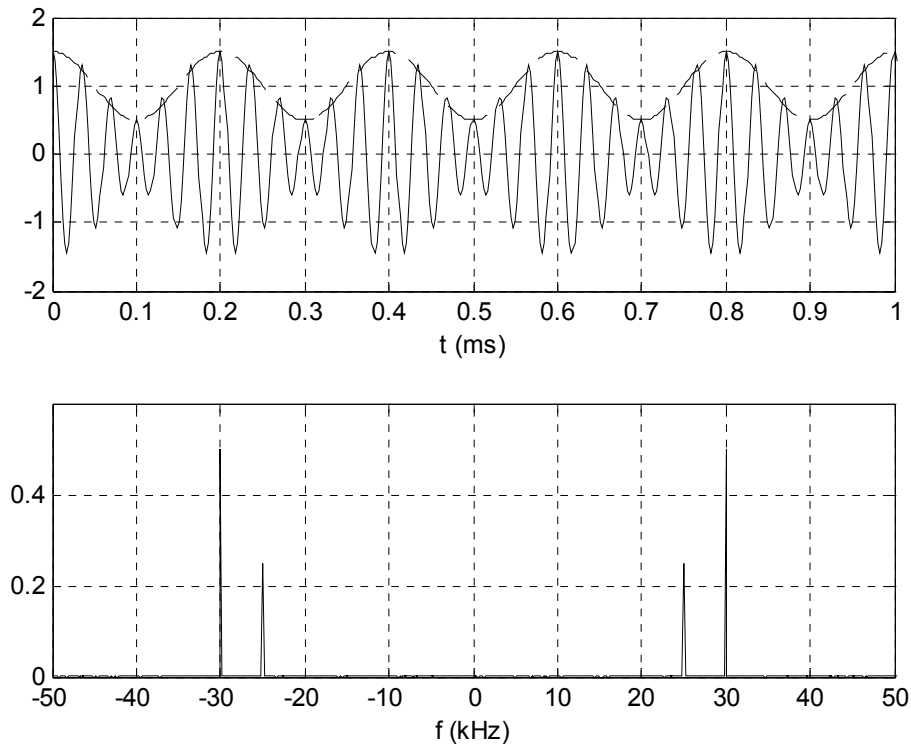


Рис 4.14. Однополосная амплитудная (SSB) модуляция
 $(\omega_0 = 30 \text{ кГц}, \Omega = 5 \text{ кГц})$

При однополосной модуляции возможно также подавление несущей частоты (полное или частичное, с передачей пилот-сигнала), что позволяет повысить КПД передатчика.

4. Угловая модуляция

При угловой модуляции в несущем гармоническом колебании $u(t) = U_m \cos(\omega \cdot t + \varphi)$ значение амплитуды колебаний U_m остается постоянным, а информация $s(t)$ переносится либо на частоту ω , либо на фазовый угол φ . В обоих случаях текущее значение фазового угла гармонического колебания $u(t)$ определяет аргумент $\psi(t) = \omega \cdot t + \varphi$, который называется *полной фазой* колебания.

5. Фазовая модуляция

При фазовой модуляции (*phase modulation* – PM) значение фазового угла постоянной несущей частоты колебаний ω_0 пропорционально амплитуде модулирующего сигнала $s(t)$. Соответственно, уравнение ФМ-сигнала определяется выражением:

$$u(t) = U_m \cos[\omega_0 t + k \cdot s(t)],$$

где k – коэффициент пропорциональности.

При $s(t) = 0$, ФМ-сигнал является простым гармоническим колебанием. С увеличением значений $s(t)$ полная фаза колебаний $\psi(t) = \omega_0 t + k \cdot s(t)$ нарастает во времени быстрее и опережает линейное нарастание $\omega_0 t$. Соответственно, при уменьшении значений $s(t)$ скорость роста полной фазы во времени спадает. В моменты экстремальных значений $s(t)$ абсолютное значение фазового сдвига $\Delta\psi$ между ФМ-сигналом и значением $\omega_0 t$ немодулированного колебания также является максимальным и носит название *девиации фазы* (вверх $\Delta\varphi_v = k \cdot s_{\max}(t)$, или вниз $\Delta\varphi_n = k \cdot s_{\min}(t)$ с учетом знака экстремальных значений модулирующего сигнала).

Для колебаний с угловой модуляцией применяется также понятие мгновенной частоты, под которой понимают производную от полной фазы по времени:

$$\omega(t) = \frac{d}{dt} \psi(t) = \omega_0 + k \cdot \frac{d}{dt} s(t).$$

Полная фаза колебаний в произвольный момент времени может быть определена интегрированием мгновенной частоты:

$$\psi(t) = \int_{-\infty}^t \omega(t) dt \text{ или } \psi(t) = \int_{-\infty}^t \omega(t) dt + \varphi_0.$$

6. Частотная модуляция

Частотная модуляция (*frequency modulation* – FM) характеризуется линейной связью модулирующего сигнала с мгновенной частотой колебаний, при которой мгновенная частота колебаний образуется сложением частоты высокочастотного несущего колебания ω_0 со значением амплитуды модулирующего сигнала с определенным коэффициентом пропорциональности k :

$$\omega(t) = \omega_0 + k \cdot s(t).$$

Соответственно, полная фаза колебаний:

$$\psi(t) = \omega_0(t) + k \int_{-\infty}^t s(t)dt \text{ или } \psi(t) = \omega_0(t) + k \int_0^t s(t)dt + \varphi_0.$$

Уравнение ЧМ-сигнала:

$$u(t) = U_m \cos \left(\omega_0(t) + k \int_0^t s(t)dt + \varphi_0 \right)$$

Аналогично ФМ, для характеристики глубины частотной модуляции используются понятия *девиации частоты вверх* $\Delta\omega_e = ks_{\max}(t)$, и *вниз* $\Delta\omega_e = ks_{\min}(t)$.

7. Однотональная угловая модуляция

Рассмотрим гармонический модулирующий сигнал с постоянной частотой колебаний Ω . Начальная фаза колебаний:

$$\varphi(t) = \beta \sin(\Omega t),$$

где β – индекс угловой модуляции, которым задается интенсивность колебаний начальной фазы. Полная фаза модулированного сигнала с учетом несущей частоты ω_o :

$$\psi(t) = \omega_0 t + \beta \sin(\Omega t). \quad (4.22)$$

Уравнение модулированного сигнала:

$$u(t) = U_m \cos(\omega_0 t + \beta \sin(\Omega t)). \quad (4.23)$$

Мгновенная частота колебаний:

$$\omega(t) = \frac{d}{dt} \psi(t) = \omega_0 + \beta \Omega \sin(\Omega t). \quad (4.24)$$

Как следует из формул (4.22) (4.23) (4.24) и начальная фаза, и мгновенная частота изменяется по гармоническому закону. Максимальное отклонение от среднего значения ω_o равно $\omega_d = \beta \Omega$, и получило название *девиации частоты*. Индекс угловой модуляции равен отношению девиации частоты к частоте модулирующего сигнала:

$$\beta = \frac{\omega_d}{\Omega}.$$

Различия между частотной и фазовой модуляцией проявляются при изменении частоты Ω модулирующего сигнала.

При фазовой модуляции девиация частоты прямо пропорциональна Ω , а индекс угловой модуляции от частоты модулирующего сигнала не зависит:

$$\beta = \text{const}, \quad \omega_d = \beta\Omega.$$

При ЧМ постоянным параметром модуляции является девиация частоты, при этом индекс модуляции обратно пропорционален частоте модулирующего сигнала:

$$\omega_d = \text{const}, \quad \beta = \frac{\omega_d}{\Omega}$$

Формулу (4.24) однотоновой модуляции можно преобразовать к виду:

$$\begin{aligned} u(t) = & U_m \cos(\beta \sin(\Omega t)) \cos(\omega_0 t) - \\ & - U_m \sin(\beta \sin(\Omega t)) \sin(\omega_0 t). \end{aligned} \quad (4.25)$$

При малых значениях индекса угловой модуляции ($\beta \ll 1$, узкополосная модуляция) имеют место приближенные равенства:

$$\cos(\beta \sin(\Omega t)) \approx 1, \quad \sin(\beta \sin(\Omega t)) \approx \beta \sin(\Omega t)$$

При их использовании в (4.25), получаем:

$$\begin{aligned} u(t) \approx & U_m \cos(\omega_0 t) + \frac{\beta U_m}{2} \cos[(\omega_0 + \Omega) \cdot t] + \\ & + \frac{-\beta U_m}{2} \cos[(\omega_0 - \Omega) \cdot t]. \end{aligned} \quad (4.26)$$

Как можно заметить (рис. 4.15), амплитудные спектры однотоновых ФМ и ЧМ сигналов при $\beta \ll 1$ практически аналогичны АМ сигналам и также содержат верхнюю и нижнюю боковые частоты $\omega_0 + \Omega$ и $\omega_0 - \Omega$.

С ростом индекса модуляции полоса частот, занимаемая сигналом, расширяется. Практическая ширина спектра сигнала с угловой модуляцией определяется по формуле:

$$\Pi_{\text{практ}} = 2(\beta + 1)\Omega.$$

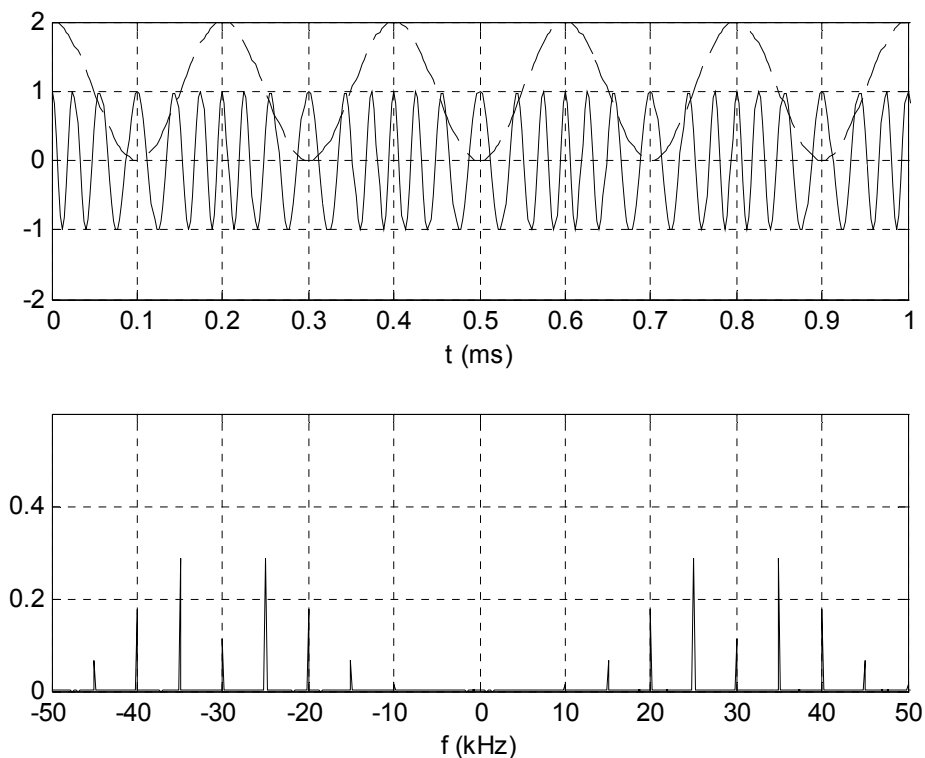


Рис 4.15 Частотная модуляция (FM)
 $(\omega_0 = 2\pi \cdot 30 \text{ кГц}, \Omega = 2\pi \cdot 5 \text{ кГц}, \beta = 2)$

Формирование реальных сигналов, как правило, выполняется при $\beta \gg 1$, при этом эффективная ширина спектра равна удвоенной девиации частоты:

$$\Pi_{\text{практ}} \approx 2\beta\Omega = 2\omega_d.$$

Отсюда следует, что по сравнению с АМ-сигналами, полоса частот которых равна 2Ω , для передачи сигналов с угловой модуляцией требуется полоса частот, в β раз большая. С другой стороны, именно это свойство ЧМ и ФМ-сигналов обеспечивает их большую помехоустойчивость по сравнению с АМ сигналами.

Сигналы с многотональной угловой модуляцией отличаются еще большей сложностью спектрального состава. В их спектре присутствуют не только боковые частоты с гармониками частот модулирующего сигнала, но и боковые комбинационные частоты типа $\omega_0 \pm \Omega_1 \pm \Omega_2 \pm \dots \pm \Omega_i$, со всеми возможными комбинациями частот модулирующего сигнала Ω_i . При непрерывном спектре

модулирующего сигнала спектры ЧМ и ФМ-сигналов также становятся непрерывными.

8. Квадратурная модуляция

Квадратурная модуляция позволяет модулировать несущую частоту одновременно двумя сигналами – путем модуляции амплитуды несущей одним сигналом и фазы несущей другим сигналом (рис. 4.16). Уравнение результирующих колебаний амплитудно-фазовой модуляции:

$$s(t) = u(t) \cos(\omega_0 t + \varphi(t)).$$

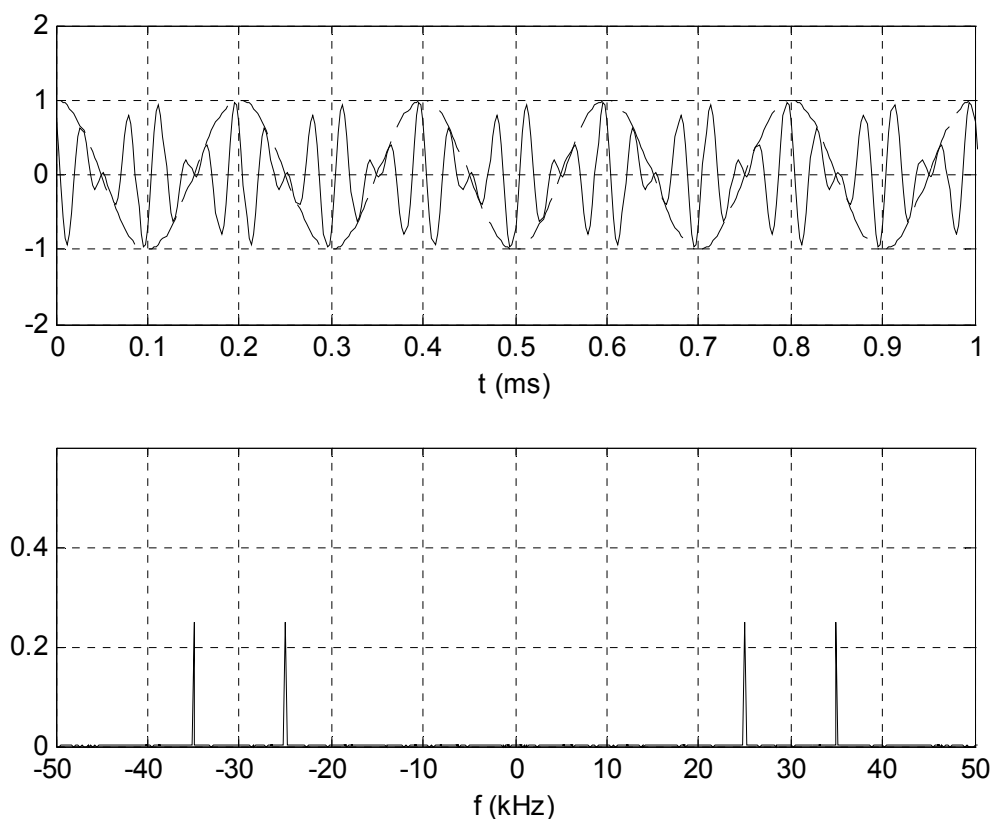


Рис 4.16. Квадратурная амплитудная модуляция (QAM)
 $(\omega_0 = 2\pi \cdot 30 \text{ кГц}, \Omega = 2\pi \cdot 5 \text{ кГц})$

Сигнал $s(t)$ обычно формируют в несколько другой последовательности, с учетом последующей демодуляции. Раскроем косинус суммы и представим сигнал в виде суммы двух АМ-колебаний.

$$s(t) = u(t) \cos(\omega_0 t) \cos \varphi(t) - u(t) \sin(\omega_0 t) \sin \varphi(t).$$

При $a(t) = u(t) \cos \varphi(t)$ и $b(t) = -u(t) \sin \varphi(t)$, сигналы $a(t)$ и $b(t)$ могут быть использованы в качестве модулирующих сигналов несущих колебаний $\cos(\omega_0 t)$ и $\sin(\omega_0 t)$, сдвинутых по фазе на 90° относительно друг друга:

$$s(t) = a(t) \cos(\omega_0 t) - b(t) \sin(\omega_0 t).$$

Полученный сигнал называется квадратурным, а способ модуляции – *квадратурной* модуляцией (КАМ).

Спектр квадратурного сигнала может быть получен непосредственно по уравнению балансной модуляции для суммы двух сигналов:

$$S(\omega) = \frac{1}{2} A(\omega + \omega_0) + \frac{1}{2} A(\omega - \omega_0) - \frac{1}{2} jB(\omega + \omega_0) + \frac{1}{2} jB(\omega - \omega_0).$$

Демодуляция квадратурного сигнала соответственно выполняется умножением на два опорных колебания, имеющих сдвиг фаз относительно друг друга на 90° :

$$s_1(t) = s(t) \cos \omega_0 t = \frac{1}{2} a(t) + \frac{1}{2} a(t) \cos 2\omega_0 t + \frac{1}{2} b(t) \sin 2\omega_0 t,$$

$$s_2(t) = s(t) \sin \omega_0 t = \frac{1}{2} b(t) + \frac{1}{2} a(t) \sin 2\omega_0 t - \frac{1}{2} b(t) \cos 2\omega_0 t.$$

Низкочастотные составляющие $a(t)$ и $b(t)$ выделяются фильтром низких частот. Как и при балансной амплитудной модуляции, для точной демодуляции сигналов требуется точное соблюдение частоты и начальной фазы опорного колебания.

9. Линейная частотная модуляция

Сигнал с линейной частотной модуляцией – это импульс, высокочастотное заполнение которого имеет линейный характер. ЛЧМ-сигналы имеют замечательное свойство. Если сигнал подать на частотно-зависимую линию задержки, время задержки сигнала которой велико на малых частотах (в начальной части ЛЧМ-сигнала) и уменьшается по мере нарастания частоты в ЛЧМ-сигнале, то на выходе такой линии происходит "сжатие" сигнала в один период высокочастотного колебания путем суммирования амплитудных значений всех периодов сигнала. При этом происходит увеличение

амплитуды выходного сигнала и уменьшение статистических шумов, так как суммируемые одновременно по этим же периодам шумы не коррелированы [5].

Пусть длительность импульса равна τ_u , точка $t = 0$ соответствует центру импульса. Пусть частота заполнения линейно нарастает от начала импульса к его концу с ускорением $\mu (\text{с}^{-2})$, тогда девиация частоты за время длительности импульса и полная фаза сигнала:

$$\Delta\omega = \mu \cdot \tau_u,$$

$$\psi(t) = \omega_0 t + \mu \cdot t^2 / 2.$$

Уравнение ЛЧМ-сигнала:

$$u(t) = \begin{cases} 0, & t < -\tau_u / 2 \\ U_m \cos(\omega_0 t + \mu t^2 / 2), & -\tau_u / 2 \leq t \leq \tau_u / 2. \\ 0, & t > \tau_u / 2 \end{cases}$$

4.2.2. Методы демодуляции

1. Демодуляция АМ-сигналов

Демодуляция АМ-сигналов может выполняться несколькими способами. Рассмотрим двухполупериодное детектирование (вычисление модуля сигнала) с последующим сглаживанием полученных однополярных полупериодов несущей фильтром низких частот.

При детектировании спектр модулированного сигнала становится однополярным, переходит на основную несущую частоту $2\omega_0$ и уменьшается по энергии почти в пять раз. Основная часть энергии переходит в область низких частот и распределяется между постоянной составляющей и выделенной гармоникой сигнала модуляции. Между постоянной составляющей и выделенной гармоникой энергия распределяется в зависимости от значения коэффициента модуляции M . При $M = 1$ энергии равны, при $M = 0$ (в отсутствие сигнала модуляции) вся энергия переходит на постоянную составляющую.

Кроме этих составляющих в спектре появляются также 2-я, 3-я и более высокие гармоники детектированного модулированного сигнала. Демодуляторы сигнала выделяют после детектирования только низкочастотный информационный сигнал и подавляют все остальные частоты, включая постоянную составляющую.

Очевидно также, что в случае перемодуляции сигнала исходный информационный сигнал будет восстанавливаться с ошибкой.

Другой распространенный метод – *синхронное детектирование*. При синхронном детектировании модулированный сигнал умножается на опорное колебание с частотой несущего колебания. Без учета начальных фаз колебаний:

$$\begin{aligned} y(t) &= u(t) \cos \omega_0 t = U(t) \cos \omega_0 t \cdot \cos \omega_0 t = \\ &= \frac{1}{2} U(t) + \frac{1}{2} U(t) \cos 2\omega_0 t. \end{aligned} \quad (4.27)$$

Как следует из (4.27), сигнал разделяется на два слагаемых, первое из которых повторяет исходный модулирующий сигнал, а второе повторяет модулированный сигнал на удвоенной несущей частоте $2\omega_0$. Форма новой несущей при синхронном детектировании является чистой гармоникой, в отличие от двухполупериодного детектирования, где новая несущая содержит дополнительные гармоники более высоких частот. Физический амплитудный спектр сигналов после демодуляции подобен спектру двухполупериодного детектирования, но однозначно соотносится со спектром входного модулированного сигнала: амплитуды гармоник модулированного сигнала на частоте $2\omega_0$ в два раза меньше амплитуд входного сигнала, постоянная составляющая равна амплитуде несущей частоты ω_0 и не зависит от глубины модуляции, амплитуда информационного демодулированного сигнала в два раза меньше амплитуды исходного модулирующего сигнала. Особенностью синхронного детектирования является независимость от глубины модуляции, т.е. коэффициент модуляции сигнала может быть больше единицы.

Однако при синхронном детектировании требуется точное совпадение фаз и частот опорного колебания демодулятора и несущей гармоники АМ-сигнала.

При сдвиге фазы опорного колебания на $\Delta\omega$ относительно несущей частоты выходной сигнал демодулятора оказывается умноженным на косинус фазовой ошибки:

$$y(t) = U(t) \cos \omega_0 t \cdot \cos(\omega_0 t - \Delta\varphi) = \frac{1}{2} U(t) \cos(-\Delta\varphi) + \\ + \frac{1}{2} U(t) \cos(2\omega_0 t - \Delta\varphi)$$

и амплитуда сигнала уменьшается, а при $\Delta\omega = \pi/2$ становится равной нулю.

При сдвиге частоты между несущим и опорным колебаниями сигнал демодулятора оказывается умноженным на гармоническое колебание с разностной частотой:

$$y(t) = U(t) \cos \omega_0 t \cdot \cos(\omega_0 t - \Delta\omega) = \frac{1}{2} U(t) \cos(-\Delta\omega \cdot t) + \\ + \frac{1}{2} U(t) \cos((2\omega_0 - \Delta\omega)t),$$

при этом выходной сигнал демодулятора начинает пульсировать (beat – биения) с частотой биений $\Delta\omega$.

Для частотной и фазовой синхронизации между несущим и опорным колебаниями в составе демодуляторов обычно используются следящие системы фазовой автоподстройки опорной частоты (см. раздел 4.2.3).

2. Демодуляция сигналов с угловой модуляцией

Методы демодуляции сигналов с угловой модуляцией сложнее демодуляции сигналов АМ.

При демодуляции цифровых сигналов обычно используется метод формирования комплексного аналитического сигнала с помощью преобразования Гильберта:

$$u_a(t) = u(t) + ju_h(t),$$

где $u_h(t)$ – аналитически сопряженный сигнал или квадратурное дополнение сигнала $u(t)$, которое вычисляется сверткой сигнала $u(t)$ с оператором Гильберта $(\frac{1}{\pi} \cdot t)$ — фазовращателем:

$$u_h(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{u(t') dt'}{(t - t')}.$$

Полная фаза колебаний представляет собой аргумент аналитического сигнала:

$$\psi(t) = \arg(u_a(t)).$$

Дальнейшие операции определяются видом угловой модуляции. При демодуляции ФМ сигналов из фазовой функции вычитается значение немодулированной несущей $\omega_0 t$:

$$\varphi(t) = \psi(t) - \omega_0 t.$$

При частотной модуляции фазовая функция дифференцируется с вычитанием из результата значения частоты ω_0 :

$$\varphi(t) = \frac{d}{dt} \psi(t) - \omega_0.$$

Метод может применяться и в реальном масштабе времени, но с определенной степенью приближения, поскольку оператор Гильберта слабо затухает.

Обычно в реальном масштабе времени входной сигнал умножается на два опорных колебания со сдвигом фазы между колебаниями в 90° :

$$\begin{aligned} u_1(t) &= u(t) \cos \omega_0 t = U_m \cos(\omega_0 t + \varphi(t)) \cos \omega_0 t = \\ &= \frac{1}{2} U_m \cos \varphi(t) + \frac{1}{2} \cos(2\omega_0 t + \varphi(t)), \\ u_2(t) &= u(t) \sin \omega_0 t = U_m \cos(\omega_0 t + \varphi(t)) \sin \omega_0 t = \\ &= \frac{1}{2} U_m \sin \varphi(t) + \frac{1}{2} \sin(2\omega_0 t + \varphi(t)). \end{aligned}$$

Из этих двух сигналов фильтрами низких частот выделяются низкочастотные колебания, и формируется аналитический сигнал:

$$u_a(t) = \frac{1}{2} U_m \cos \varphi(t) - \frac{1}{2} j U_m \sin \varphi(t).$$

Аргумент этого аналитического сигнала, как и в первом случае, представляет полную фазу колебаний, обработка которой выполняется аналогично.

4.2.3. Фазовая автоподстройка частоты

Схемы фазовой автоподстройки частоты (ФАПЧ) используется в передатчиках и приемниках, при аналоговой и цифровой модуляции, а также при передаче цифровых сигналов. Широко применяются схемы ФАПЧ при фазовой и частотной демодуляции. Также схема ФАПЧ позволяет обеспечить точную настройку, частотную селекцию и фильтрацию без использования громоздких элементов, обычно используемых в схемах детектирования. ФАПЧ представляет собой систему управления с петлей обратной связи, в которой параметрами регулирования являются частота или фаза сигнала. Структурная схема петли ФАПЧ (рис. 4.17) содержит три основных блока:

1. Фазовый компаратор или фазовый детектор (ФД).
2. Петлевой фильтр, фильтр низких частот (ФНЧ).
3. Генератор, управляемый напряжением (ГУН).

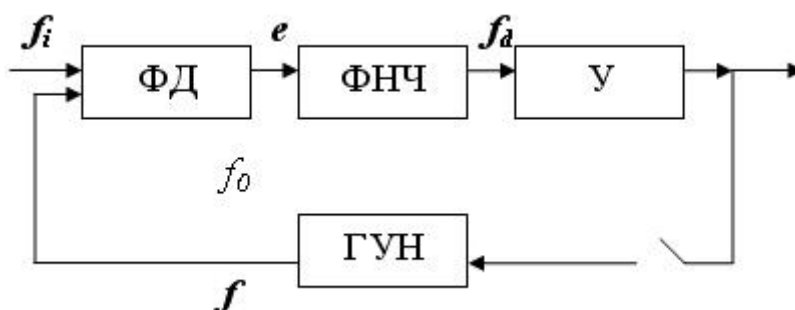


Рис. 4.17. Структурная схема ФАПЧ

В основе работы системы ФАПЧ лежит фазовая синхронизация сигналов подстраиваемого и опорного генераторов.

При отсутствии внешнего опорного сигнала или когда петля обратной связи разомкнута, ГУН работает на некоторой начальной частоте (частоте предварительной настройки) f_0 , которую также называют собственной частотой. Собственная частота является настраиваемым параметром системы ФАПЧ. Перед тем, как система

ФАПЧ войдет в режим слежения, должен произойти захват частоты. Когда на вход системы ФАПЧ поступает внешний входной сигнал с частотой f_i , компаратор сравнивает частоту внешнего сигнала с частотой выходного сигнала ГУН f_0 .

На выходе фазового компаратора формируется напряжение ошибки e , зависящее от рассогласования частот Δf и от разности фаз двух сигналов на входе. При реализации ФД наиболее часто применяются следующие четыре варианта:

1. Flip-Flop детектор (использует RS-триггер, переключаемый положительными переходами через ноль входных сигналов).

2. Детектор, использующий теорему отсчетов Шеннона-Котельникова-Найквиста (берутся отсчеты одного входного сигнала с частотой Найквиста и перемножаются со вторым сигналом).

3. Детектор перехода через ноль (использует моменты перехода через ноль).

4. Детектор опережения и запаздывания (представляет разницу фаз входного и опорного сигналов в виде однобитной последовательности).

Напряжение сигнала ошибки через фильтр нижних частот и усилитель воздействует на управляющий элемент ГУН. Если собственная частота ГУН f_0 достаточно близка к частоте внешнего опорного сигнала f_i , то под действием обратной связи в схеме ФАПЧ ГУН синхронизируется, то есть захватывает внешний входной сигнал. Поэтому выходная частота ГУН – это сумма или разность его собственной частоты и разницы между внешней опорной частотой и собственной частотой ГУН. Выходная частота ГУН f определяется выражением:

$$f = f_0 \pm \Delta f,$$

где f – выходная частота ГУН, Гц; f_0 — собственная частота ГУН, Гц; $\Delta f = f_0 - f_i$ Гц; f_i — внешняя опорная частота, Гц.

Таким образом, система ФАПЧ имеет три рабочих состояния:

1. автономное состояние при разомкнутой петле обратной связи;

2. режим захвата;
3. режим слежения.

При подаче на вход схемы сигнала с угловой модуляцией на выходе ФНЧ после захвата частоты и фазы формируется сигнал, пропорциональный модулирующему сигналу, что позволяет применять схему ФАПЧ для демодуляции.

В автономном состоянии внешний сигнал опорной частоты отсутствует или петля обратной связи разомкнута. При этом ГУН генерирует сигнал на своей собственной частоте. В состоянии захвата, система ФАПЧ должна иметь внешний опорный сигнал и замкнутую петлю обратной связи. Из состояния захвата система ФАПЧ со временем переходит в режим слежения частоты. В режиме слежения частота ГУН отслеживает частоту внешнего входного сигнала (т.е. равна ей). Когда система ФАПЧ находится в состоянии удержания, частота ГУН следует за изменениями частоты внешнего опорного сигнала.

Рассмотрим работу схемы подробнее.

Когда внешний сигнал $V_i \sin(2\pi f_i t + \theta_i)$ поступает на фазовый компаратор, он сравнивается с выходным сигналом ГУН $V_0 \sin(2\pi f_0 t + \theta_0)$. Сначала частоты этих двух сигналов не равны ($f_0 \neq f_i$). Сравнение выполняется путем перемножения сигналов в ФД, где внешний сигнал и сигнал с выхода ГУН создают комбинационные частоты (т.е. суммарные и разностные частоты). Поэтому на выходе фазового компаратора присутствуют: частота с выхода ГУН f , их сумма $(f + f_i)$ и разность $(f - f_i)$ (см. также раздел 4.2.1).

Фильтр нижних частот подавляет входные частоты f_i и f_0 , а также их суммарную частоту $f_0 + f_i$. Таким образом, на выходе ФНЧ присутствует разностная частота $f_d = f_0 - f_i$. Сигнал биений усиливается, а затем поступает на вход ГУН, изменяя выходную частоту ГУН пропорционально величине и полярности своего напряжения. Передаточная функция ГУН определяется выражением:

$$K_0 = \frac{\Delta f}{\Delta V}$$

где K_0 – передаточная функция, Гц/В; ΔV — изменение входного управляющего напряжения, В; Δf — изменение выходной частоты, Гц.

Поскольку выходная частота ГУН изменяется, амплитуда и частота разностного сигнала также пропорционально изменяется. После определенного числа циклов подстройки выходная частота ГУН становится равной внешней опорной частоте и происходит захват частоты системой ФАПЧ. Как только захват частоты произошел, разностная частота на выходе ФНЧ становится равной 0 Гц (напряжение постоянного тока), а его величина и полярность будут пропорциональны разности фаз между внешним опорным сигналом и выходным сигналом ГУН. Напряжение постоянного тока обеспечивает входное смещение для ГУН, поддерживая режим слежения частоты внешнего сигнала.

Пока захват частоты не произошел, фазовый компаратор работает как частотный компаратор, и только после захвата он становится фазовым компаратором. После захвата частоты величина сдвига фаз между внешним сигналом и выходной частотой ГУН конвертируется в фазовом компараторе в постоянное напряжение смещения u_d , фильтруется, усиливается, а затем поступает обратно в ГУН, где поддерживает процесс слежения частоты (режим синхронизации). Следовательно, для поддержания этого режима необходимо поддерживать сдвиг фаз между внешним опорным сигналом и выходным сигналом ГУН. Время, требуемое для достижения режима удержания, называют временем захвата или временем вхождения в синхронизацию.

Полоса захвата и полоса удержания, отслеживания — это два ключевых параметра системы ФАПЧ, которые характеризуют её рабочий диапазон частот.

Область захвата определяется как диапазон частот, сосредоточенный вокруг собственной частоты ГУН f_n , в котором система ФАПЧ в состоянии произвести захват частоты и синхронизироваться с внешним входным сигналом. В зависимости от конструкции системы ФАПЧ, ширины полосы пропускания фильтра нижних частот и коэффициента усиления петли обратной связи, область захвата находится, как правило, в диапазоне между $0,5 f_0$ и $1,7 f_0$.

Область удержания определяется как диапазон частот, сосредоточенный вокруг собственной частоты ГУН, в котором система области захвата и удержания прямо пропорциональны коэффициенту усиления по постоянному току петли обратной связи системы ФАПЧ. Область захвата никогда не бывает больше и почти всегда меньше, чем область удержания.

4.2.4. Модуляция (манипуляция) цифровых сигналов

Числа при передаче информации в цифровой форме с периодом T поступают от источника информации и называются *символами* (symbol), а частота передачи символов – *символьной скоростью* (symbol rate) $f_T = 1/T$. В практике передачи данных распространена двоичная (binary) последовательность символов, где числа передаются значениями 0 и 1.

Каждому из возможных символов устанавливается определенный набор параметров несущего колебания, которые поддерживаются постоянными на интервале T до прихода следующего символа. Это означает преобразование последовательности чисел в ступенчатый сигнал, который используется в качестве модулирующего сигнала. Соответственно, параметры несущего колебания, на которые переносится сигнал, меняются скачкообразно. Такой способ модуляции несущей обычно называется *манипуляцией* (keying).

1. Амплитудная манипуляция

Амплитудно-манипулированные сигналы простейшего типа представляют собой последовательности импульсов, разделенные паузами. Форма огибающей радиоимпульсов в общем случае может быть произвольной, паузы могут отличаться по длительности от радиоимпульсов.

На рис. 4.18 приведен пример амплитудно-манипулированного сигнала: $u(t) = U_m \cos(2\pi f_0 t)$, с прямоугольной П-формой огибающей.

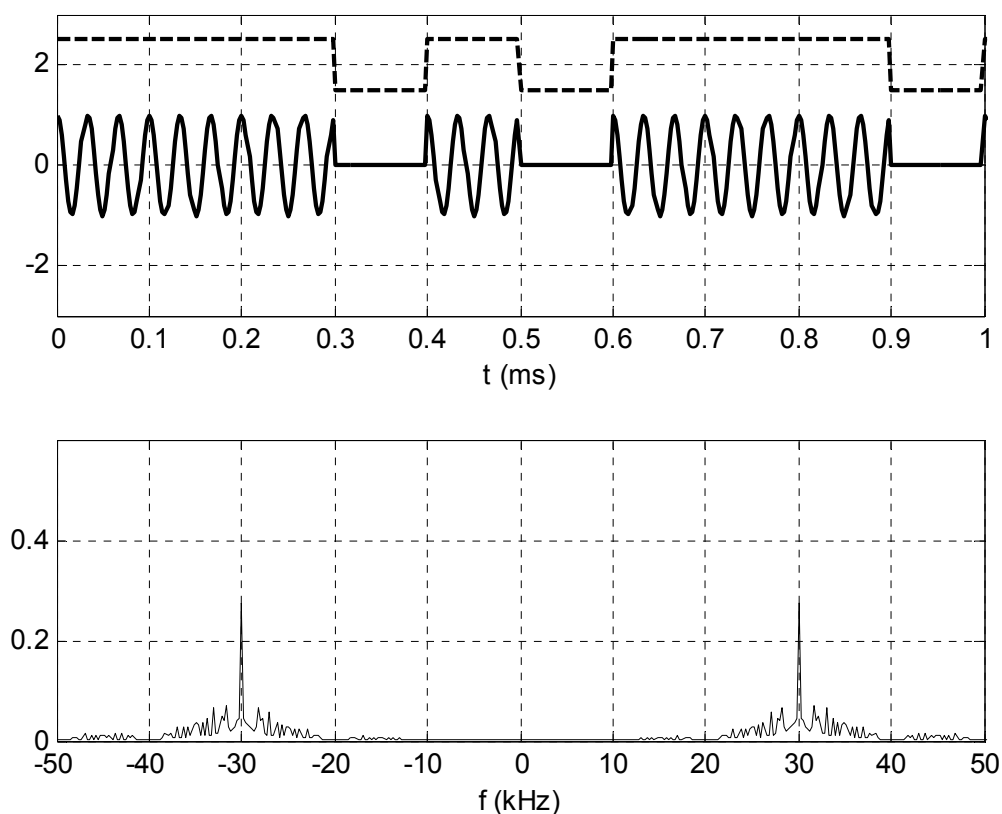


Рис 4.18. Цифровая амплитудная манипуляция (АМП)
($T_d = 0,1$ мс, $\omega_c = 2\pi \cdot 30$ кГц)

Спектр АМП-сигнала образуется сверткой спектра огибающей функции (в данном случае – спектра прямоугольного импульса) со спектром несущей (дельта-функции на частоте f_0). Модуль спектральной плотности сигнала приведен на рис. 4.18 внизу. Спектр прямоугольного импульса слабо затухает и бесконечен в частотной области, поэтому его использование в качестве огибающей АМП-

сигнала не рекомендуется, т.к. при передаче данных частотный диапазон канала передачи данных ограничивается значимой частью спектра, ширина которого устанавливается по допустимой степени искажения приемных сигналов.

2. Угловая манипуляция

Угловая манипуляция использует частотные методы, в которых каждому возможному значению передаваемого символа сопоставляется значение частоты гармонической несущей. При этом в точках сопряжения интервалов посылок могут происходить скачки напряжения, с соответствующим усложнением спектра модулированного сигнала. Для снижения влияния высокочастотных эффектов на границах символов можно использовать способ модуляции с кратным количеством периодов несущей в символьном интервале — минимальная частотная манипуляция (рис. 4.19). При более сложных способах, независимых от точного сопряжения несущих частот с интервалами посылок, осуществляется управление скоростью изменения фазы несущих на границах посылок (например, по гауссову закону – гауссова минимальная частотная манипуляция).

Демодуляция сигналов осуществляется корреляционными методами, т.е. путем вычисления взаимной корреляции между принимаемым сигналом и набором опорных частот, используемых при модулировании, с идентификацией символов по максимумам взаимной корреляции или с помощью цифровых схем ФАПЧ (см. также раздел 2.4.3).

Для повышения помехоустойчивости передачи данных желательно, чтобы разносимвольные посылки были некоррелированы. Если для бинарных символов 0 и 1 принять частоты посылок равными $s_0(t) = \cos \omega_0(t)$ $s_1(t) = \cos \omega_1(t)$, то их взаимная корреляционная функция при нулевом временном сдвиге определяется как:

$$B_{01} = \int_0^T s_0(t)s_1(t)dt = \frac{1}{2} \frac{\sin(\omega_1 + \omega_0)T}{\omega_1 + \omega_0} + \frac{1}{2} \frac{\sin(\omega_1 - \omega_0)T}{\omega_1 - \omega_0}.$$

При $(\omega_1 + \omega_2)T \gg 1$ первым слагаемым можно пренебречь, второе слагаемое обращается в нуль при $(\omega_1 + \omega_2)T = \pi k$, где $k = 1, 2, \dots$ – целое число. Отсюда, минимальное значение между частотами манипуляции для некоррелированных посылок определяется выражениями:

$$\Delta\omega_{\min} = \pi/T \quad \Delta f_{\min} = 1/2 T = f_T/2,$$

где f_T – символьная скорость.

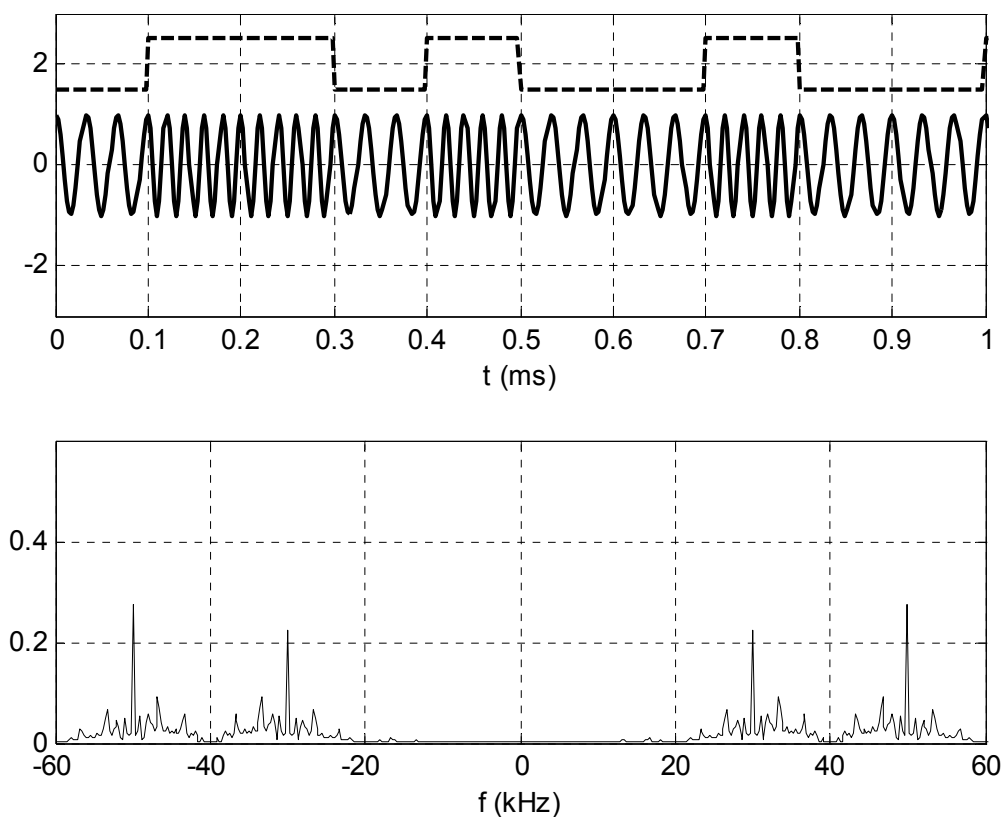


Рис. 4.19. Цифровая частотная манипуляция
($T_d = 0,1$ мс, $\omega_{c1} = 2\pi \cdot 30$ кГц, $\omega_{c2} = 2\pi \cdot 50$ кГц)

Фазовая манипуляция применяется реже (рис. 4.20), в связи со значительными сложностями измерения абсолютных значений начальных фаз в посылках.

Значительно легче детектировать относительный фазовый сдвиг в соседних посылках, поэтому обычно используется фазоразностная манипуляция.

3. Квадратурная амплитудная модуляция

При квадратурной амплитудной модуляции (КАМ) изменяется как фаза, так и амплитуда несущего сигнала. Это позволяет увеличить количество кодируемых в единицу времени бит и при этом повысить помехоустойчивость их передачи по каналу связи. В настоящее время число кодируемых информационных бит на одном интервале может достигать 8-9, а число состояний сигнала в сигнальном пространстве, соответственно – 256...512.

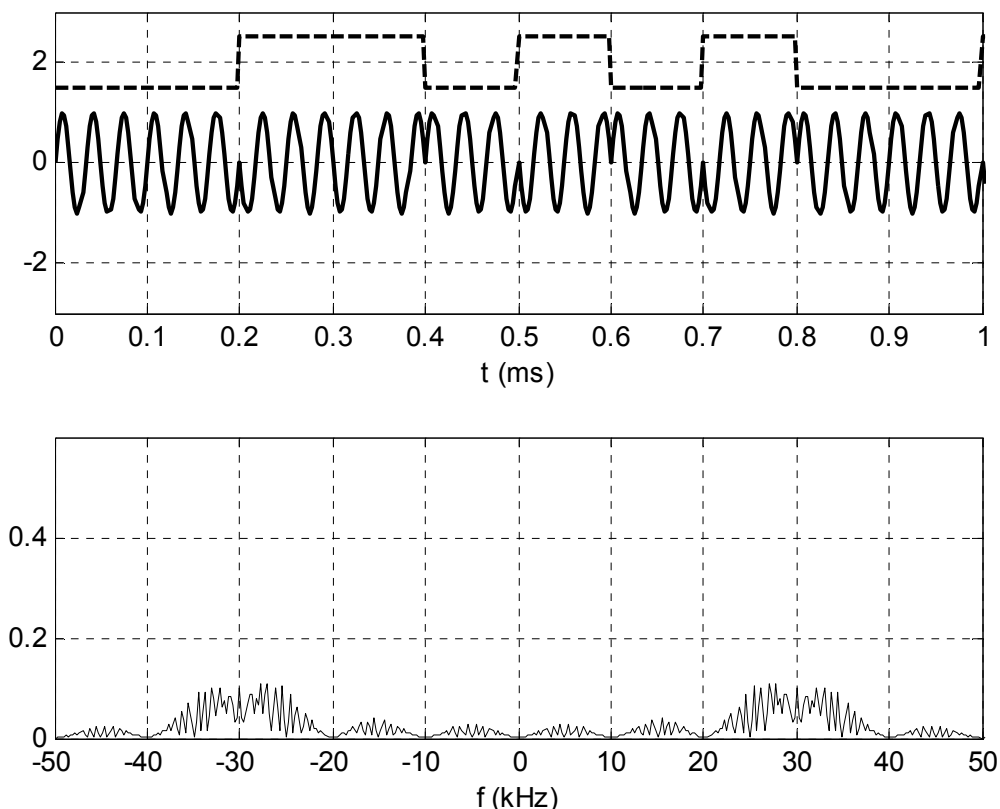


Рис. 4.20. Цифровая бифазная манипуляция
($T_d = 0,1$ мс, $\omega_c = 2\pi \cdot 3$ кГц)

Квадратурное представление сигнала заключается в выражении колебания линейной комбинацией двух ортогональных составляющих – квадратурной и синфазной:

$$S(t) = x(t) \sin(\omega \cdot t + \varphi) + y(t) \cos(\omega \cdot t + \varphi),$$

где $x(t)$ и $y(t)$ – биполярные дискретные сигналы.

Рассмотрим работу квадратурного модулятора на примере схемы формирования сигналов четырехфазной ФМ из битового потока. Исходная последовательность двоичных символов при помощи регистра сдвига разделяется на нечетные y и четные x импульсы, которые поступают на входы формирователей манипулирующих импульсов (ФМИ) квадратурного и синфазного каналов. На выходах ФМИ образуются последовательности биполярных импульсов $x(t)$ и $y(t)$ с амплитудой $\pm U_m$ и длительностью $2T$, которые поступают на входы канальных перемножителей, где они независимо друг от друга модулируют по амплитуде два одинаковых несущих колебания, имеющих сдвиг фаз 90° , т.е. находящихся в квадратуре. В результате, на их выходах формируются двухфазные (0, π) колебания.

После суммирования они образуют сигнал ФМ-4 или квадратурный ФМ-сигнал (Quadrature Phase Shift Keying – QPSK). Поскольку в каждом канале осуществляется амплитудная манипуляция, этот вид модуляции называют еще квадратурной амплитудной манипуляцией (QASK – Quadrature Amplitude Shift Keying) или просто квадратурной амплитудной модуляцией (КАМ). При одновременной смене символов в обоих каналах модулятора (с 10 на 01, или с 00 на 11) в сигнале ФМ-4 происходит скачок фазы на 180° (π). Такие скачки фазы вызывают появление высокочастотных составляющих в спектре модулированного сигнала. В результате этого при прохождении сигнала через узкополосный фильтр возникают провалы огибающей несущего колебания до нуля. Такие изменения сигнала нежелательны, поскольку приводят к увеличению энергии боковых полос и помех в канале связи.

Четырехфазная ФМ со сдвигом (OQPSK – Offset QPSK) позволяет избежать скачков фазы на 180° и, следовательно, глубокой модуляции огибающей. Формирование сигнала в модуляторе OQPSK

происходит так же, как и в модуляторе ФМ-4, за исключением того, что манипуляционные элементы информационных последовательностей $x(t)$ и $y(t)$ смещены во времени на длительность одного элемента T (рис. 4.21, б, в). Изменение фазы при таком смещении модулирующих потоков определяется лишь одним элементом последовательности, а не двумя, как при ФМ-4. В результате скачки фазы на 180° отсутствуют, так как каждый элемент последовательности, поступающий на вход модулятора синфазного или квадратурного канала, может вызвать изменение фазы на 0 , $+90^\circ$ или -90° .

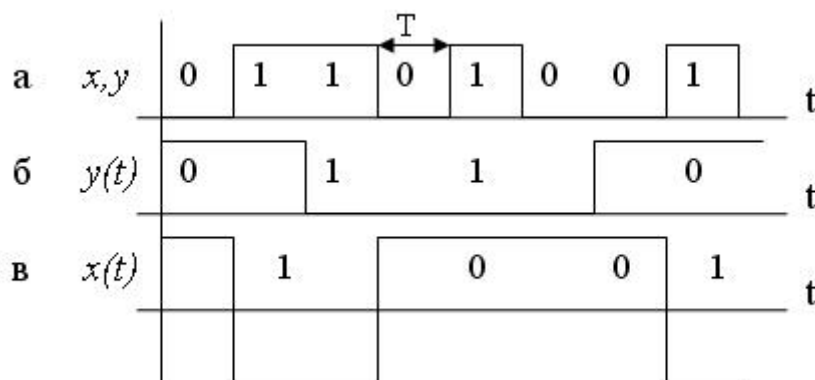


Рис. 4.21. Формирование манипулирующих сигналов

Преобразованные таким образом сигналы передаются в одном канале. Поскольку один и тот же физический канал используется для передачи двух сигналов, то скорость передачи КАМ-сигнала в отличие от АМ-сигнала в два раза выше.

Удобным способом визуализации сигналов модуляции можно считать их изображение векторами или точками в сигнальном пространстве. Совокупность сигнальных точек образует так называемое сигнальное созвездие (signal constellation).

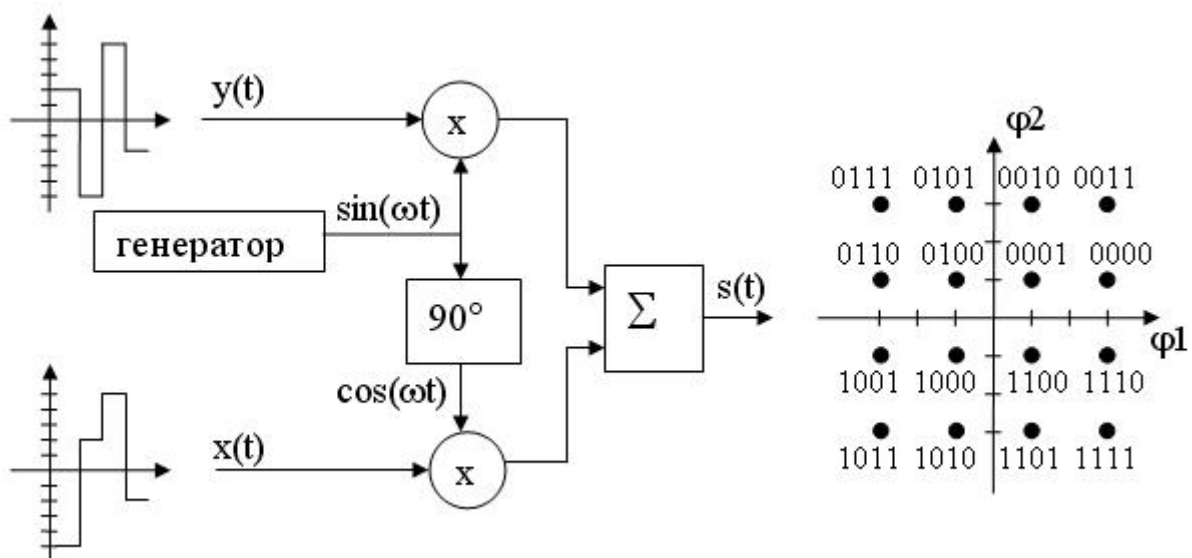


Рис. 4.22. Модуляция КАМ-16 и ее сигнальное созвездие

На рис. 4.22 показана структурная схема модулятора и диаграмма состояний (сигнальное созвездие) системы КАМ-16, в которой $x(t)$ и $y(t)$ принимают значения $\pm 1, \pm 3$ (4-х уровневая КАМ).

Существует несколько способов практической реализации 4-уровневой КАМ. Еще одна схема, реализующая данный этот вид модуляции, использует два одинаковых 4-фазных модулятора. Структурная схема такого модулятора КАМ-16 имеет вид (рис. 4.23).

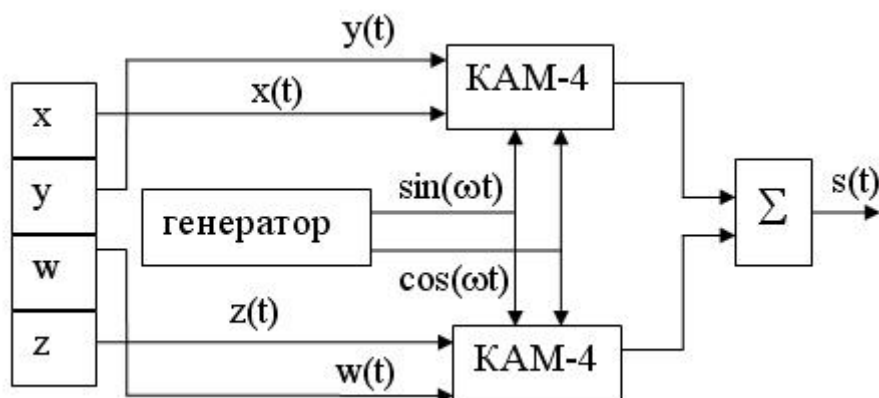


Рис. 4.23. Альтернативная схема модулятора КАМ-16

В общем случае, при формировании сигналов многопозиционной КАМ, модуляция ортогональных сигналов осуществляется в цифровом виде. Для этих целей используется два цифровых полосовых фильтра с одинаковой амплитудой входных

колебаний, но различающихся фазовым сдвигом в 90° . Уровни усиления амплитуды для каждого потока устанавливают независимо. Для системы, поддерживающей m амплитудных уровней для каждого потока, можно образовать m^2 различных комбинаций нуля и единицы.

При равном числе точек в сигнальном созвездии спектр сигналов КАМ идентичен спектру сигналов ФМ. Однако помехоустойчивость систем ФМ и КАМ различна. При одинаковом числе точек сигналы системы КАМ имеют лучшую помехозащищенность, чем сигналы системы ФМ. Основная причина этого состоит в том, что расстояние между сигнальными точками в системе ФМ меньше расстояния между сигнальными точками в системе КАМ. На рис. 4.24 представлены сигнальные созвездия систем КАМ-16 и ФМ-16 при одинаковой нормированной мощности сигнала.

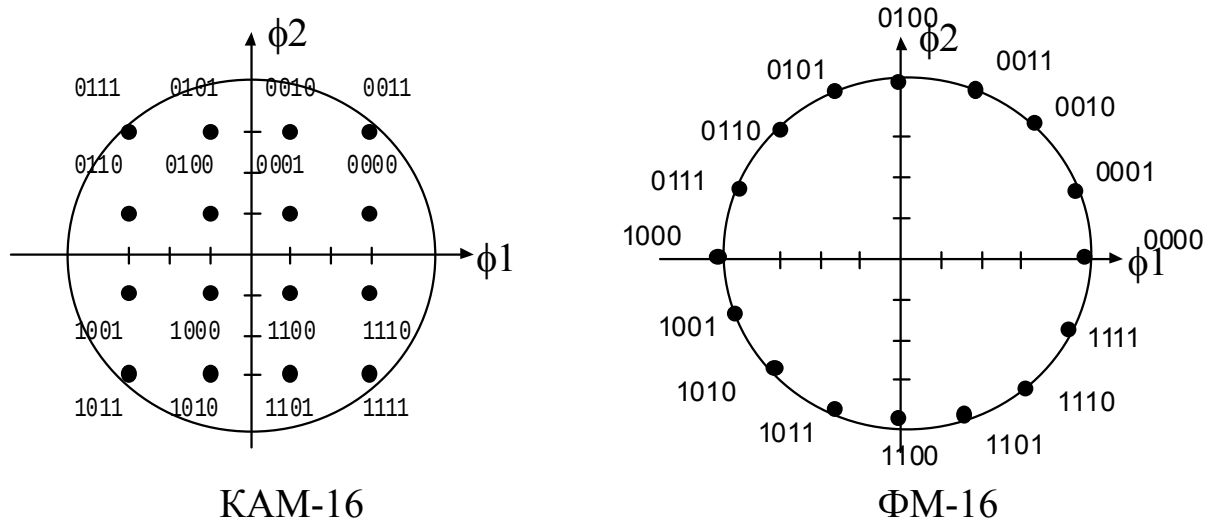


Рис. 4.24. Сигнальные созвездия КАМ-16 и ФМ-16

Расстояние между соседними точками сигнального созвездия в системе КАМ с L уровнями модуляции определяется выражением

$$d = \frac{\sqrt{2}}{L-1}.$$

Аналогично при ФМ: $d = 2 \sin \frac{\pi}{M}$, где M – число фаз.

4. Амплитудно-фазовая манипуляция с подавлением несущей

Амплитудно-фазовая модуляция с подавлением несущей (CAP – Carrier less Amplitude Phase modulation) является одним из наиболее широко используемых на цифровых абонентских или DSL (Digital Subscriber Line) линиях способов модуляции. CAP-модуляция представляет собой одну из разновидностей КАМ. Ее особенность заключается в специальной обработке модулированного информационного сигнала перед отправкой в линию. В процессе этой обработки из спектра модулированного сигнала исключается неинформативная составляющая, соответствующая сигналу несущей частоты (рис. 4.25).

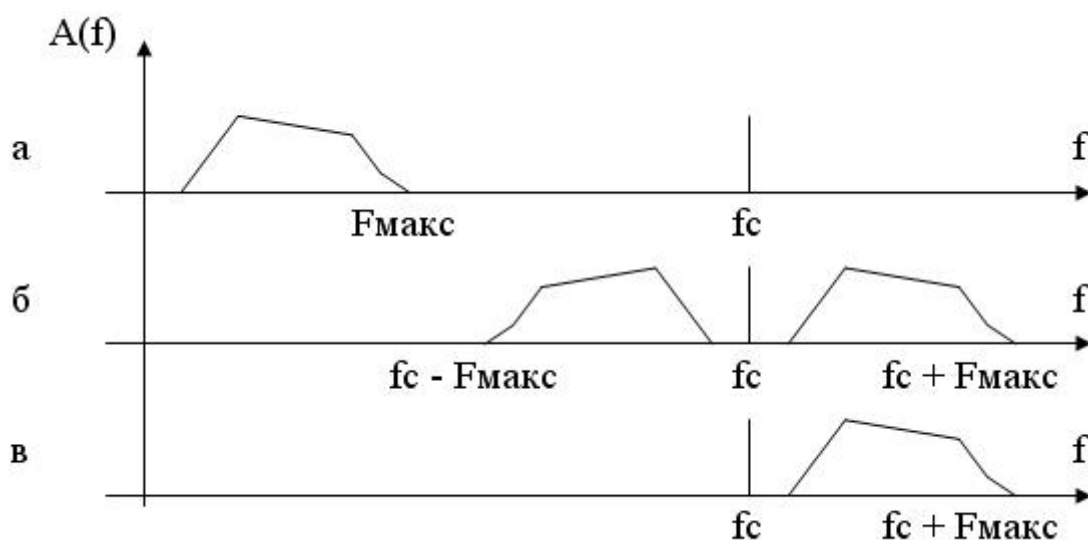


Рис. 4.25. Формирование спектра CAP-модулированного сигнала

Спектр информационного сигнала ограничен сверху частотой $F_{\text{макс}}$, частота модулируемого колебания (носителя) – f_c . После выполнения процедуры модуляции спектр полезного сигнала переносится в область частоты f_c с образованием двух зеркальных спектров сигнала относительно составляющей f_c . Для восстановления переданного сигнала на приемнике достаточно передать только одну из зеркальных компонент модулированного сигнала. Гармоника с частотой f_c не является информативной, и ее потеря не повлияет на

качество восстановленного сигнала. Схема формирования сигнала, модулированного по методу CAP, представлена на рис. 4.26.

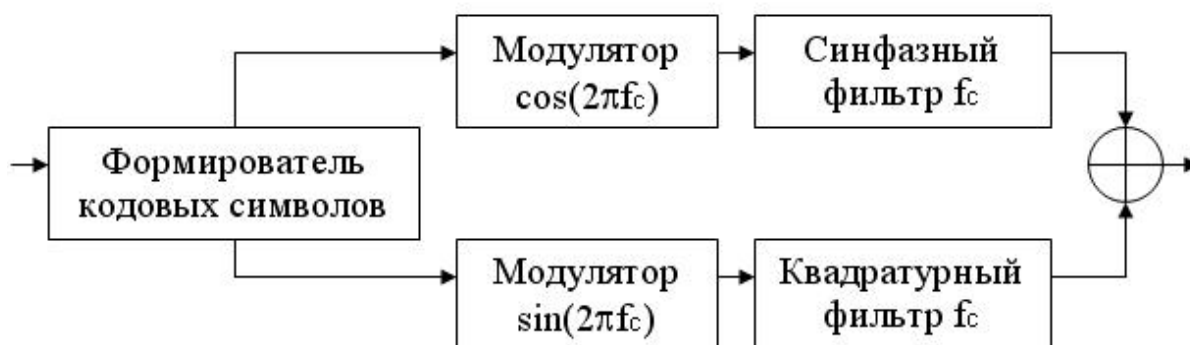


Рис. 4.26. Схема формирования CAP-модулированного сигнала

Для подавления несущего колебания используются синфазный и квадратурный фильтры. Для восстановления сформированного таким образом сигнала на приемной стороне должны быть выполнены соответствующие операции по восстановлению несущего колебания. После восстановления несущей приемник выполняет те же операции, что и приемник КАМ. Данный способ модуляции теоретически способен обеспечить максимальные значения отношения сигнал/шум.

Основным недостатком данного способа модуляции является отсутствие стандарта, определяющего процедуры, в соответствии с которыми выполняется преобразование сигнала. Одной из причин, которые приводят к сдерживанию внедрения этой технологии, является сильная поддержка альтернативной по отношению к CAP – технологии многочастотного способа модуляции.

5. Многочастотный способ модуляции

Многочастотный способ модуляции (DMT – Discrete Multi Tone) в настоящее время является одним из основных методов модуляции, используемых в наиболее перспективных технологиях xDSL – ADSL и VDSL. В DMT в отличие от CAP используется пакет несущих частот, количество которых определяется числом каналов в полосе частот, занимаемой спектром DMT-сигнала. Обычно в полосе частот, занимаемой сигналом DMT, размещается 256 частотных каналов.

Каждый из этих каналов имеет ширину 4,3 кГц и служит для организации независимой передачи данных (рис. 4.27).

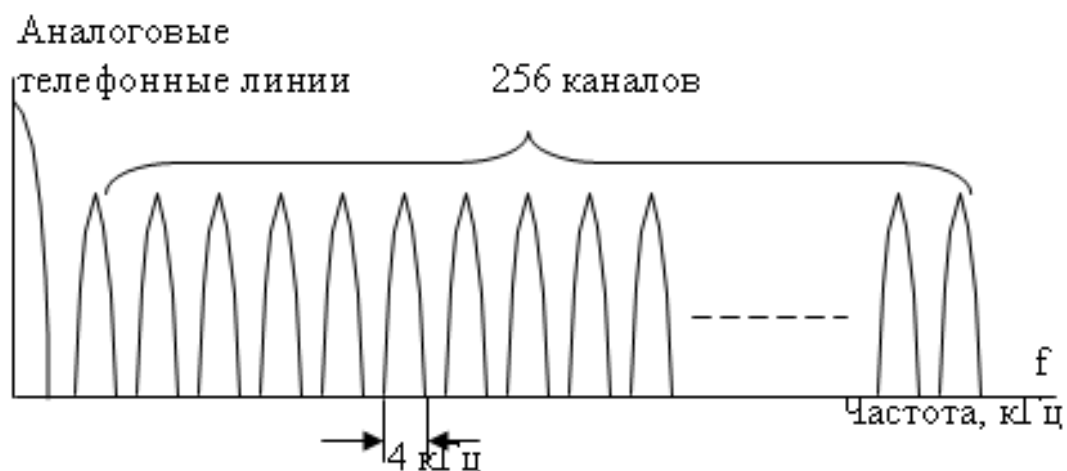


Рис. 4.27 Образование независимых каналов по методу DMT

На этапе вхождения в связь производится проверка качества линии. После чего, передатчик, исходя из уровня помех в частотном диапазоне сигнала DMT, для каждого из отдельных каналов выбирает подходящую схему модуляции. На участках с малым уровнем шумов могут быть использованы методы с большим числом уровней модуляции, такие как КАМ-64. На более зашумленных участках могут быть использованы простые алгоритмы модуляции, например, ФМ-4. Очевидно, что использование такого принципа регулирования скорости передачи данных позволяет наиболее точно согласовывать параметры модулированного сигнала с параметрами линии, по которой он будет передаваться. При передаче данных информация распределяется между независимыми каналами пропорционально их пропускной способности. Приемник выполняет операцию демультимплексирования и восстанавливает исходный информационный поток. Недостатками DMT-модуляции можно считать сложность, недостаточную технологичность и высокую стоимость ее аппаратной реализации. Однако отмеченные недостатки при постоянном развитии технологий являются все менее критичными. В подтверждение тому метод DMT утвержден в

качестве основного способа модуляции стандартов ANSI для технологий ADSL и VDSL.

6. Треллис-модуляция

Применение многопозиционной КАМ в чистом виде сопряжено с проблемой недостаточной помехоустойчивости. Поэтому во всех высокоскоростных протоколах КАМ используется совместно с решетчатым кодированием – специальным видом сверточного кодирования. В результате появился новый способ модуляции, называемый треллис-модуляцией (TCM – Trellis Coded Modulation). Выбранная определенным образом комбинация конкретной КАМ и помехоустойчивого кода обозначается термином "сигнально-кодовая конструкция" (СКК), которая позволяет повысить помехозащищенность передачи информации одновременно со снижением требований к отношению сигнал-шум на 3-6 дБ. Это объясняется увеличением вдвое числа сигнальных точек за счет добавления к информационным битам одного избыточного, образованного путем сверточного кодирования. Далее выполняется КАМ-модуляция или другая многопозиционная модуляция (ФМ-8, 8-FSK и т.д.). В процессе демодуляции принятый сигнал декодируется по алгоритму Витерби. Этот алгоритм за счет введенной избыточности и знания предыстории процесса приема позволяет по критерию максимального правдоподобия выбрать из сигнального пространства наиболее достоверную точку. Все применяемые сегодня СКК используют сверточное кодирование со скоростью $(n-1)/n$, т.е. при передаче одного сигнального элемента используется только один избыточный двоичный символ. Схема сверточного кодирования со скоростью, равной $2/3$, приведена на рис. 4.28.

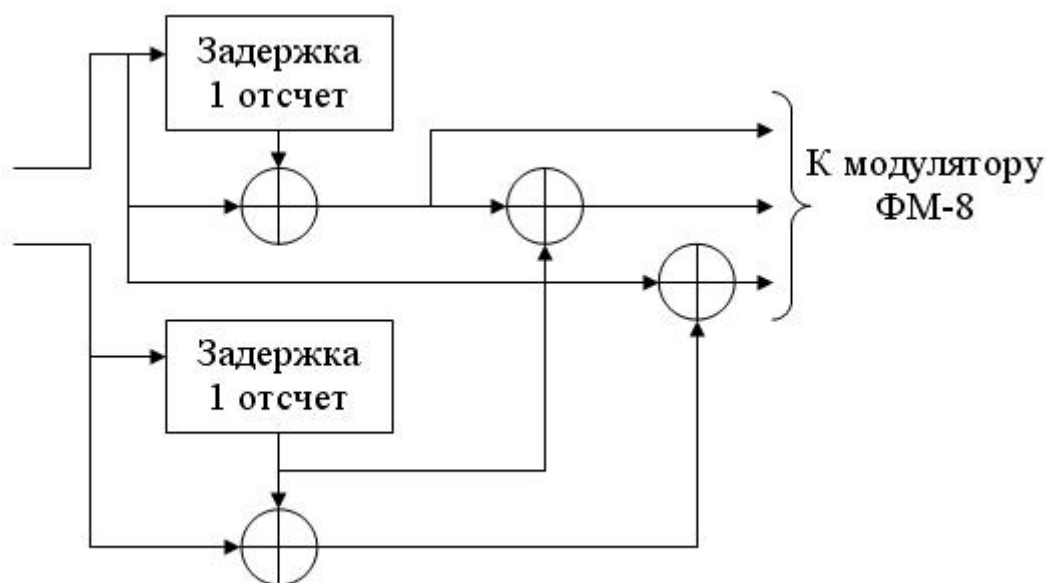


Рис. 4.28. Сверточный кодер с кодовым отношением 2/3

Каждым двум информационным битам на входе кодер сопоставляет трехсимвольные двоичные блоки на выходе, которые поступают на модулятор ФМ-8.

4.3. ТЕХНОЛОГИИ РАСШИРЕНИЯ СПЕКТРА

Основным свойством технологий расширения спектра для передачи информации является значительно более широкая полоса частот, чем это требуется при обычной узкополосной передаче. Разработано два принципиально различающихся между собой метода использования широкой полосы частот – *метод прямой последовательности* (Direct Sequence Spread Spectrum – DSSS) и *метод частотных скачков* (Frequency Hopping Spread Spectrum – FHSS).

Современное состояние беспроводной связи определяется стандартом IEEE 802.x. Разработкой и совершенствованием стандарта занимается рабочая группа по беспроводным локальным сетям (Working Group for Wireless Local Area Networks) комитета по стандартизации Института Инженеров Электротехники и Электроники (Institute of Electrical and Electronic Engineers, IEEE).

Стандарт определяет один тип протокола доступа к среде МАС-уровня и три различных протокола для физических (РНУ) каналов.

На физическом уровне стандарт допускает использование одного из двух типов радиоканалов и одного типа канала инфракрасного диапазона. Оба типа радиоканалов используют технологию расширения спектра, приводящую к уменьшению среднего значения спектральной плотности мощности сигнала благодаря распределению энергии в полосе частот, более широкой, чем необходимо для обеспечения заданной скорости передачи. Эта технология позволяет уменьшить уровень создаваемых помех и обеспечивает повышенную помехоустойчивость приема.

4.3.1. Метод частотных скачков

Предусмотрена скорость передачи 1 Мбит/с (опционально 2 Мбит/с). Версия 1 Мбит/с использует двухуровневую гауссову частотную модуляцию (2GFSK), а версия 2 Мбит/с — четырехуровневую (4GFSK). При GFSK к цифровому сигналу перед модуляцией применяется фильтр Гаусса:

$$y(n) = x(n) * g(n) = \sum_{n=1}^N x(N - n + 1)g(n) = \sum_{n=1}^N x(N - n + 1) \cdot \sqrt{\frac{a}{\pi}} (e^{-an^2})$$

где $g(n)$ – импульсная характеристика фильтра Гаусса.

При скорости 1 Мбит/с частота сигнала изменяется на длительности символа сообщения, равной 1 мкс, по гауссову закону от номинального значения до значения +170 кГц и возвращается к номинальному значению. Для передачи нуля частота сигнала изменяется на величину –170 кГц. Для скорости 2 Мбит/с предусмотрено четыре уровня отклонения частоты (+225, +75, –75, –225 кГц), поэтому каждая элементарная посылка (символ) переносит два бита сообщения. Ширина спектра сигнала при такой модуляции равна 1 МГц, независимо от скорости передачи. Это дает возможность использовать для передачи семидесяти девяти частотных позиций в диапазоне от 2402 до 2480 МГц с шагом 1 МГц.

Для расширения спектра частота сигнала изменяется по псевдослучайному закону не реже одного раза в 400 мс.

4.3.2. Метод прямой последовательности

Второй тип радиоканала — Direct Sequence Spread Spectrum (DSSS) Radio PHY. В этом варианте предусматривается передача со скоростями 1 и 2 Мбит/с. При скорости передачи 1 Мбит/с используется двоичная фазовая манипуляция — Binary Phase Shift Keying (BPSK). Единичный бит представляется 11-элементным кодом Баркера вида 11100010010, а нулевой бит — инверсным кодом Баркера. Элементарные символы кода Баркера не переносят информации, биты передаются сразу всем кодом Баркера — прямым или инверсным. Это позволяет придать сигналу свойства шума, обеспечивающие помехоустойчивость. Ширина спектра такого сигнала составляет 22 МГц. Для скорости 2 Мбит/с стандарт предусматривает квадратурную фазовую манипуляцию — QPSK. На длительности символа сообщения в этом случае передаются два бита. Для этого необходимо уже не два, а четыре различных сигнала. Поэтому вместе с основным несущим колебанием используется дополнительное, сдвинутое относительно него по фазе на 90° . Фаза каждого из этих колебаний управляется прямой или инверсной последовательностью Баркера, и оба колебания складываются. Таким образом, на длительности символа сигнал имеет четыре степени свободы, позволяющие передавать два бита. При этом скорость передачи увеличивается вдвое при сохранении той же полосы частот, что и при двоичной передаче. Для передачи сигнала DSSS используется одна из четырнадцати перекрывающихся частотных полос, определенных стандартом в общей полосе частот 83,5 МГц.

4.3.3. Сравнение методов частотных скачков и прямой последовательности

Метод частотных скачков, так же как и метод прямой последовательности, обеспечивает конфиденциальность и некоторую помехозащищенность передач. С другой стороны, поскольку при

использовании метода частотных скачков, в отличие от метода прямой последовательности, на каждом подканале передача ведется на достаточно большой мощности (сравнимой с мощностью обычных узкополосных передатчиков), то этот метод создает помехи другим видам радиопередачи.

4.3.4. Ортогональное частотное разделение

При более высоких скоростях передачи применяется принципиально иной метод обработки сигнала – ортогональное частотное разделение каналов с мультиплексированием (Orthogonal Frequency Division Multiplexing, OFDM). Идея данного метода заключается в том, что поток передаваемых данных распределяется по множеству частотных подканалов и передача ведется параллельно на всех этих подканалах. При этом высокая скорость передачи достигается именно за счет одновременной передачи данных по всем каналам, а скорость передачи в отдельном подканале может быть относительно небольшой.

Поскольку в каждом из частотных подканалов скорость передачи данных можно сделать не слишком высокой, это создает предпосылки для эффективного подавления межсимвольной интерференции.

При частотном разделении каналов необходимо, чтобы ширина отдельного канала была, с одной стороны, достаточно узкой для минимизации искажения сигнала в пределах отдельного канала, а с другой — достаточно широкой для обеспечения требуемой скорости передачи. Кроме того, для экономного использования всей полосы канала, разделяемого на подканалы, желательно как можно более плотно расположить частотные подканалы, но при этом избежать межканальной интерференции, чтобы обеспечить полную независимость каналов друг от друга. Частотные каналы, удовлетворяющие перечисленным требованиям, называются ортогональными. Несущие сигналы всех частотных подканалов (а

точнее, функции, описывающие эти сигналы) ортогональны друг другу.

Рассмотренный способ деления широкополосного канала на ортогональные частотные подканалы называется ортогональным частотным разделением с мультиплексированием. Одним из ключевых преимуществ метода OFDM является сочетание высокой скорости передачи с эффективным противостоянием многолучевому распространению. Сама по себе технология OFDM не устраняет многолучевого распространения, но создает предпосылки для устранения эффекта межсимвольной интерференции. Неотъемлемой частью технологии OFDM является защитный интервал (Guard Interval, GI) — циклическое повторение окончания символа, пристраиваемое в начале символа.

Защитный интервал является избыточной информацией и в этом смысле снижает полезную (информационную) скорость передачи, но именно он служит защитой от возникновения межсимвольной интерференции. Эта избыточная информация добавляется к передаваемому символу в передатчике и отбрасывается при приеме символа в приемнике.

Наличие защитного интервала создает временные паузы между отдельными символами, и если длительность охранного интервала превышает максимальное время задержки сигнала в результате многолучевого распространения, то межсимвольной интерференции не возникает. При использовании технологии OFDM длительность защитного интервала составляет одну четвертую длительности самого символа. При этом сам символ имеет длительность 3,2 мкс, а охранный интервал — 0,8 мкс. Таким образом, длительность символа вместе с защитным интервалом составляет 4 мкс.

5. ОБЗОР СОВРЕМЕННЫХ МЕТОДОВ АНАЛОГО-ЦИФРОВОГО ПРЕОБРАЗОВАНИЯ

5.1. КЛАССИФИКАЦИЯ УСТРОЙСТВ АНАЛОГО- ЦИФРОВОГО ПРЕОБРАЗОВАНИЯ ПО ФАКТОРУ ТОЧНОСТЬ/БЫСТРОДЕЙСТВИЕ

Среди современных устройств аналого-цифрового преобразования обычно выделяют четыре основных типа: АЦП последовательных приближений, интегрирующие и сигма-дельта АЦП, АЦП параллельного и полупараллельного преобразования; архитектура субполосных АЦП является относительно новой технологией, совмещающей высокую разрешающую способность и скорость преобразования (рис. 5.1). Среди субполосных АЦП выделяют два типа: АЦП с временным и с частотным разделением сигнала на субполосы [1].

Так как принцип частотного разделения сигнала является основой реализации быстрого алгоритма вейвлет-преобразования, субполосные АЦП с временным и, в особенности, частотным разделением сигнала далее рассмотрены более подробно.

Определим кратко основные характеристики современных АЦП.

Наиболее быстродействующими являются параллельные и полупараллельные АЦП, архитектура которых представляет собой набор компараторов. Параллельные АЦП проводят одно сравнение входного сигнала на 2^n компараторах и обеспечивают 10 бит разрешения при 100 Мотсч./сек.; полупараллельные выполняют несколько шагов параллельного сравнения, при этом разрешение до 14 бит достигается при скорости преобразования 1 Мотсч./сек.

Принцип последовательных сравнений входного сигнала с опорным напряжением за несколько итераций позволяет повысить точность преобразования; типовое значение разрешающей способности АЦП последовательных приближений составляет

16-18 бит, однако, при достаточно невысокой скорости порядка 100 к отсч./сек.

Интегрирующие и сигма-дельта АЦП являются наиболее точными устройствами аналого-цифрового преобразования. В случаях, когда скорость преобразования не является решающим условием, применение АЦП этого типа позволяет строить высокоточные измерительные каналы. Их разрешающая способность достигает 24 и более бит.



Рис. 5.1 Классификация современных АЦП по фактору точность/быстродействие

5.2. СУБПОЛОСНЫЕ АЦП С ВРЕМЕННЫМ РАЗДЕЛЕНИЕМ СИГНАЛА

Объединение скоростных свойств параллельных АЦП и точности АЦП последовательных приближений привело к созданию субполосных структур АЦП (рис. 5.2).

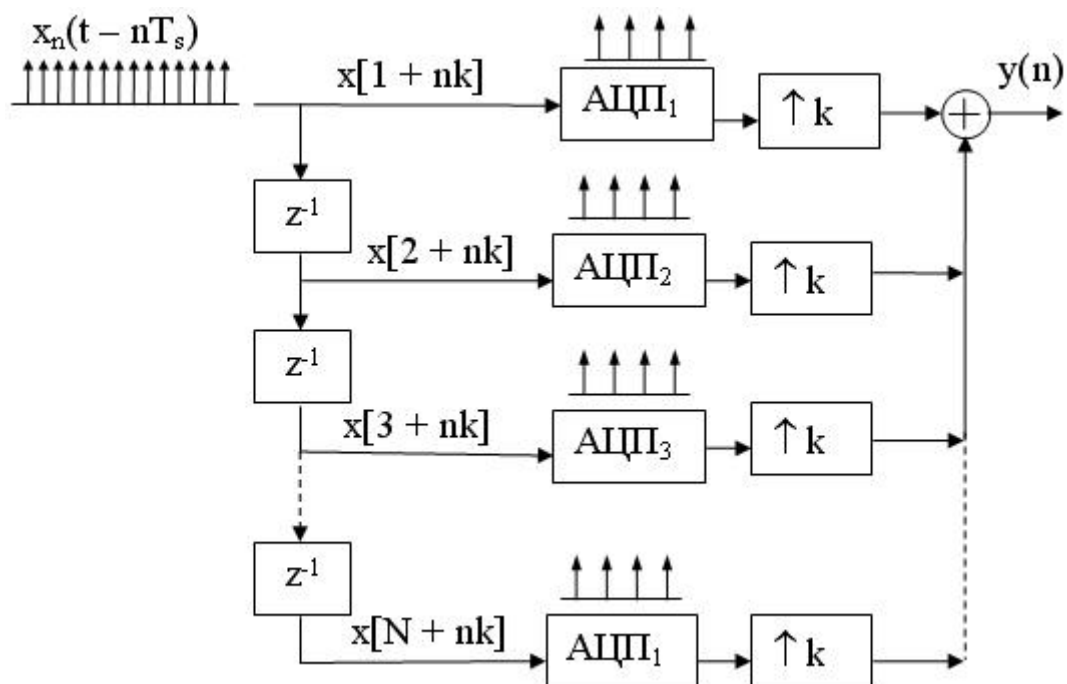


Рис. 5.2. Субполосный АЦП с временным разделением сигнала

В зарубежной литературе, как правило, используется термин – time-interleaving ADC (АЦП с параллельным временем преобразования). Данные системы были разработаны в середине 90-х годов и развиваются в основном компаниями Hewlett-Packard и Tektronix. Они представляют собой многоканальную структуру, состоящую из k одинаковых АЦП последовательных приближений. АЦП в каждом канале относительно соседнего запускаются с задержкой во времени T_s , равной интервалу дискретизации входного сигнала. Таким образом, первый отсчет входного сигнала поступает на вход первого канала, второй – второго и т.д. Тогда $k+1$ отсчет снова поступает на вход первого канала через интервал kT_s . и т.д. Для получения выходной кодовой последовательности выходы каналов суммируются.

Общая скорость преобразования системы $1/T_s$ может быть достаточно высокой, тогда как АЦП в каждом канале работает на частоте в k раз ниже. Разрешающая способность системы при этом определяется разрядностью АЦП. Данный принцип построения

высокоскоростных высокоточных АЦП обладает, однако, рядом недостатков:

1. В частотной области каждый АЦП действует во всем диапазоне входного сигнала. Это означает, что в точке суммирования на выходе погрешности линейности и шумы квантования всех АЦП складываются. Это свойство снижает, в итоге, общее разрешение системы.

2. Требуются точные цепи управления временными задержками, недостатки реализации которых также снижают разрешающую способность за счет внесения аддитивной погрешности.

3. Необходимы точные компоненты, на которых реализуется АЦП, так как рассогласование коэффициентов усиления каналов вносит мультипликативную погрешность.

Первый недостаток является методическим, обусловленным архитектурой этого типа АЦП. Он свойственен данным АЦП в целом и не может быть устранен в рамках принятого принципа временного разделения каналов. Отчасти эта погрешность снижается введением в каналы схем компенсации нелинейности.

Структуры АЦП с частотным разделением каналов, устраняющие указанный методический недостаток, явились прямым улучшением АЦП с временным разделением.

Типовые аддитивные и мультипликативные погрешности могут быть уменьшены путем совершенствования технологии реализации цепей временной задержки, подбора компонентов со сходными характеристиками (например, встроенных в микропроцессоры многоканальных АЦП или АЦП в интегральном исполнении), компенсацией нелинейности.

В идеальном случае разрешающая способность описанной системы равна разрешению АЦП в каналах. Влияние типовых погрешностей приводит к потере 2–3 бит.

Типовые значения для субполосных АЦП составляют 8 - 10 Г отсч./сек. при 8-битном разрешении для сигнала в полосе 0 - 2 ГГц.

Принцип построения АЦП с временным разделением сигнала повышает скоростные свойства преобразователя и расширяет его динамический диапазон.

5.3. УЛУЧШЕНИЕ ТОЧНОСТНЫХ ХАРАКТЕРИСТИК БЫСТРЫХ АЦП. СУБПОЛОСНЫЕ АЦП С ЧАСТОТНЫМ РАЗДЕЛЕНИЕМ СИГНАЛА

Разделение сигнала в частотной области на субполосы представляет собой важный раздел цифровой обработки сигналов и изображений, радиотехники, теории связи. Широкое применение метод частотного разделения имеет в теории обработки речевых сигналов для построения частотных скремблеров [9]. Главное достоинство метода состоит в независимой обработке сигнала каждой субполосы. Погрешности каждого канала относятся, в данном случае, только к своей субполосе и не действуют по всему частотному диапазону. Разделение сигнала выполняется набором, в общем случае, полосно-пропускающих фильтров. Такая линейка фильтров называется в литературе блоком или банком фильтров. Для восстановления исходного сигнала также требуется банк фильтров (рис. 5.3) [1].

Ограничение частотного диапазона сигнала в каждой субполосе в k раз позволяет также в k раз понизить частоту дискретизации kf_s в соответствии с теоремой Котельникова. Блоки цифровой обработки работают на частоте f_s в k раз ниже частоты входного сигнала, что позволяет применять более медленные точные устройства. При этом общий частотный диапазон системы в k раз шире частотного диапазона каждого канала (рис. 5.4).

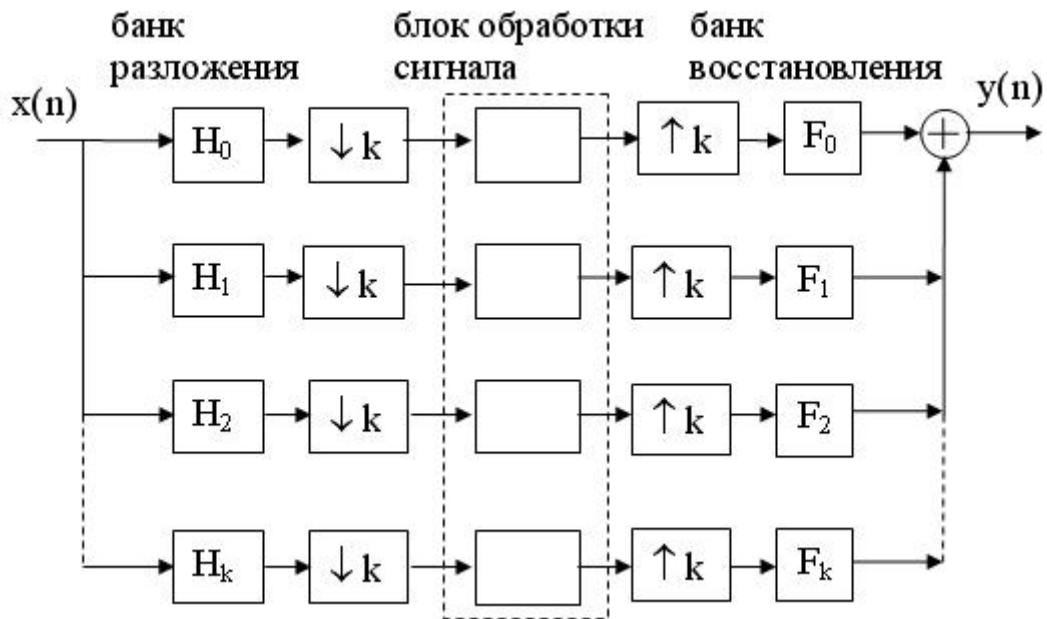


Рис. 5.3. Структура цифровой обработки сигнала с частотным разделением с помощью банка фильтров

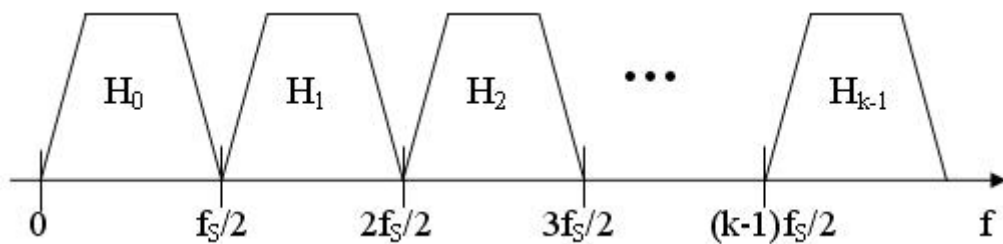


Рис. 5.4. АЧХ k -канального банка фильтров. Исходная частота дискретизации kf_s . Рабочая частота дискретизации каждого канала f_s

Первоначально такая структура была полностью дискретно-временной. По этой причине ее непосредственное применение в АЦП было невозможным. Для создания АЦП с частотным разделением сигнала потребовалась адаптация структуры. Создание *гибридного банка фильтров*, где банк разложения является аналоговым (линейкой активных аналоговых фильтров на операционных усилителях или на переключаемых конденсаторах), а банк восстановления – цифровым, послужило толчком к дальнейшему совершенствованию класса субполосных АЦП (рис. 5.5).

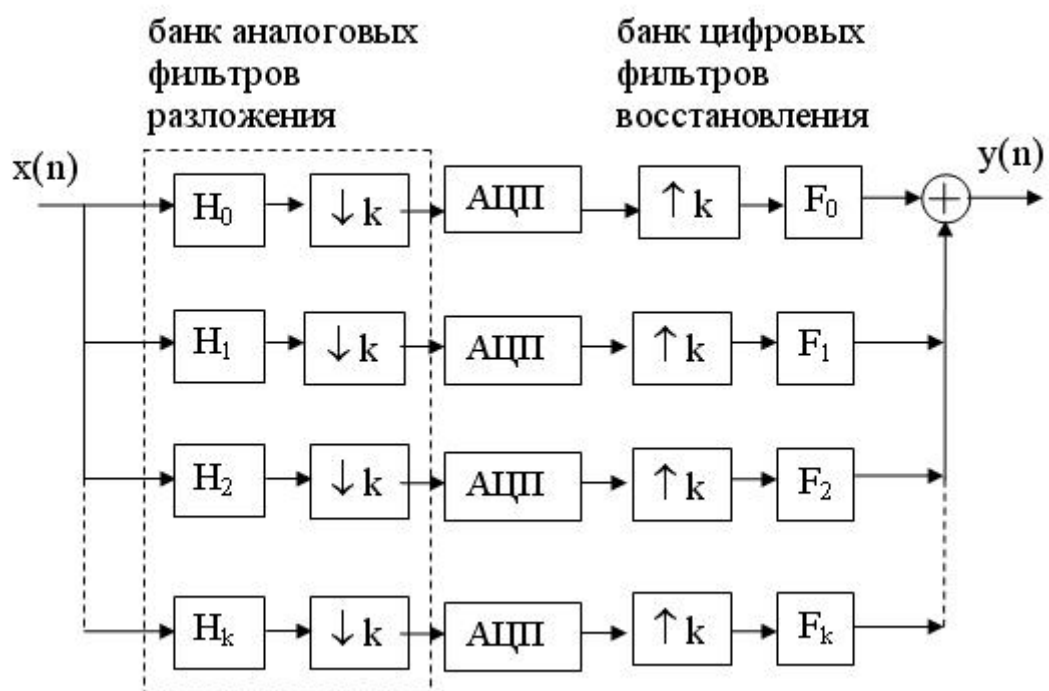


Рис. 5.5. АЦП с частотным разделением сигнала на основе гибридного банка фильтров

6. ПОСТРОЕНИЕ УЗЛА АЦП НА ОСНОВЕ СИГМА-ДЕЛЬТА ПРЕОБРАЗОВАНИЯ

По принципу действия сигма-дельта АЦП относятся к классу интегрирующих АЦП, однако, имеют ряд характерных особенностей. Принцип действия сигма-дельта АЦП основан на использовании передискретизации при выборке отсчетов входного сигнала, применении одно- или многобитного квантователя и цифрового фильтра (рис. 6.1). Выходной код, формируемый квантователем, обрабатывается цифровым фильтром. Фильтр выполняет усреднение значения сигнала за интервал времени. В предположении, что полоса интереса полезного сигнала находится на низких частотах, особенности архитектуры сигма-дельта АЦП позволяют вытеснить шум, вносимый квантователем, в область высоких частот, где он удаляется при фильтрации. Преобразование сигнала типовой структурой сигма-дельта модулятора с обратной связью дало название этому типу АЦП: значение сигнала на выходе вычитается из входного (*операция Δ*), а разница интегрируется в качестве входа в следующем такте (*операция Σ*).

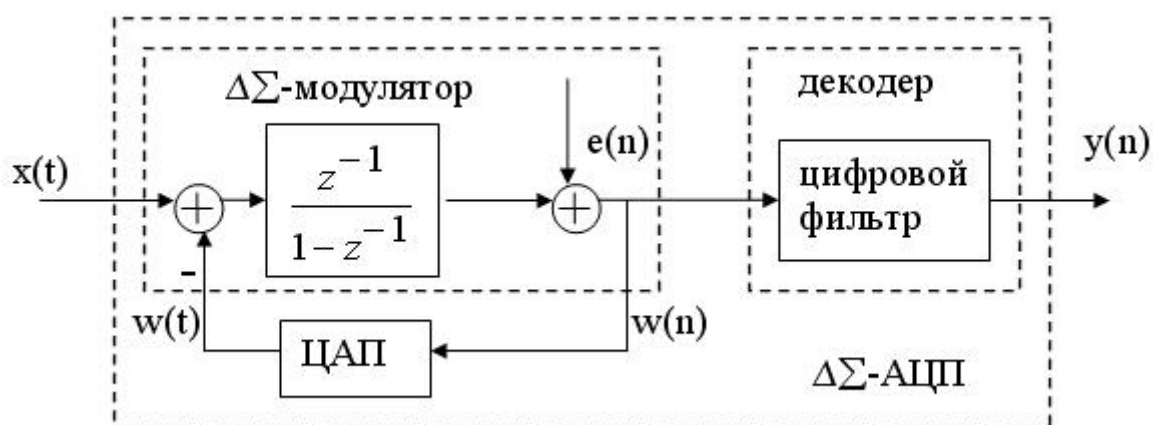


Рис. 6.1. Структурная схема сигма-дельта АЦП

Разрешающая способность этого типа АЦП обычно составляет 16-24 бит, при скорости преобразования до 100 кГц.

В структуре сигма-дельта АЦП выделяют два основных узла: сигма-дельта модулятор и декодер (фильтр). По типу системы

структуры сигма-дельта АЦП бывают непрерывно-временные и дискретно-временные. Существует два схемотехнических принципа построения сигма-дельта АЦП: использование аналоговых интеграторов и переключаемых (летающих) конденсаторов. Наибольшие трудности вызывает обеспечение устойчивости каскадов при создании сигма-дельта модуляторов высоких (выше второго) порядков. Совершенствование структур сигма-дельта АЦП в основном касается модуляторов, тогда как вид цифрового фильтра практически не изменялся: наиболее распространенным является применение фильтров типа sinc. Однако, совершенствование процедуры цифровой фильтрации также может способствовать улучшению функциональности сигма-дельта АЦП. В частности, разработка алгоритмов фильтрации, воздействующих не только на шум квантования, но и на шум, действующий на входе преобразователя. В классических структурах сигма-дельта АЦП удаление этого шума выполняется при предварительной фильтрации.

Данная задача особенно актуальна в том случае, когда полезный сигнал и шум имеют перекрывающиеся частотные диапазоны. Действующий на входе модулятора белый шум преобразуется также как и полезный сигнал. Проблема разделения смеси сигнала с шумом в сигма-дельта АЦП была поставлена в работах по построению оптимального декодера и решалась в соответствии с критерием максимума отношения сигнал/шум.

В задачах, где сигнал имеет сложную форму (речь, звук, сигнал электрокардиограмма, томограмма и т.д.) помимо удаления шумов требуется обеспечить наименьшее искажение полезного сигнала.

В данной работе были исследованы методы и алгоритмы нелинейной оптимальной фильтрации на основе вейвлет-преобразования, позволяющие выполнять разделение сигнала и шума в одной полосе. Описание свойств шума квантования было выполнено во многих работах, посвященных сигма-дельта АЦП [1]. При этом обеспечивается наименьшее сглаживание особенностей

сигнала, неизбежное при применении алгоритмов линейной фильтрации. К достоинствам этого метода можно отнести и вычислительную экономичность, так как число операций умножения и сложения прямо пропорционально числу отсчетов входного сигнала.

6.1. ПРИМЕНЕНИЕ ПЕРЕДИСКРЕТИЗАЦИИ ПРИ АНАЛОГО-ЦИФРОВОМ ПРЕОБРАЗОВАНИИ

АЦП с передискретизацией сигнала состоит из следующих узлов: в общем случае n -битного квантователя и цифрового фильтра, работающих на частоте, превышающей максимальную частоту в спектре входного сигнала более чем дважды, и прореживающего блока, снижающего частоту дискретизации (рис. 6.2).

Наличие квантователя делает устройство нелинейным, однако широко используется линейная модель АЦП, в которой блок квантователя замещен источником аддитивного белого шума в предположении об отсутствии корреляции внутри шумового процесса и между сигналом и шумом.

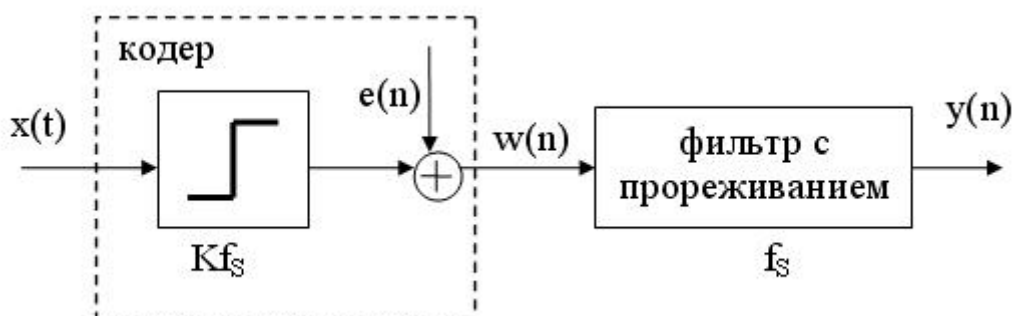


Рис. 6.2. Структурная схема АЦП с передискретизацией

Пусть входной сигнал $x(t)$ ограничен по полосе до некоторой частоты $f < \frac{f_s}{2}$, где f_s – критическая частота дискретизации (частота Найквиста),

$K = \frac{f_q}{f_s}$, где K – коэффициент передискретизации,

f_q – рабочая частота квантователя.

Пусть входной сигнал также ограничен по уровню $|x(Kf_s)| < U_{\max}$.

При квантовании вносится шум, распределенный вдоль полосы частот $f < Kf_s$.

Известное значение мощности шума квантования

$$S = \frac{q^2}{12}, \quad (6.1)$$

где q — величина ступени квантователя, можно преобразовать к виду:

$$S = \frac{1}{12} \left(\frac{2U_{\max}}{2^n} \right)^2, \quad (6.2)$$

записав в явном виде значение для ступени квантователя $q = \frac{2U_{\max}}{2^n}$,

где n — число уровней квантователя.

Спектральная плотность мощности, полученная в результате интегрирования по частоте от 0 до Kf_s составляет:

$$S_N(f) = \frac{1}{6Kf_s} \left(\frac{2U_{\max}}{2^n} \right)^2. \quad (6.3)$$

Заменяя выражение в скобках для ступени квантователя на q , запишем

$$S = \frac{2q^2}{3Kf_s}. \quad (6.4)$$

Фильтр ослабляет частоты $f > \frac{f_s}{2}$, после чего выполняется снижение частоты дискретизации в K раз.

Исходя из этих соотношений, можно сделать вывод о том, что при передискретизации шум квантования распределен по полосе в K раз более широкой. Мощность шума квантования после фильтрации снижается в K раз.



Рис. 6.3. Влияние изменения частоты дискретизации на мощность шума квантования

$$S_F = \frac{1}{12K} \left(\frac{2U_{\max}}{2^n} \right)^2 = \frac{1}{12} \left(\frac{2U_{\max}}{K^{\frac{1}{2}} 2^n} \right)^2 = \frac{1}{12} \left(\frac{2U_{\max}}{2^{1/2 \log K} \cdot 2^n} \right)^2 =$$

$$= \frac{1}{12} \left(\frac{2U_{\max}}{2^{n+1/2 \log K}} \right)^2. \quad (6.5)$$

Результирующее отношение сигнал/шум увеличивается в K раз при одновременном повышении разрешающей способности преобразователя $n_K = n + \frac{1}{2} \log K$ (рис. 6.3).

6.2. ФОРМИРОВАНИЕ СПЕКТРА ШУМА КВАНТОВАНИЯ С ПОМОЩЬЮ ОБРАТНОЙ СВЯЗИ

Введение отрицательной обратной связи в АЦП с передискретизацией (рис. 6.4) позволяет вытеснить шум квантования в область высоких частот.

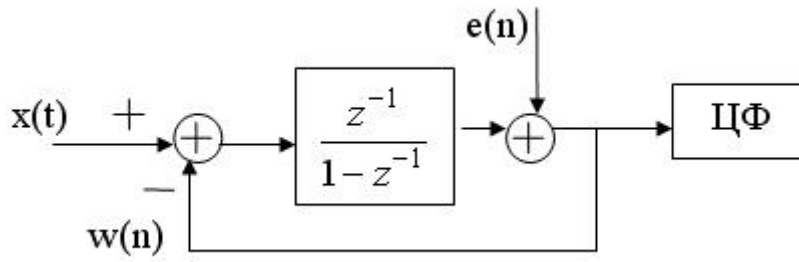


Рис. 6.4. Сигма-дельта модулятор I порядка с обратной связью

Передаточная функция для полезного сигнала (STF) сигма-дельта модулятора I порядка имеет вид:

$$A(z) = \frac{\frac{z^{-1}}{1 - z^{-1}}}{1 + \frac{z^{-1}}{1 - z^{-1}}} = z^{-1}, \quad (6.6)$$

для шума квантования (NTF):

$$B(z) = \frac{1}{1 + \frac{z^{-1}}{1 - z^{-1}}} = 1 - z^{-1} \quad (6.7)$$

при $z = e^{j2\pi f / Kf_s}$ запишем

$$B(f) = 1 - e^{j2\pi f / Kf_s} \quad STF = 1 - NTF. \quad (6.8)$$

При преобразовании линейной моделью спектральной плотности мощности имеем:

$$S_{III} = S * |B(f)|^2, \quad |B(f)|^2 = \left| 1 - e^{j2\pi f / Kf_s} \right|^2 = 4 \sin^2 \frac{\pi \cdot f}{Kf_s}. \quad (6.9)$$

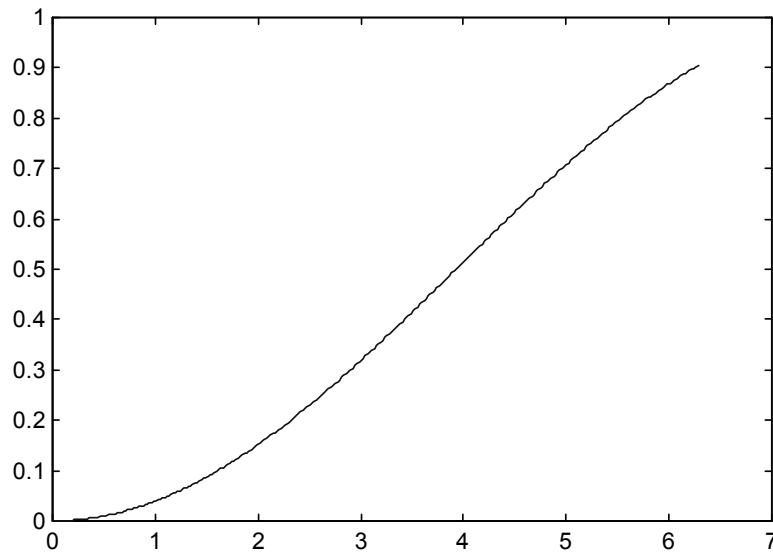


Рис. 6.5. Спектральная плотность мощности шума квантования в полосе сигнала

Спектральная плотность мощности шума квантования в полосе сигнала имеет следующий вид (рис. 6.5).

$$S_{ш} = \frac{2U_0^2}{3Kf_s} 4 \sin^2 \frac{\pi \cdot f}{Kf_s} = \frac{8U_0^2 \sin^2 \frac{\pi f}{Kf_s}}{3Kf_s}, \quad (6.10)$$

интегрируя по полосе частот от 0 до $f_s/2$ получим дисперсию шума квантования:

$$D_{ш} = \frac{U_0^2 \pi^2}{9} \frac{1}{K^3}. \quad (6.11)$$

Отношение сигнал/шум имеет вид:

$$SNR = -3,41 + 30 \log K, \quad \text{или} \quad 9 \text{ дБ / октаву}. \quad (6.12)$$

Данное теоретически полученное значение соответствует идеальному фильтру в составе сигма-дельта АЦП.

Исследования в области цифровой обработки двоичной последовательности сигма-дельта модулятора показали, что предположение об отсутствии корреляции не всегда справедливо. Вблизи границ диапазона преобразования очевидно возрастание

мощности шума квантования, вследствие корреляции между ошибками квантования.

6.3. ФОРМИРОВАНИЕ СПЕКТРА ШУМА КВАНТОВАНИЯ С ПОМОЩЬЮ УСРЕДНЯЮЩИХ ОКОН

Наиболее просто реализуется фильтр типа sinc^M , представляющий собой свертку M прямоугольных окон. Его передаточная функция:

для sinc^1 :

$$W_D(z) = \frac{1 - z^{-N}}{N(1 - z^{-1})}, \quad (6.13)$$

где N — длина фильтра (окно Дирихле),

для sinc^2

$$W_B(z) = \frac{(1 - z^N)^2}{N^2(1 - z^{-1})^2} \text{ (окно Бартлета)}. \quad (6.14)$$

Тогда результирующая мощность шума на выходе АЦП с sinc^1

$$S = S_{ш} * |W_D(z)|^2 = \frac{8U_0^2 \sin^2 \frac{\pi Nf}{Kf_S}}{3NKf_S}. \quad (6.15)$$

Аналогично, для sinc^2 :

$$S = S_{ш} * |W_B(z)|^2 = \frac{8U_0^2 \sin^4 \frac{\pi Nf}{2Kf_S}}{3N^4 Kf_S \sin^2 \left(\frac{\pi f}{Kf_S} \right)}. \quad (6.16)$$

Применение весовых функций цифровых фильтров в сигма-дельта АЦП, декоррелирующих сигнал, способствует понижению мощности шума квантования во всем диапазоне преобразования.

Цифровые фильтры для вейвлет-преобразования, обладающего свойством, подобно всем ортогональным преобразованиям, декоррелировать сигнал, в то же время подготавливают двоичную

последовательность к нелинейной фильтрации, удаляющей шум на входе преобразователя.

С увеличением порядка модулятора изменяется передаточная функция для шума квантования и, следовательно, спектральная плотность мощности.

Для сигма-дельта модулятора II порядка:

$$S_{\text{ш}} = \frac{32U_0^2 \sin^4 \frac{\pi f}{Kf_S}}{3Kf_S} \quad D_{\text{ш}} = \frac{U_0^2 \pi^4}{15} \frac{1}{K^5}. \quad (6.17)$$

Сигма-дельта модулятор II порядка обладает хорошей устойчивостью и обнаруживает меньшую корреляцию ошибок квантования. Увеличение разрешающей способности повышает требования к управляющим цепям интеграторов и требует защиты интеграторов и квантователя от перегрузок.

Применение сигма-дельта модуляторов высших порядков позволяет получить высокую разрешающую способность при малом количестве бит квантователя. Это свойство усиливается с увеличением порядка модулятора, однако, поддержание устойчивости таких структур является сложной задачей.

7. ДИСКРЕТНОЕ ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЕ

Теоретическая база метода дискретного вейвлет-преобразования (ДВП) является уже хорошо изученной областью математики, тогда как его приложение к обработке сигналов стало развиваться с конца 80-х и в 90-е годы в результате создания быстрого алгоритма вейвлет-преобразования – субполосного разделения сигнала многоуровневым банком фильтров [6]. К основным достоинствам метода следует отнести принципиально более информативное и компактное представление сигнала, и, как следствие, снижение вычислительных затрат при его реализации. В работах по математической статистике были проведены исследования изменения энтропии сигнала при его разложении, в частности в вейвлет-базисе, на основании чего, по степени информативности, вейвлет-преобразование сигнала мы можем отнести к методам оптимального кодирования [6].

С точки зрения ЦОС вейвлет-преобразование состоит в представлении сигнала в виде набора время-частотных окон различной разрешающей способности [6]. Это позволяет локализовать такие особенности сигнала (выбросы, всплески, разрывы, паузы и др.), обнаружение которых может представлять затруднение при использовании, например, такого распространенного метода как быстрое преобразование Фурье. В связи с этим наибольшее распространение вейвлет-преобразование сигнала имеет в области исследования детерминированных сигналов сложной формы, а также динамики случайных, в частности нестационарных, процессов.

Субполосные фильтры, служащие для реализации быстрого алгоритма вейвлет-преобразования представляют интерес и при их обычном использовании в качестве фильтров низких (высоких) частот.

Статистические методы анализа случайных функций позволяют разделить сигнал и шум, находящиеся в одном диапазоне частот, т.е. модифицировать процедуру фильтрации. Методы фильтрации, основанные на применении вейвлет-преобразования как подготовительного этапа обработки, относятся к категории нелинейных методов фильтрации. Однако, в отличие от известных нелинейных средств (медианная фильтрация, кепстральный анализ) вейвлет-фильтрация не требует реализации сложных алгоритмов статистической обработки, больших временных и вычислительных затрат, работы с вероятностными моделями и пр. Эффективная компрессия и фильтрация сигналов являются одними из распространенных проблем в создании измерительных систем и могут успешно решаться путем применения методов цифровой обработки сигнала, основанных на вейвлет-преобразовании. Очевидно, что любая обработка сигнала в реальном времени выполняется с помощью вычислительных устройств над дискретными сигналами. Поэтому непрерывное вейвлет-преобразование также действует в дискретном времени над конечной реализацией сигнала. Причина того, что данное преобразование называется непрерывным (в отличие от дискретного вейвлет-преобразования) состоит в том, что изменение сдвига k может происходить сколь угодно плавно и определяется разрядной сеткой вычислителя. В дискретном вейвлет-преобразовании изменение шкального множителя α зависит от номера шкалы как $\alpha=2^j$, сдвиг k равен интервалу дискретизации сигнала.

7.1. БЫСТРЫЙ АЛГОРИТМ ДИСКРЕТНОГО ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЯ

Дискретным вейвлет-преобразованием функции $f(t) \in L_2(\mathbf{R})$ называется следующее ее разложение (для наглядности поэтапно рассмотрим процесс разложения):

$$\begin{aligned}
f^1(t) &= \sum_{1,k} \psi\left(\frac{t-k}{2}\right) \cdot d_{1,k} + \sum_{1,k} \varphi\left(\frac{t-k}{2}\right) \cdot s_{1,k} \\
f^2(t) &= \sum_{1,k} \psi\left(\frac{t-k}{2}\right) \cdot d_{1,k} + \sum_{2,k} \psi\left(\frac{t-k}{4}\right) \cdot d_{2,k} + \sum_{2,k} \varphi\left(\frac{t-k}{4}\right) \cdot s_{2,k} \\
f^3(t) &= \sum_{1,k} \psi\left(\frac{t-k}{2}\right) \cdot d_{1,k} + \sum_{2,k} \psi\left(\frac{t-k}{4}\right) \cdot d_{2,k} + \\
&+ \sum_{3,k} \psi\left(\frac{t-k}{8}\right) \cdot d_{3,k} + \sum_{3,k} \varphi\left(\frac{t-k}{8}\right) \cdot s_{3,k} \\
f^4(t) &= \sum_{1,k} \psi\left(\frac{t-k}{2}\right) \cdot d_{1,k} + \sum_{2,k} \psi\left(\frac{t-k}{4}\right) \cdot d_{2,k} + \sum_{3,k} \psi\left(\frac{t-k}{8}\right) \cdot d_{3,k} + \\
&+ \sum_{4,k} \psi\left(\frac{t-k}{16}\right) \cdot d_{4,k} + \sum_{4,k} \varphi\left(\frac{t-k}{16}\right) \cdot s_{4,k},
\end{aligned}$$

тогда на j -том шаге имеем:

$$f^j(t) = \sum_{j=1}^N \sum_k \psi\left(\frac{t-k}{2^j}\right) \cdot d_{j,k} + \sum_k \varphi\left(\frac{t-k}{2^N}\right) \cdot s_{N,k}, \quad (7.1)$$

где:

$$\psi(t-k) = (-1)^n \varphi(t-N+k) \quad (7.2)$$

является семейством базисных функций, j, k – соответственно глубина разложения и шаг суммирования, а $d_{j,k}$ и $s_{j,k}$ – коэффициенты разложения:

$$d_{j,k} = \sum_k \psi\left(\frac{t-k}{2^j}\right) f(t) \quad s_{j,k} = \sum_k \varphi\left(\frac{t-k}{2^j}\right) f(t). \quad (7.3)$$

Коэффициенты разложения $d_{j,k}$ и $s_{j,k}$ для каждого значения j и k отличны от нуля также на протяжении лишь нескольких шагов суммирования, т.е. короткой области сигнала.

Для выполнения дискретного вейвлет-преобразования был разработан быстрый алгоритм, в котором переход к представлению сигнала $f(t)$ с переменным разрешением по частоте и времени выполняется с помощью структуры фильтров.

Все фильтры разложения $H_0^j(z)$, $F_0^j(z)$ и восстановления $H_1^j(z)$, $F_1^j(z)$ имеют коэффициенты, одинаковые с фильтрами на первой ступени разложения/восстановления. Таким образом, расчет фильтров выполняется однократно. Данная процедура соответствует определению базисной вейвлет-функции.

Шкалирование выполняется за счет изменения частоты дискретизации. Количество ступеней банка соответствует числу этапов шкалирования вейвлет-функции. В силу простоты реализации и минимальной сложности вычислений (порядка N – длины сигнала) данный алгоритм называется быстрым вейвлет-преобразованием (БВП).

Создание алгоритма БВП сделало возможным применение на практике преимуществ вейвлет-преобразования:

1. локализованного частотно-временного представления произвольного сигнала с одновременным информативным разрешением по частоте и времени;
2. компактной записи сигнала в виде комбинации коэффициентов разложения, конечной для большинства сигналов, обусловленной сходной формой сигнала и базисных функций;
3. эффективной компрессии сигнала;
4. нелинейной фильтрации;

Наращивание банка выполняется обычно по низкочастотной ветви (рис. 7.1).

Однако, в ряде случаев, например, для сигнала электрокардиограммы целесообразно изменить структуру дерева (рис. 7.2).

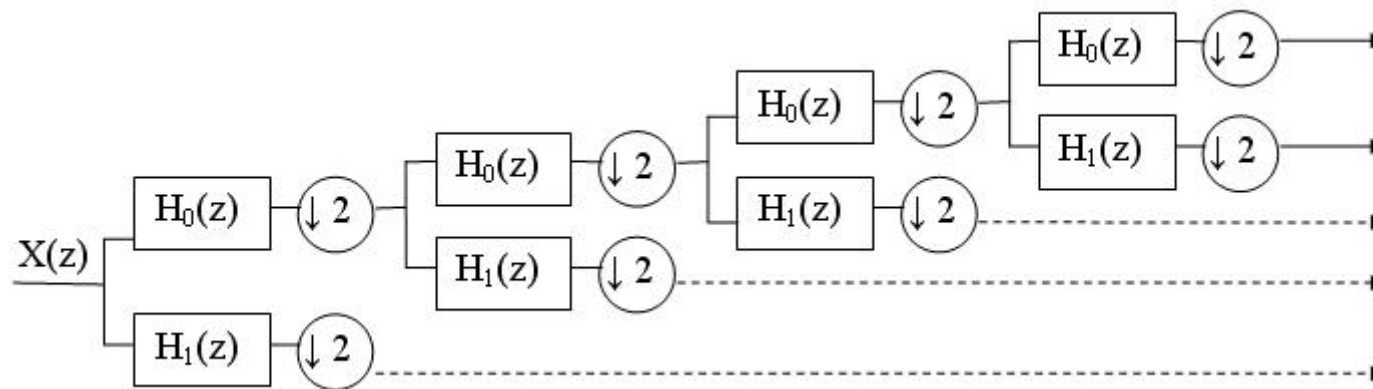


Рис. 7.1. Каскадирование банка фильтров по низкочастотной ветви

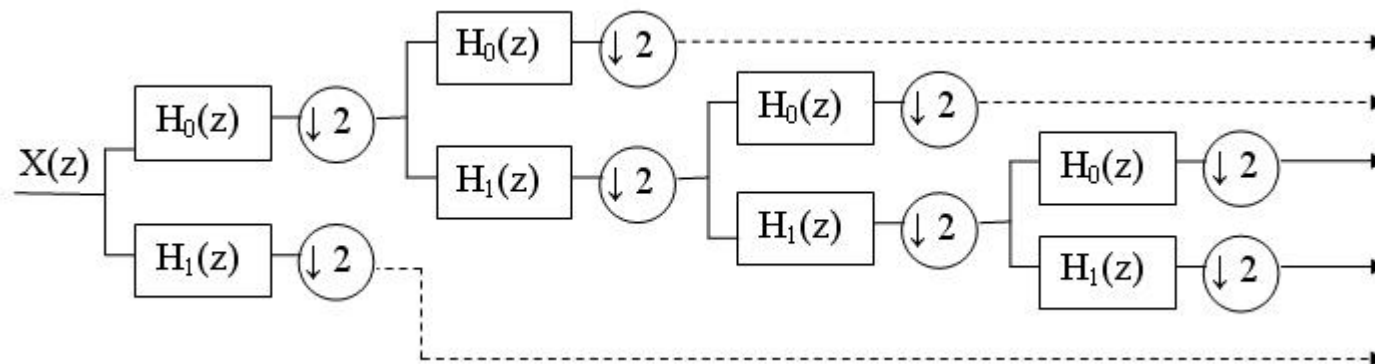


Рис. 7.2. Пример оптимизированного банка фильтров

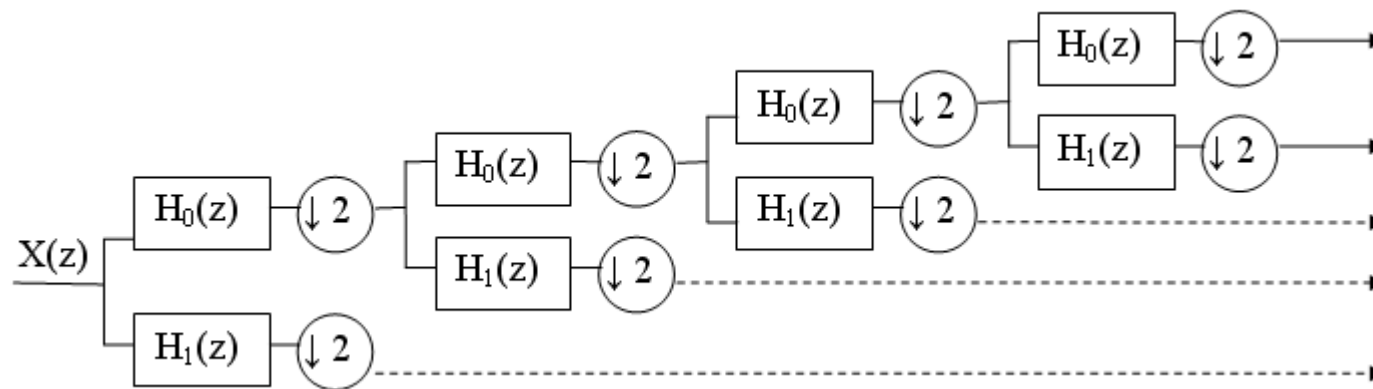


Рис. 7.3. Банк фильтров для прямого дискретного ВП сигнала

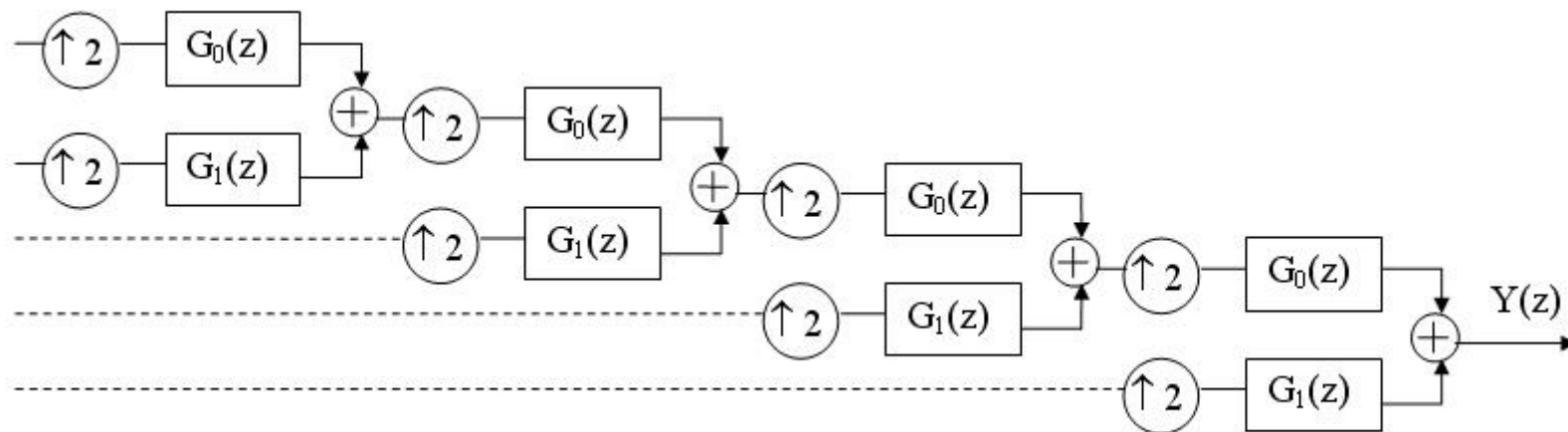


Рис. 7.4 Банк фильтров для обратного дискретного ВП сигнала

Теоретически выбор оптимально информативного пути разложения может быть определен путем оценки энтропийной характеристики каждого узла (уровня) разложения:

$$\begin{aligned}\Xi^j &= \sum_k \psi(2^{-j}x - k) \cdot f(x) \cdot \log[\psi(2^{-j}x - k) \cdot f(x)] = \\ &= \sum_k d_{jk} \cdot \log d_{jk},\end{aligned}\tag{7.4}$$

где d_{jk} – вейвлет-коэффициенты j -го узла. Формирование дерева идет по пути уменьшения энтропийной оценки.

Обратное преобразование может быть получено с помощью дуальной структуры (рис. 7.3, 7.4):

Число уровней разложения j может быть произвольным.

Здесь $\varphi(\frac{t-k}{2^j})$ и $\psi(\frac{t-k}{2^j})$ – импульсные характеристики (ИХ)

пары фильтров; множитель 2^j соответствует растяжению ИХ фильтров, т.е. сужению их полосы пропускания, шаг суммирования k определяется интервалом дискретизации входного сигнала. Последовательности $\{s_{j,k}\}$ и $\{d_{j,k}\}$ получаются при свертке $f(t)$ с ИХ фильтров. Расчет ИХ вейвлет-фильтров изложен подробно в приложении. Разработчик также может получить готовые коэффициенты вейвлет-фильтров, пользуясь средствами пакетов Matlab, Matematika и др.

При представлении ВП в виде структуры фильтров видно, что в процессе ВП из сигнала выделяется его основная низкочастотная огибающая – $s_{N,k}$ и ряд его особенностей, сравнительно более высокочастотных – $d_{j,k}$. Описанная техника вейвлет-преобразования позволяет варьировать длительность временного окна. Для определения низкочастотных составляющих сигнала нужны более длительные интервалы времени наблюдения сигнала, чем при выявлении его высокочастотных компонент, где требуются относительно короткие временные окна.

7.2. СРЕДСТВА ОБРАБОТКИ СИГНАЛОВ, ОСНОВАННЫЕ НА ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИИ

7.2.1. Применение вейвлет-преобразования для сжатия сигналов

Сжатие методом вейвлет-преобразования относится к методам сжатия с частичной потерей информации. Принцип компрессии основан на том, что в сжимаемом сигнале имеются такие детали, потеря которых не ухудшает качества восприятия сигнала. К таким деталям относятся, например, части аудио- и видеосигналов, не воспринимаемые зрением или слухом. Оптимизация кодирования повторяющихся или близких по значению коэффициентов разложения также приводит к сжатию сигнала. Исследования изменения энтропии сигнала при переходе к его ВП показали, что информативность вейвлет-коэффициентов весьма высока, на основании чего ВП относят к методам оптимального кодирования. Очевидно, что часть теряемой информации всегда различна и определяется выбранной стратегией сжатия.

1. Компрессия на основе жесткой пороговой обработки коэффициентов

По полученному массиву вейвлет-коэффициентов $\{d_{j,k}\}$ строится гистограмма (рис. 7.5, а) и сортируется по возрастанию (рис. 7.5, б)

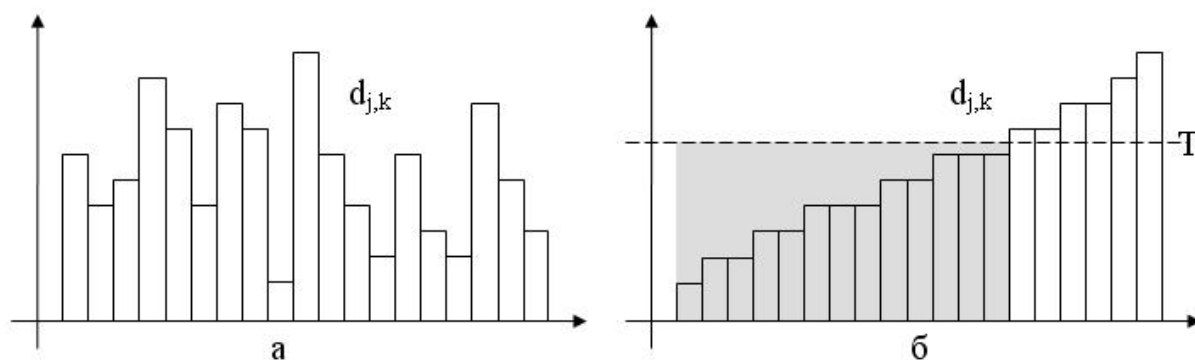


Рис. 7.5. Сортировка $d_{j,k}$

Далее подбирается значение порога T . Используются следующие статистические критерии выбора порогового значения:

1. фиксированный порог: $T = (\sqrt{2 \ln N}) * \sigma$, где N – длина $\{d_{j,k}\}$,
 σ – уровень шума, вносимый при округлении коэффициентов,
2. критерий минимума среднеквадратической ошибки,
3. критерий минимума максимальной ошибки.

Условие пороговой обработки вейвлет-коэффициентов при сжатии методом ВП называется жесткой пороговой обработкой (hard thresholding) и имеет следующий вид (рис. 7.6).

$$d_{j,k} = \begin{cases} 0, & \text{если } |d_{j,k}| \leq T \\ d_{j,k}, & \text{если } |d_{j,k}| > T \end{cases}$$

Таким образом, коэффициенты, не превышающие пороговое значение, обнуляются и число коэффициентов ВП сигнала, участвующих в обратном преобразовании сокращается.

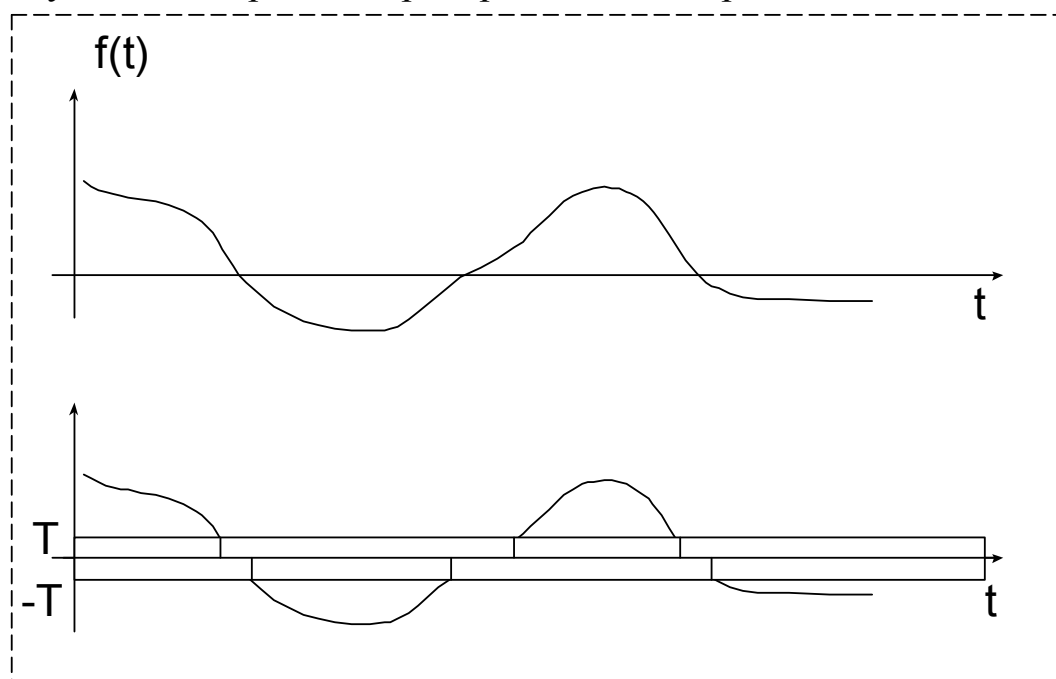


Рис. 7.6. Жесткая пороговая обработка

Коэффициенты сжатия, предельно достижимые при такой стратегии компрессии, очень высоки: правдоподобное восстановление сигнала может быть выполнено по 2 – 4 % коэффициентов, что означает коэффициент сжатия 25 : 1, 50 : 1. Объясняется этот факт тем, что при вейвлет-преобразовании в *каждом* коэффициенте содержится информация обо *всем* сигнале.

Это свойство имеет еще одно преимущество: шум округления, вносимый от квантования коэффициентов подобен белому шуму. Это особенно важно при сжатии изображений, так как человеческой моделью зрения лучше воспринимается такой шум, чем, например, шум в виде клеток, возникающий при высоких коэффициентах сжатия в стандартах JPEG, MPEG.

2. Энтропийные методы компрессии

Данные методы используют вейвлет-коэффициенты разложения, округление которых выполняется адаптивным квантователем. Исходный сигнал делится на блоки, для каждого блока определяется среднее значение и диапазон разброса путем поиска наибольшего и наименьшего в блоке. Значение среднего и величина разброса определяют ступень адаптивного квантователя. Округленные коэффициенты затем кодируются методом RLC (run length coding), который выполняет сжатие повторяющихся бит коэффициентов. После этого на основе анализа числа повторений кодовых последовательностей выполняется кодирование сигнала по Хаффману. Структура компрессора с адаптивным квантователем и блоками кодеров RLC и Хаффмана имеет следующий вид (рис. 7.7).



Рис. 7.7. Структура энтропийного компрессора

7.2.2. Удаление шумов при вейвлет-преобразовании

1. Классическая вейвлет-фильтрация

Удаление шума из сигнала всегда основывается на построении некоторой модели шума.

Обобщенную модель сигнала с шумом можно записать в виде:

$$s(t) = f(t) + \sigma \cdot e(t), \quad (7.5)$$

где $f(t)$ – полезный сигнал, $e(t)$ – белый гауссовский шум, σ – уровень шума. Фильтрация выполняется путем пороговой обработки вейвлет – коэффициентов $d_{j,k}$. Т.к. вейвлет-преобразование выполняется над сигналом $s(t)$, то в вейвлет-коэффициентах содержится информация о смеси сигнала с шумом. Таким образом, задачей фильтрации является удаление из $d_{j,k}$ шумовой составляющей, при этом полагается, что коэффициенты образующей $\{s_{N,k}\}$ относятся только к полезному сигналу и не содержат шума. Действующий в сигнале шум изменяет не только частотный состав сигнала, но и его амплитуду, особенно, если в самом сигнале есть высокочастотные компоненты. Метод, применяемый при вейвлет-фильтрации, называется гибкой пороговой обработкой (soft thresholding). Отсортированные по возрастанию $\{d_{j,k}\}$ преобразуются согласно условию гибкой пороговой обработки, которое имеет следующий вид (рис. 7.8).

$$d_{j,k} = \begin{cases} 0, & \text{если } d_{j,k} \leq T \quad (1 \text{ этап}) \\ \text{sign}(d_{j,k}) * (|d_{j,k}| - T), & \text{если } d_{j,k} > T \quad (2 \text{ этап}) \end{cases}$$

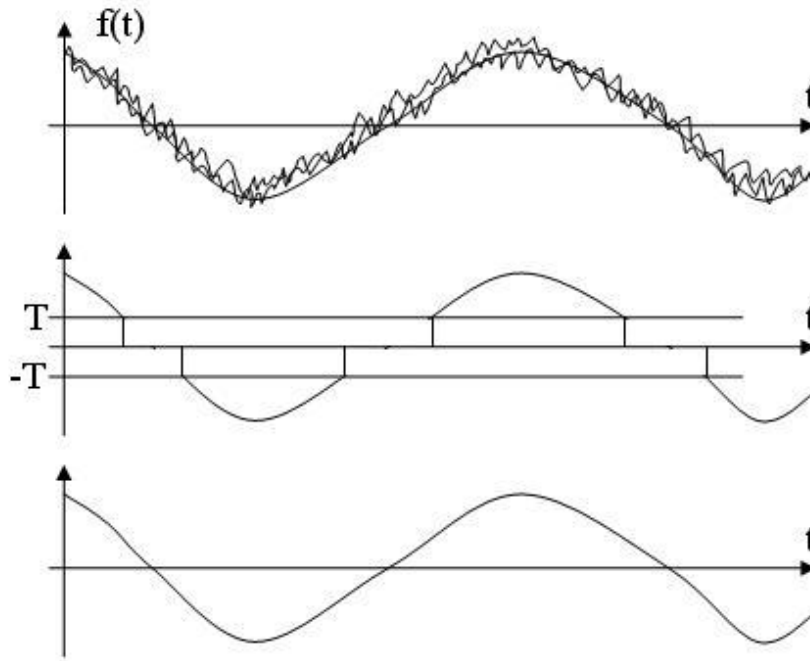


Рис. 7.8. Гибкая пороговая обработка

Если в сигнале присутствует белый шум, то его спектральная плотность одинакова. При субполосном разделении сигнала, его энергия распределяется пропорционально ширине полосы. Следовательно, энергия шума логарифмически (каждый раз вдвое) убывает с увеличением глубины разложения. Порог T_0 , определенный для самой высокочастотной полосы, должен изменяться как:

$$T_j = \frac{T_0}{2^{j/2}}, \quad (7.6)$$

где j – номер уровня разложения.

В результате преобразования из всех $d_{j,k}$ удаляется шумовая компонента, а энергия сигнала уменьшается на величину энергии удаленного шума. В том случае, если шум не белый, вклад шумовой составляющей может быть различным, и вычисление порога производят в соответствии с априорными предположениями о природе шума.

Статистические критерии выбора порога:

1. *фиксированный*: $T = (\sqrt{2 \ln N}) * \sigma$, где N – длина $\{d_{j,k}\}$, σ – уровень шума, принятый в модели, N – длина реализации сигнала.

2. критерий *минимума среднеквадратической ошибки*,

$$T: \sqrt{\sum_i \frac{(f(i) - \overline{f(i)})^2}{N-1}} \rightarrow \min, \text{ где } N - \text{длина реализации сигнала.}$$

Этот метод является экспериментальным и позволяет сравнивать значения вектора порогов, оптимизируя порог в соответствии с минимумом дисперсии (или СКО).

3. критерий *минимума максимальной ошибки*.

$$T: \sum_i \max(f(i) - \overline{f(i)}) \rightarrow \min, \text{ критерий минимума максимальной}$$

ошибки применяется в тех случаях, когда нет априорной информации о модели шума. Данный критерий определяет погрешность в наихудшем случае и является наиболее жестким критерием.

Вычислительная сложность метода пропорциональна N . Такая методика фильтрации подходит как для стационарных, так и для нестационарных сигналов. В первом случае применение классической фильтрации также дает прекрасные результаты. В случае нестационарного сигнала или наличия в нем высокочастотных компонент фильтр ВП обеспечивает лучшее подавление шума при сохранении особенностей формы сигнала.

2. Совершенствование методов вейвлет-фильтрации

Одним из путей усиления алгоритма может служить использование *непрореженных* (undecimated) вейвлет-коэффициентов. При вычислении вейвлет-преобразования выполняется только свертка сигнала со шкалирующей и вейвлет-функциями, изменение масштаба сигнала не происходит.

Вычисляется вейвлет-преобразование сигнала

$$f(t) = \sum_{j=1}^N \sum_k \psi\left(\frac{t-k}{2^j}\right) \cdot d_{j,k} + \sum_k \varphi\left(\frac{t-k}{2^N}\right) \cdot s_{N,k}, \text{ но из структуры}$$

фильтров, выполняющих эту операцию, исключены блоки прореживания и интерполяции. В результате увеличивается избыточность, и, следовательно, информативность представления

сигнала. Вычислительная сложность алгоритма составляет $N \log N$, где N – размер реализации.

В условиях, когда модель шума неизвестна или шум небелый, применяется оптимальный алгоритм фильтрации [10]. Данный алгоритм был предложен для решения задачи удаления шума, подчиняющегося распределению Пуассона, т.е. шума, пропорционального уровню сигнала. Алгоритм работает только не в реальном масштабе времени. Для оптимизации вейвлет-фильтра выполняются следующие предварительные операции.

Выбирается априорно некоторое значение порога h для гибкой обработки.

Из реализации сигнала последовательно выделяются подреализации, состоящие из всех отсчетов кроме одного.

Выполняется вейвлет-фильтрация по обычному правилу, в результате чего получают фильтрованные оценки всего сигнала $\tilde{f}^{(k)}$ без k -того отсчета.

Минимизируется ошибка оценивания:

$$e^k(h) = \left\| \frac{1}{N-1} \tilde{f}^{(k)} - f^{(k)} \right\|^2 - \text{ошибка оценивания всего сигнала}$$

по фильтрованной подреализации,

$$V(h) = \sum_{k=1}^N e^k(h) \rightarrow \min, - \text{функция ошибки оценивания.}$$

Показывается, что СКО стационарного шума с произвольным законом распределения, наложенного на сигнал $f(k)$, определяется на основании следующего вейвлет-разложения сигнала:

$$\sigma^2 = \sum_k \psi^2\left(\frac{t-k}{2^j}\right) \cdot f(k), \text{ т.е. путем свертки сигнала с квадратом}$$

вейвлет-функции на первом (самом высокочастотном) уровне разложения.

Оптимальное h по минимуму $V(h)$ имеет вид:

$$h^{opt} = \frac{d_{j,k} - \sigma^2}{d_{j,k}}, \text{ это означает, что при неизвестной природе}$$

шума порог вычисляется не глобально и не по уровням разложения как для белого шума, *для каждого коэффициента* в отдельности.

Данный алгоритм иллюстрирует условность понятия "вейвлет-фильтр", т.к. здесь неприменимы качественные характеристики фильтра как процесса. Под понятием вейвлет-фильтра подразумевается правило, позволяющее выделить с определенной точностью некоторый случайный процесс из совокупности нескольких, в общем случае, случайных процессов.

Анализ средств нелинейной фильтрации показывает, что все они носят эвристический характер. Простота вычислений и эффективность обработки способствуют расширению области их применения. Вследствие нелинейности нельзя строго разграничить влияние вейвлет-фильтрации на сигнал и шум, как нельзя дать строго качественного описания вейвлет-фильтра. Данное соображение применимо ко всем методам нелинейной фильтрации (медианной фильтрации, кепстральному анализу). Исключение составляют только методы нелинейной оптимальной марковской фильтрации.

8. ОБРАТНЫЕ ЗАДАЧИ ПРИ ИЗМЕРЕНИЯХ И УПРАВЛЕНИИ

8.1. ОБЩАЯ ЧАСТЬ

Если $x(v)$ – измеряемая величина, изменяющаяся при изменении аргумента v и представляющая собой сигнал на входе средства измерений, A – оператор, реализуемый средством измерений, $y(v)$ – отклик средства измерений на входной сигнал $x(v)$, то при отсутствии каких-либо погрешностей и помех имеет место равенство

$$Ax(v) = y(v).$$

Обратная задача, как она формулируется в математике, заключается в решении этого уравнения относительно сигнала $x(v)$, то есть в нахождении обратного оператора A^{-1} , с помощью которого можно определить $x(v)$:

$$x(v) = A^{-1}y(v).$$

Применительно к процессу измерений эта задача является прямой задачей измерений.

Обычно оператор A искажает форму сигнала $x(v)$ и известен неточно, поэтому, по образному выражению В. Г. Кнорринга, решение обратной задачи – это определение предмета при разглядывании его через неровное и кривое стекло.

В теории автоматического управления довольно часто возникает задача найти такой управляющий сигнал $x(v)$, действующий на входе регулятора (позиционера), при действии которого траектория сигнала на выходе была в точности равна заданной траектории $y(v)$. В такой постановке оператор A есть оператор, реализуемый регулятором (позиционером). Математическая формулировка этих двух задач одинакова, но по-существу эти задачи принципиально различны, поскольку первая из них отличается от второй своим ярко выраженным познавательным характером. Кроме того, если решение

неединственно, в этом случае из множества решений можно выбрать такое, которое проще реализуется.

Оператор любого физически реализуемого устройства, будь то средство измерений или регулятор, есть вполне непрерывный оператор, который отображает ограниченное множество функций в компактное множество. Оператор $\mathbf{A}^{-1}$, обратный вполне непрерывному оператору $\mathbf{A}$, не является непрерывным оператором. Для практики это обстоятельство является решающим, поскольку, во-первых, оператор $\mathbf{A}$ известен неточно, с некоторой (а именно, инструментальной) погрешностью γ , во-вторых, выходной сигнал $y(v)$ также содержит погрешности $\delta(v)$, и в этих условиях исходное операторное равенство имеет вид:

$$\mathbf{A}_\gamma x(v) = y(v) + \delta(v), \quad (8.1)$$

где $\mathbf{A}_\gamma$ – реальный оператор применяемого средства измерений.

Оператор $\mathbf{A}_\gamma$ неизвестен, известен лишь оператор $\mathbf{A}$, поэтому приходится решать уравнение

$$\mathbf{A}x(v) = y(v) + \delta(v). \quad (8.2)$$

Если обозначить решение уравнения (8.2) через $x_\delta(v)$, то поскольку обратный оператор $\mathbf{A}^{-1}$ не является непрерывным, мы не можем применить к решению этого уравнения стандартной формулировки, применяемой к непрерывным операторам, а именно, для любого сколь угодно малого наперед заданного положительного числа ε можно найти такое число δ , что если квадрат нормы погрешностей $\|\delta(v)\|^2$ будет меньше δ , то будет выполнено неравенство $\|x(v) - x_\delta(v)\|^2 \leq \varepsilon$.

Напротив, из-за того, что оператор $\mathbf{A}^{-1}$ не непрерывен, при сколь угодно малом заданном числе δ и при сколь угодно малой норме невязки $\|\mathbf{A}x(v) - y_\delta(v)\|^2$, где $y_\delta(v) = y(v) + \delta(v)$ гарантировать малость нормы погрешности решения $\|x(v) - x_\delta(v)\|^2$ невозможно. В

этом заключается математическая некорректность решения уравнения (8.2). Такие задачи называются задачами, некорректными по Адамару.

Для нахождения приближенных регулярных решений подобных задач А. Н. Тихонов [13] предложил метод, который он назвал *методом регуляризации*.

Предлагается использовать вместо оператора A другой, близкий к нему и немного модифицированный оператор A_α , такой, чтобы он и обратный ему операторы были непрерывны. Модифицировать оператор необходимо незначительно, таким образом, чтобы решение уравнения (8.2) отличалось от "точного" не слишком сильно, но в то же время не выходило за пределы области функций, гладких в некотором смысле. В качестве показателя гладкости функций А. Н. Тихонов использовал ограниченность норм их производных до k -го порядка включительно. Общая идея метода заключается в том, чтобы поиск решения заключался в минимизации *стабилизирующего функционала* [13].

$$\Omega[x(v)] = \alpha_0 \|x(v)\|^2 + \alpha_1 \|x'(v)\|^2 + \dots + \alpha_k \|x^{(k)}(v)\|,$$

ограничивающего гладкость решения, при условии, что норма невязки $\|Ax(v) - y_\delta(v)\|^2$ равна норме погрешности $\|\delta(v)\|^2$. Эта задача есть задача поиска условного экстремума (минимума), и в общем случае она решается методом неопределенных множителей Лагранжа [13]:

$$x_r(v) = \arg \min \left[\|Ax(v) - y_\delta(v)\|^2 + a_0 \|x(v)\|^2 + a_1 \|x'(v)\|^2 + \dots + a_k \|x^{(k)}(v)\|^2 \right]. \quad (8.3)$$

Минимизируемый функционал называется *сглаживающим* [13]. Коэффициенты (неопределенные множители) $\alpha_0, \alpha_1, \dots, \alpha_k$ определяют степень ограничений на нормы соответствующих производных решения.

В конкретных случаях на практике применяют одночленный стабилизирующий функционал с неравным нулю одним

коэффициентом, который называется *параметром регуляризации*. Если применяется только коэффициент (параметр регуляризации) $\alpha_p > 0$, то процесс нахождения решения называется регуляризацией p -го порядка. С увеличением параметра регуляризации решение оказывается более гладким, но при этом возрастает отличие регуляризованного решения от точного. Аналогичным образом гладкость решения зависит и от погрешности $\delta(v)$: с увеличением погрешности решение становится более гладким. Из этого следует, что значение параметра регуляризации и норма погрешности исходных данных должны быть согласованы.

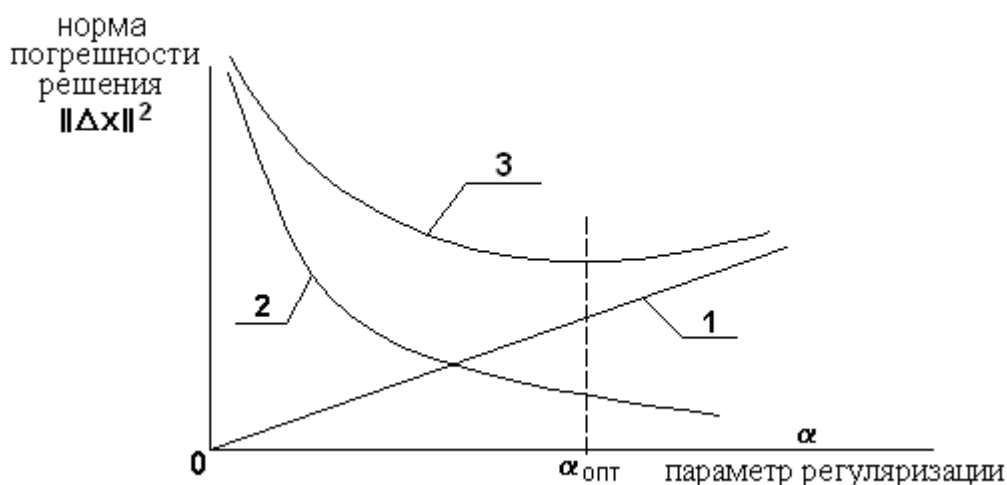


Рис. 8.1 Влияние параметра регуляризации на погрешность решения некорректной задачи

На рис. 8.1 представлены кривые, показывающие характер зависимости погрешностей регуляризованного решения, вызванные изменением параметра регуляризации: погрешности от модификации оператора (кривая 1), погрешности, вызванной некорректностью задачи (кривая 2) и общей погрешности решения (кривая 3). Видно, что существует оптимальное значение параметра регуляризации, при котором погрешность решения минимальна. Но определение этого значения весьма затруднительно.

На практике обычно применяют нулевой и первый порядки регуляризации.

Так, при нулевом порядке регуляризации, то есть при одном не равном нулю коэффициенте α_0 ограничивается только квадрат нормы на значения решения, и сглаживающий функционал имеет вид:

$$\|Ax(v) - y_\delta(v)\|^2 + \alpha_0 \|x(v)\|^2. \quad (8.4)$$

Минимум этого функционала достигается, когда его вариация равна нулю, то есть

$$2 \cdot A^* (Ax(v) - y_\delta(v)) + 2 \cdot \alpha_0 \cdot x(v) = 0, \quad (8.5)$$

откуда получаем уравнение Эйлера для нахождения регулярного решения

$$A^* Ax_r(v) + \alpha_0 x_r(v) = A^* y_\delta(v). \quad (8.6)$$

В этом случае регулярное решение находится, как

$$x_r(v) = (A^* A + \alpha_0)^{-1} A^* y_\delta(v), \quad (8.7)$$

где A^* – оператор, сопряженный оператору A .

Из (8.7) видно, что оператор A заменен на близкий ему (при малом α_0) оператор $(A^*)^{-1} (A^* A + \alpha_0) = (A + A^{*-1} \alpha_0)$. Этот оператор имеет непрерывный обратный оператор при любом $\alpha_0 > 0$.

Если имеется необходимость найти решение с ограничением на норму первой производной (порядок регуляризации $p = 1$), то отыскивается минимум сглаживающего функционала

$$\|Ax(v) - y_\delta(v)\|^2 + \alpha_1 \|x'(v)\|^2. \quad (8.8)$$

Уравнение Эйлера, обеспечивающее нулевую вариацию этого функционала, имеет вид

$$A^* Ax_r(v) - \alpha_1 x_r''(v) = A^* y_\delta(v). \quad (8.9)$$

Из этого уравнения невозможно выразить регулярное решение в конечном аналитическом виде, поскольку при интегральном операторе A оно является интегро-дифференциальным уравнением.

Коэффициенты α_0 и α_1 , как уже было упомянуто выше, именуются *параметрами регуляризации*. Выбор этих параметров в каждом конкретном случае при отсутствии априорных сведений о решении неопределен. Мало того, решение уравнения (8.9) может быть получено только, если известна априори вторая производная этого решения, что практически невозможно. Если знать априори вторую производную искомого решения, то это решение может быть получено повторным интегрированием. Операция интегрирования является математически корректной и может быть реализована на цифровых ЭВМ с высокой степенью точности.

При использовании регуляризации более высоких порядков возникают более серьезные проблемы.

Во избежание этих проблем следует действовать в соответствии с неформальным существом метода А. Н. Тихонова [13], а именно, начинать поиск решения в области очень гладких функций и проверять каждое регуляризованное таким образом решение на удовлетворение неравенству $\|Ax_r(v) - y_\delta(v)\|^2 \leq \|\delta(v)\|^2$. Если это неравенство не удовлетворяется, разрешить решению быть менее гладким и снова проверить выполнение неравенства. Приемлемым решением будет такое, при котором окажется, что $\|Ax_r(v) - y_\delta(v)\|^2 = \|\delta(v)\|^2$. Тем самым гладкость решения (то есть параметр регуляризации) и погрешность исходных данных окажутся согласованными. Подобная процедура реальна, поскольку область, в которой стабилизирующий функционал $\Omega[x(v)]$ ограничен, является монотонной [13].

Вариация гладкости решения может быть осуществлена путем изменения или параметра регуляризации, или погрешности (в зависимости от задачи).

В настоящем разделе рассмотрены изначальные общие формальные положения метода регуляризации решения некорректных задач, предложенного А. Н. Тихоновым.

В дальнейшем мы будем рассматривать практически приемлемые способы решения некоторых некорректных задач, возникающих в практике измерений и автоматического управления, при следующих условиях.

1. Оператор A есть оператор свертки. Именно такие операторы реализуются в подавляющем большинстве средств измерений и средств управления.

2. Правые части уравнений (1) и регуляризованные решения таких уравнений принадлежат функциональному нормированному пространству L_2 , норма функций в котором определяется следующим образом:

$$\|z(v)\|_{L_2}^2 = \int_a^b [z(v)]^2 dv.$$

В дальнейшем индекс L_2 мы будем опускать.

8.2. ФОРМУЛИРОВКА МЕТОДА РЕГУЛЯРИЗАЦИИ А.Н. ТИХОНОВА В ПРОСТРАНСТВЕ ФУРЬЕ- ОБРАЗОВ

Если оператор A есть оператор свертки, то исходное точное равенство записывается в виде

$$y(v) = \int_{v_0}^v h(v-u)x(u)du \quad \text{или} \quad y(v) = \int_{-\infty}^{+\infty} h(v-u)x(u)du.$$

Первое из этих равенств описывает идеальное линейное физически реализуемое средство измерений или регулятор, второе описывает, например, спектрометрические, хроматографические и тому подобные процессы, плотности распределения суммы независимых случайных величин. Разностное ядро приведенных операторов будем называть аппаратной функцией соответствующего устройства. Преобразование Фурье свертки есть произведение преобразований Фурье участников свертки, то есть

$$\Phi y(v) = y(j\omega) = \Phi h(v)\Phi x(v) = H(j\omega)x(j\omega), \quad (8.10)$$

где $H(j\omega) = \Phi h(\nu)$ – преобразование Фурье аппаратной функции, для средств измерений и средств регулирования $H(j\omega)$ есть комплексная частотная характеристика (КЧХ) соответствующего средства.

Фурье-образ уравнения (8.1) имеет вид

$$H_\gamma(j\omega) \cdot x(j\omega) = y(j\omega) + \delta(j\omega), \quad (8.11)$$

где $H_\gamma(j\omega)$ – Фурье-образ ядра реального оператора (для средств измерений – реальная КЧХ), отличающийся от Фурье-образа ядра оператора A (для средств измерений – от нормированной КЧХ) $H(j\omega)$ не более, чем на величину погрешности $\gamma(j\omega)$.

Преобразование Фурье ядра сопряженного оператора A^* есть $H(-j\omega)$, поэтому, например, выражение (8.7) в пространстве Фурье-образов записывается следующим образом:

$$x_r(j\omega) = \frac{H(-j\omega)}{|H(j\omega)|^2 + \alpha_0} y_\delta(j\omega) \quad (8.12)$$

Выполним преобразование Фурье равенства (8.9). Учитывая, что в пространстве Фурье-образов производная выражается множителем ω , получим:

$$|H(j\omega)|^2 x_r(j\omega) - \alpha_1 \cdot \omega^2 x_r(j\omega) = H(-j\omega) \cdot y_\delta(j\omega),$$

откуда в этом случае можно получить Фурье-образ регуляризованного решения:

$$x_r(j\omega) = \frac{H(-j\omega)}{|H(j\omega)|^2 + \alpha_1 \cdot \omega^2} y_\delta(j\omega) \quad (8.13)$$

Сейчас видно, что при наших условиях в общем случае можно получить Фурье-образ регуляризованного решения, гладкость которого ограничена всеми k производными:

$$x_r(j\omega) = \frac{H(-j\omega)}{|H(j\omega)|^2 + \alpha_0 + \alpha_1 \cdot \omega^2 + \alpha_2 \cdot \omega^4 + \dots + \alpha_k \cdot \omega^{2k}} y_\delta(j\omega) \quad (8.14)$$

Еще раз заметим, что на практике прибегают к регуляризации p -го порядка, полагая, что отличен от нуля только один коэффициент α_p знаменателя (8.14).

Последнее равенство имеет место по той причине, что в соответствии с теоремой Парсеваля

$$\Phi \left[\int_a^b [z(v)]^2 dv \right] = K \int_{-\infty}^{\infty} |z(j\omega)|^2 d\omega$$

Поэтому, учитывая, что одинаковый постоянный коэффициент у всех слагаемых функционала (8.3) не влияет на положение минимума, получим выражение для регуляризованного решения в пространстве Фурье-образов:

$$x_r(j\omega) = \arg \min_{x(j\omega)} \left[\int_{-\infty}^{\infty} |H(j\omega)x(j\omega) - y_\delta(j\omega)|^2 d\omega + \alpha_0 \int_{-\infty}^{\infty} |x(j\omega)|^2 d\omega + \right. \\ \left. + \alpha_1 \int_{-\infty}^{\infty} |x(j\omega)|^2 d\omega + \alpha_2 \omega^4 \int_{-\infty}^{\infty} |x(j\omega)|^2 d\omega + \dots + \alpha_k \omega^{2k} \int_{-\infty}^{\infty} |x(j\omega)|^2 d\omega \right], \quad (8.15)$$

откуда следует (8.14).

Обратим внимание на то обстоятельство, что выражения (8.12) – (8.14) можно непосредственно использовать в тех случаях, когда исходное уравнение в конкретной задаче записано в виде

$$y(v) = \int_{-\infty}^{+\infty} h(v-u)x(u)du.$$

Такие задачи возникают при желании восстановить спектр или хроматограмму, освободив полученную запись этих функций от искажений, внесенных аппаратными функциями применяемых спектрометров или хроматографов. Подобную задачу сформулировал в свое время Рэлей, назвав ее *редукцией к идеальному прибору*.

Для решения задачи, которая формулируется, как решение уравнения

$$y(v) = \int_{v_0}^v h(v-u)x(u)du \quad \text{или} \quad y(v) = \int_{-\infty}^v h(v-u)x(u)du,$$

непосредственное применение формул (8.12) – (8.14) в реальном времени непригодно, поскольку использование в числителе аргумента $(-j\omega)$ соответствует в пространстве преобразов Фурье отрицательным значениям аргумента ядра $h(u)$, что невозможно. Подобные задачи возникают при решении обратных задач при измерениях или управлении, когда ядро $h(u)$ есть импульсная переходная характеристика (весовая функция) применяемого технического средства.

Общее замечание к настоящему пункту.

Выше приведены формулы для нахождения регуляризованного решения в пространстве Фурье-образов. Поэтому для получения окончательного решения во всех приведенных случаях необходимо применить к $x_r(j\omega)$ обратное преобразование Фурье. После этого может быть вычислена норма невязки $\|Ax_r(v) - y_\delta(v)\|^2$ и проверена гипотеза

$$\|Ax_r(v) - y_\delta(v)\|^2 \leq \|\delta(v)\|^2.$$

Имея в виду, что эта гипотеза может быть отвергнута как при слишком больших, так и при слишком малых значениях параметра регуляризации, следует воспользоваться монотонностью области, в которой ограничен функционал $\Omega[x(v)]$. Начинать процедуру необходимо с самых гладких решений, то есть при достаточно больших значениях параметра регуляризации и уменьшать его значение не слишком быстро, особенно вблизи окончательного решения. Так необходимо поступать, чтобы не переходить границу критического значения слишком далеко. Тем самым будут согласованы значения параметра регуляризации и погрешность исходных данных. При некорректности задачи чересчур малое значение невязки грозит большим отличием регуляризованного решения от "истинного".

8.3. ПРИНЦИП МИНИМАЛЬНОГО МОДУЛЯ

В разделе 8.1 изложена общая идея метода регуляризации, а именно, искать решение $x_r(v)$ путем минимизации *стабилизирующего функционала*

$$\Omega[x(v)] = \alpha_0 \|x(v)\|^2 + \alpha_1 \|x'(v)\|^2 + \dots + \alpha_k \|x^{(k)}(v)\|,$$

ограничивающего гладкость решения, при условии, что норма невязки $\|Ax_r(v) - y_\delta(v)\|^2$ равна норме погрешности $\|\delta(v)\|^2$.

В пространстве Фурье-образов функционал $\Omega[x(v)]$ отображается в функционал

$$\begin{aligned} \Omega[x(j\omega)] &= \alpha_0 \int_{-\infty}^{\infty} |x(j\omega)|^2 d\omega + \alpha_1 \int_{-\infty}^{\infty} |\omega \cdot x(j\omega)|^2 d\omega + \alpha_2 \int_{-\infty}^{\infty} |\omega^2 x(j\omega)|^2 d\omega + \dots + \alpha_k \int_{-\infty}^{\infty} |\omega^{2k} x(j\omega)|^2 d\omega = \\ &= \int_{-\infty}^{\infty} |x(j\omega)|^2 d\omega \cdot [\alpha_0 + \alpha_1 \omega^2 + \alpha_2 \omega^4 + \dots + \alpha_p \omega^{2p}] \end{aligned} \quad (8.16)$$

Мы видим, что минимизация функционала (8.16) по $x(j\omega)$ с неизбежностью влечет за собой только минимизацию модуля $|x(j\omega)|$.

Поскольку мы отыскиваем Фурье-образ регуляризованного решения путем минимизации (8.15) по $x(j\omega)$, то можно надеяться на то, что с учетом погрешности оператора $\mathbf{A}$ (пересчитанной в пространство Фурье-образов) регуляризованное решение можно отыскать, обеспечивая минимальность его модуля при нахождении его из (8.11) непосредственным делением.

В самом деле, из (8.11) можно получить:

$$x_\delta(j\omega) = \frac{y_\delta(j\omega)}{H_\gamma(j\omega)} \quad (8.17)$$

Однако, в силу некорректности задачи, существования погрешностей $\gamma(j\omega)$ и $\delta(j\omega)$, пересчитанных в пространство Фурье-образов, а также отсутствия непрерывности оператора $\mathbf{A}$ полученный в результате Фурье-образ, как и его прообраз $x(v)$, вряд ли будут иметь физический смысл, соответствующий задаче.

Во избежание этого неудобства предлагается выполнять деление (8.17) так, чтобы при каждом значении ω получить результат, обладающий минимальным модулем из всех возможных результатов, порождаемых числителем и знаменателем (8.17), каждый из которых при этом же значении ω может находиться в любом месте своих областей, ограниченных предельными погрешностями $\gamma(j\omega)$ и $\delta(j\omega)$. Рис. 8.2 иллюстрирует высказанное предложение при некотором (любом) значении ω_i .

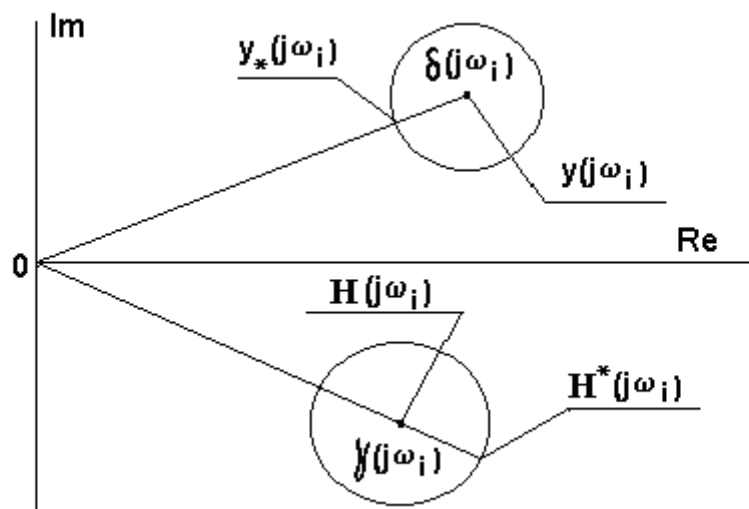


Рис. 8.2. Иллюстрация принципа минимального модуля

Для обеспечения минимального модуля Фурье-образа решения при некотором значении ω_i предлагается в качестве числителя брать значение $y_*(j\omega_i)$, которое соответствует значению правой части, лежащему на границе зоны погрешности $\delta(j\omega_i)$ в точке, ближайшей к началу координат комплексной плоскости. В качестве знаменателя при том же значении ω_i предлагается брать значение $H^*(j\omega_i)$, которое соответствует значению Фурье-образа ядра, лежащему на границе зоны погрешности $\gamma(j\omega_i)$ в точке, наиболее удаленной от начала координат комплексной плоскости. Таким образом, при каждом заданном значении ω_i должно вычисляться частное

$$x_{min}(j\omega) = \frac{y^*(j\omega)}{H^*(j\omega)},$$

обладающее минимальным модулем.

Если такое деление выполнять во всех точках аргумента ω , то полученный результат в целом будет обладать наименьшим модулем при выполнении условий нахождения числителя и знаменателя, каждого из них, в своей зоне погрешностей. При этом знаменатель никогда не будет равным нулю. Числитель окажется равным нулю тогда, когда его значение не будет превышать погрешности, с которой он известен.

Ясно, что при увеличении предела допускаемой погрешности $\delta(j\omega)$ условия гладкости будут ужесточаться, получающийся Фурье-образ решения будет иметь меньший модуль, а значит, решение будет более гладким.

Применение принципа минимального модуля эффективно для решения задач в следующих условиях.

1. Известен вид аппаратной функции $h(u)$, но неизвестны ее параметры. Характеристики погрешности $\delta(v)$ недостаточно известны, известны лишь оценки погрешности сверху.

2. Известна аппаратная функция и предельные значения ее погрешности $\gamma(v)$. Характеристики погрешности $\delta(v)$ недостаточно известны, известны лишь оценки погрешности сверху.

Рассмотрим решение обратной задачи в хроматографии (метод разделения компонентов смесей веществ). Точность результатов количественного химического анализа многокомпонентных смесей существенно снижается при пересечении близких пиков хроматограммы (график зависимости изменения концентрации компонентов смеси или пропорциональной ей величины от длительности процесса разделения). Такие пересечения происходят вследствие недостаточной разрешающей способности применяемых приборов, повысить которую можно двояко – как за счет усовершенствования хроматографических колонок, детекторов, так и

применяя математические методы разделения пиков. Математически это означает, что зарегистрированная хроматограмма есть результат свертки аппаратной функции применяемого прибора и реальной характеристики исследуемого вещества. Поскольку оператор свертки есть линейный интегральный оператор с разностным ядром, для разделения пересекающихся пиков и тем самым для повышения разрешающей способности этих приборов достаточно решить интегральное уравнение относительно истинной характеристики. Однако, эта задача не является математически корректной, и для получения физически осмысленного решения необходимо применять методы регуляризации, которые для общего случая разработаны академиком А. Н. Тихоновым. Но в данном случае задача осложняется тем, что аппаратная функция хроматографа неизвестна, и поэтому необходимо модифицировать общие методы. В качестве путей модификации методов А. Н. Тихонова применительно к хроматографии могут быть применены нейронные сети и вейвлет-анализ.

Первая задача имеет место при попытках исключить искажения, вносимые аппаратной функцией спектрометра или хроматографа в зарегистрированный результат. Эти искажения приводят к пересечению спектроскопических или хроматографических пиков, из-за чего погрешность определения концентраций отдельных компонентов определяется с большой погрешностью, достигающей 50 % и более.

В этом случае задача решается следующим образом.

На первом шаге задаются наиболее широкой аппаратной функцией из возможного диапазона и наибольшей погрешностью. Аппаратная функция и погрешность преобразуются по Фурье и выполняется деление в соответствии с изложенным принципом минимального модуля. Результат деления подвергается обратному преобразованию Фурье и вычисляется невязка. Для сравнения невязки с погрешностью проверяется гипотеза о соотношении между невязкой

и погрешностью $\delta(v)$. Если в результате сравнения оказывается, что невязка больше, уменьшается ширина аппаратной функции и (или) погрешность, то процедура повторяется. Тем самым на каждом следующем шаге гладкость области решений должна монотонно ухудшаться. Остановка процесса может произойти по признаку равенства невязки и погрешности или по виду полученного решения (решение не должно быть меньше нуля, не должно возникать ложных пиков и других признаков потери физического смысла). Суждение о нарушении физического смысла принимает опытный исследователь.

Вторая задача может иметь место также при спектроскопических или хроматографических исследованиях. Но можно себе представить с такими же условиями и иные задачи, например, задачу устранения динамических искажений при прохождении сигнала через динамическое звено (средство измерений или регулятор). В этом случае последовательность действий примерно такая же. Но вместо вариаций аппаратной функции приходится монотонно уменьшать погрешности числителя и знаменателя, обеспечивая тем самым на каждом шаге устойчивость решения. Остановка этого процесса выполняется так же, как и в первом случае.

В качестве примера применения принципа минимального модуля на рис. 8.3 приведен результат решения модельной задачи, характерной для спектроскопических исследований, где аппаратная функция является симметричной и существует при положительных и отрицательных значениях аргумента. При решении задачи было принято, что аппаратная функция имеет форму гауссианы, но ее ширина, как и погрешность регистрации спектрограммы, неизвестны. Кривая 1 – смоделированная (истинная) кривая, кривая 2 – свертка с аппаратной функцией

$$k(v) = \frac{8,5}{1 + 72,25v^2},$$

то есть модель зарегистрированного выходного сигнала, кривая 3 – решение обратной задачи при следующих значениях дисперсии погрешностей: дисперсия погрешности регистрации $\sigma^2 = 10^{-4} e \partial^2$., дисперсия погрешности, с которой известна аппаратная функция $\sigma_k^2 = 10^{-7}$.

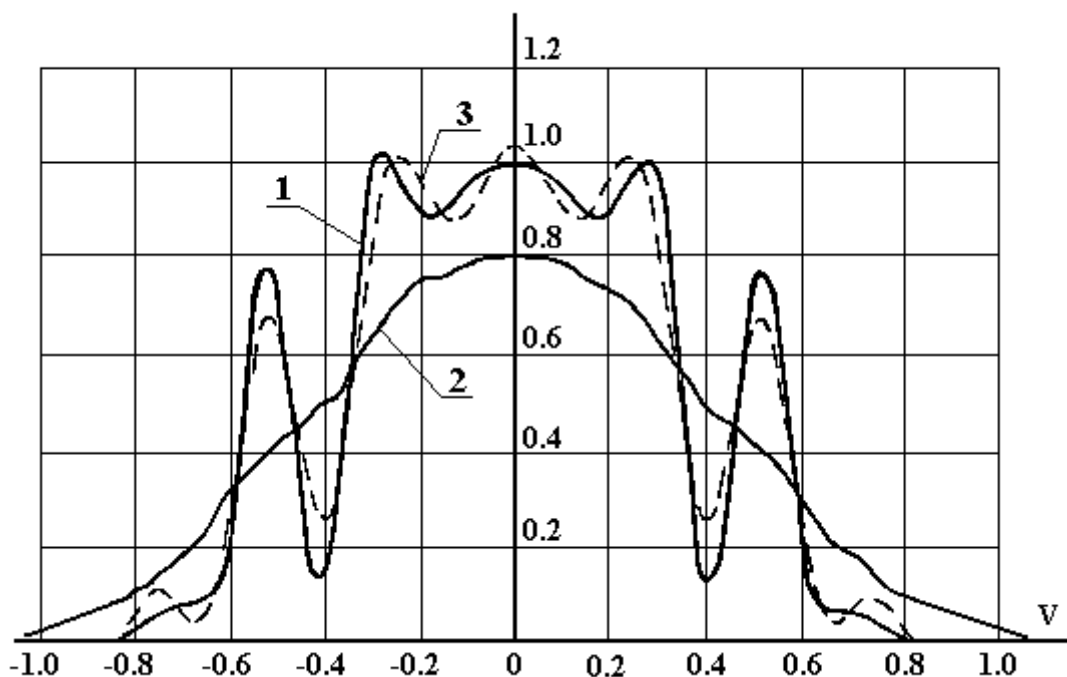


Рис. 8.3. Пример решения обратной задачи при спектрометрических исследованиях

8.4. НЕЙРОСЕТЕВОЙ ПОДХОД РЕШЕНИЯ ОБРАТНОЙ ЗАДАЧИ

Физико-химический процесс сорбции и элюирования химических компонентов в хроматографии, как и аппаратную функцию хроматографа принято описывать гауссианами. В зарегистрированной хроматограмме пики расширяются за счет действия аппаратной функции хроматографа, которая не может быть идеальной. Это действие есть результат свертки фактической хроматограммы с аппаратной функцией. На рис. 8.4 приведен пример хроматограммы взрывчатых веществ.

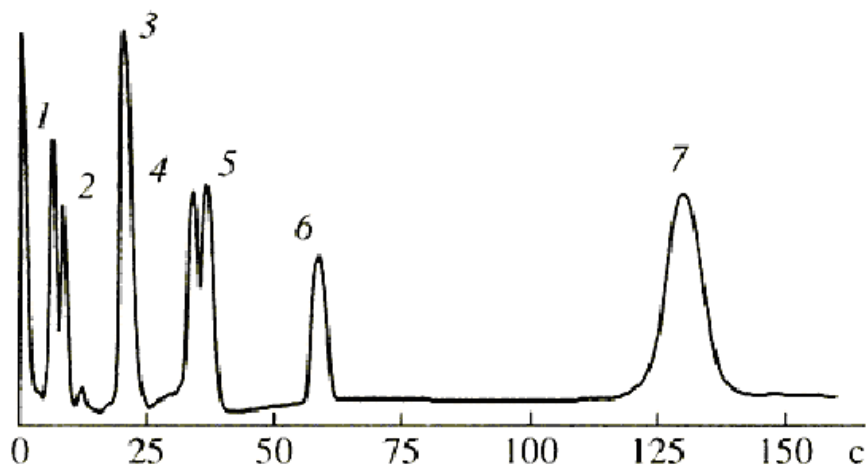


Рис. 8.4. Скоростной анализ паров взрывчатых веществ на поликапиллярной колонке при температуре 170°C.

Поликапиллярная колонка длиной всего 22 см позволяет за 2,5 минуты обнаружить и идентифицировать следовые количества паров взрывчатых веществ:

1 – 2,6-динитротолуол, 2 – 2,4-динитротолуол, 3 – 2,4,6-тринитротолуол, 4 – 3,4,5-тринитротолуол, 5 – 2,3,4-тринитротолуол, 6 – гексоген, 7 – тетрил

На вход измерителя приходит сигнал $x(t)$, на выходе получаем сглаженный сигнал $y(\tau) = \int_{-\infty}^{\infty} K(\tau - t)x(t)dt$, где известная функция $K(\tau - t)$ характеризует измерительное устройство. Задача состоит в том, чтобы по $y(\tau)$ восстановить $x(t)$.

Представим исходный сигнал в виде нейросетевого разложения $x(t) = \sum_{i=1}^N c_i v(t, a_i)$, где v – базисная функция, c_i и a_i – искомые параметры (веса сети). Тогда для выходного сигнала получаем другое

нейросетевое разложение $y(\tau) = \sum_{i=1}^N \tilde{c}_i \tilde{v}(\tau, \tilde{a}_i)$, где

$\tilde{c}_i \tilde{v}(\tau, \tilde{a}_i) = c_i \int_{-\infty}^{\infty} K(\tau - t)v(t, a_i)dt$. Так как выходной сигнал считается

известным, мы можем по нему определить параметры $\tilde{c}_i$ и $\tilde{a}_i$, обучая соответствующую нейронную сеть, а далее по этим параметрам восстановить c_i и a_i , тем самым определяя входной сигнал $x(t)$. Исходная задача является некорректной и её регуляризация осуществляется обычным для нейросетевого подхода образом – ограничением на параметры $\tilde{a}_i$ при обучении сети для выходного сигнала.

Так как входной сигнал имеет вид локализованных всплесков, для его аппроксимации разумно воспользоваться сетью радиальных базисных функций – RBF (radial basis function).

Аппаратная функция имеет вид $K(s) = \alpha \cdot e^{-\beta^2 s^2}$. В этом случае в качестве базисной функции разумно также использовать гауссиан $v(t, a_i) = e^{-a_i^2 (t-t_i)^2}$. В этом случае функция $\tilde{v}$ имеет такой же вид, а коэффициенты c пересчитываются в соответствии с формулами

$$a_i = \frac{\tilde{a}_i \beta}{\sqrt{\beta^2 - \tilde{a}_i^2}}, \quad t_i = \tilde{t}_i, \quad c_i = \frac{\tilde{c}_i \sqrt{a_i^2 + \beta^2}}{\sqrt{\pi}}. \quad \text{При этом должно}$$

выполняться условие $\beta > \tilde{a}_i$, а для регуляризации параметр $\tilde{a}_i$, также как и в предыдущем случае, должен быть отделён от β некоторой фиксированной величиной.

Реально задан не сам сигнал на выходе измерителя, а его дискретизация $y\left(jT/n\right)$, связанная с входным сигналом равенством

$$y\left(jT/n\right) = \frac{T}{n} \sum_{i=1}^n K\left(j - i/n\right) x\left(iT/n\right). \quad \text{Обучение сети для выходного}$$

сигнала проводим минимизацией функционала

$$J = \sum_{j=1}^n \left(y\left(jT/n\right) - \sum_{i=1}^N \tilde{c}_i v_i\left(jT/n, \tilde{a}_i\right) \right)^2, \quad \text{далее вычисляем параметры}$$

нейронной сети для входного сигнала по приведённым выше формулам и получаем аппроксимацию для входного сигнала [15].

Однако аппаратная функция хроматографа обычно неизвестна. Для этого случая предлагается следующая рабочая процедура.

1. По экспериментальным точкам зарегистрированной хроматограммы определяются положения пиков, их амплитуды и характеристики их ширины. Выполняется аппроксимация хроматограммы.

2. Задается очень узкая аппаратная функция, характеристика ширины которой β в несколько раз больше, чем самое большое значение из характеристик $\tilde{a}_i$. По приведенным выше формулам вычисляются новые значения c_i и a_i , по этим значениям строится функция

$$x(t) = \sum_{i=1}^N c_i v(t, \mathbf{a}_i). \quad (8.18)$$

3. Для косвенного контроля результата вычисляется свертка функции (8.18) с аппаратной функцией и результат сопоставляется с зарегистрированной хроматограммой.

4. После выполнения пп. 2, 3 аппаратная функция немного расширяется (уменьшается β) и эти пункты повторяются вновь.

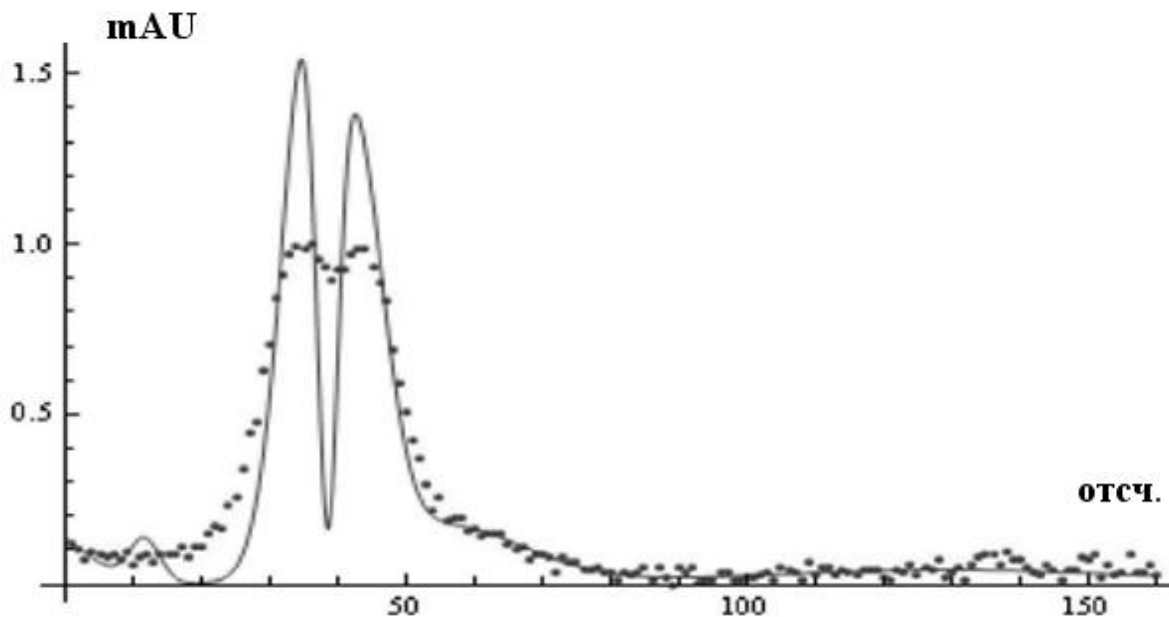


Рис. 8.5. Измеренный (точки) и восстановленный (сплошной) сигнал

5. При повторении пп. 2, 3 может возникнуть опасная ситуация, которая заключается в том, что значение аппаратной функции, уменьшаясь, будет приближаться к одному из значений a_i . В результате амплитуда соответствующего пика будет резко возрастать, а его ширина – уменьшаться. В этом случае этот пик можно удалить и выполнить п. 3 для оставшихся пиков. В качестве критерия удаления пика можно использовать:

- возрастание пика до высоты самого высокого,
- сужение его до 0.1 значения ширины самого узкого из остальных,
- достижение характеристики β величины $(1.1 \div 1.05) a_i$.

Выбор того или иного критерия придется делать на основе попыток.

Результаты численных экспериментов восстановления хроматограмм приведены на рис. 8.5 и 8.6.

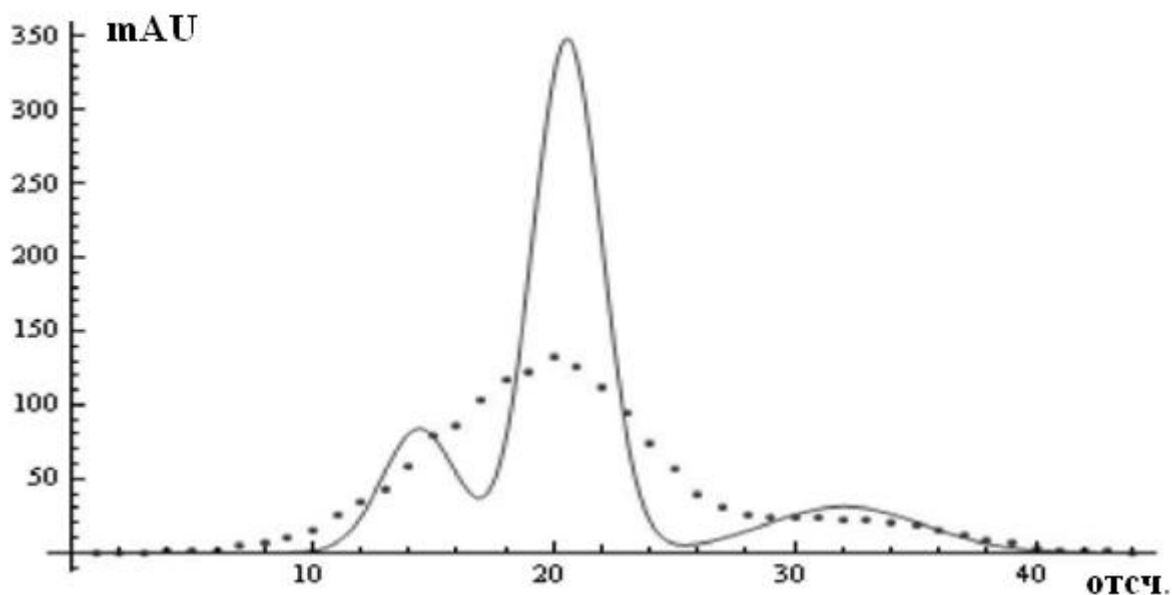


Рис. 8.6. Измеренный (точки) и восстановленный (сплошной) сигнал

8.5. МОДИФИЦИРОВАННЫЙ ОПТИМАЛЬНЫЙ ФИЛЬТР ВИНЕРА ДАЕТ РЕГУЛЯРИЗОВАННОЕ РЕШЕНИЕ

Классический оптимальный винеровский фильтр предназначен для выделения полезного сигнала из его смеси со случайным процессом (помехой). При этом предполагается, что и сигнал $x(t)$ и помеха $\delta(t)$ суть стационарные случайные процессы, для которых определены и известны спектральные плотности $S_{xx}(\omega)$ и $S_{\delta\delta}(\omega)$, соответственно.

Здесь мы рассмотрим модификацию фильтра Винера, которая заключается в том, что мы будем отыскивать сигнал $x(t)$, претерпевший преобразование линейным оператором свертки A , результат которого смешан со случайной помехой $\delta(t)$ в соответствии с равенством (см. также раздел 8.1):

$$Ax(t) = y(t) + \delta(t),$$

где помеха представляет собой случайную погрешность регистрации (измерения) выходного сигнала или реализации выходного сигнала (траектории) регулятора. Такой фильтр мы вправе назвать *обратным фильтром* Винера.

В частотной области (в области Фурье-образов) это равенство преобразуется:

$$K(j\omega) \cdot x(j\omega) = y(j\omega) + \delta(j\omega), \quad (8.19)$$

где $K(j\omega)$ – комплексная частотная характеристика (КЧХ) средства измерения или регулятора, которая на практике не бывает известной точно. Однако, погрешность, с которой известна КЧХ, нецелесообразно описывать случайным процессом, поскольку эта погрешность носит, как правило, систематический характер. К учету погрешности, с которой известна КЧХ, мы вернемся позже.

Фильтр Винера строится по критерию минимальной дисперсии погрешности восстановления сигнала $x(t)$. Полагая, что сигнал $x(t)$ и помеха $\delta(t)$ суть независимые случайные процессы, запишем

выражение для дисперсии погрешности $\Delta x(t)$ восстановления сигнала $x(t)$ в следующем виде:

$$D[\Delta x(t)] = \frac{1}{2\pi} \int_{-\infty}^{\infty} |H(j\omega)K(j\omega) - 1|^2 S_{xx}(\omega) d\omega + \frac{1}{2\pi} \int_{-\infty}^{\infty} |H(j\omega)|^2 S_{\delta\delta}(\omega) d\omega. \quad (8.20)$$

Данный функционал имеет один минимум, который можно отыскать, приравняв нулю его вариацию по $H(j\omega)$. Поскольку подинтегральные выражения больше нуля, вариация этого функционала равна нулю тогда, когда равна нулю вариация подинтегрального выражения:

$$2|K(j\omega)|^2 H(j\omega)S_{xx}(\omega) + K(-j\omega)S_{xx}(\omega) + 2H(j\omega)S_{\delta\delta}(\omega) = 0, \quad (8.21)$$

откуда КЧХ оптимального фильтра

$$H(j\omega) = \frac{K(-j\omega)}{|K(j\omega)|^2 + S_{\delta\delta}(\omega)/S_{xx}(\omega)}. \quad (8.22)$$

Отсюда получаем Фурье-образ решения (см. также [13]):

$$x_r(j\omega) = \frac{K(-j\omega)}{|K(j\omega)|^2 + S_{\delta\delta}(\omega)/S_{xx}(\omega)} \cdot y_{\delta}(j\omega), \quad (8.23)$$

где $y_{\delta}(j\omega) = y(j\omega) + \delta(j\omega)$ – Фурье-образ сигнала, зарегистрированного на выходе средства измерений или сигнал на выходе регулятора.

Выражение (8.23) очень похоже на формулы (8.12) – (8.14) и в соответствии с материалом п. 8.2 представляет собой Фурье-образ регуляризованного решения, если при увеличении частоты ω степень убывания спектральной плотности сигнала $S_{xx}(\omega)$ больше, чем степень убывания спектральной плотности помехи (случайной погрешности) $S_{\delta\delta}(\omega)$. Порядок регуляризации в этом случае будет равен полуразности степеней убывания спектральных плотностей $S_{xx}(\omega)$ и $S_{\delta\delta}(\omega)$.

Теперь поговорим об учете погрешности, с которой известна КЧХ $K(j\omega)$. Пусть известны допускаемые границы, в которых может находиться модуль КЧХ (то есть амплитудная характеристика АЧХ) реального устройства, допущенного к эксплуатации. Тогда для тех значений частот, на которых вычисляется (8.23), следует брать: в числителе – нижнее значение КЧХ в установленных границах, в знаменателе – верхнее значение КЧХ в установленных границах.

Что касается практического применения обратного фильтра Винера, то ему присущи те же недостатки, которые отмечены в разделе 8.2, а именно, КЧХ с отрицательным значением аргумента, стоящей в числителе (8.23), не соответствует физически реализуемое техническое устройство. Обратная фильтрация в соответствии с (8.23) возможна лишь после того, как в распоряжении исследователя оказывается вся реализация процесса $y_\delta(j\omega)$. В практике систем управления и вообще в реальном времени такая ситуация вовсе недопустима в силу ее бессмысленности.

Подобный подход к решению обратной задачи, основанный на винеровской оптимальной фильтрации, может быть распространен и на такие типичные случаи измерений, когда на входе инерционного средства измерений кратковременно действует измеряемая величина, ее сигнал искажается инерционным средством измерений и этот сигнал желательно восстановить. Примером таких измерений может служить, например, измерения мгновенных значений давления внутри цилиндра двигателя внутреннего сгорания во время взрыва смеси. В подобных случаях сигнал измеряемой величины не является ни случайным, ни стационарным. Для таких случаев придется использовать:

1. вместо дисперсии – математическое ожидание квадрата погрешности $M[\Delta x^2(t)]$,
2. вместо спектральной плотности сигнала измеряемой величины – математическое ожидание квадрата модуля ее спектра

(Фурье-образа), вычисленного на конечном интервале времени измерения T $S_{xx}^T(\omega) = M[x^T(j\omega) \cdot x^T(-j\omega)]$,

3. вместо спектральной плотности погрешности измерения выходного сигнала – математическое ожидание квадрата модуля его спектра (Фурье-образа), вычисленного на конечном интервале времени измерения T $S_{\delta\delta}^T(\omega) = M[\delta^T(j\omega) \cdot \delta^T(-j\omega)]$.

В этих условиях выражение (8.20) может быть переписано в виде:

$$M[\Delta x(t)] = \frac{1}{2\pi} \int_{-\infty}^{\infty} |H(j\omega)K(j\omega) - 1|^2 S_{xx}(\omega) d\omega + \frac{1}{2\pi} \int_{-\infty}^{\infty} |H(j\omega)|^2 S_{\delta\delta}(\omega) d\omega. \quad (8.24)$$

Вариация этого функционала приравняется нулю

$$2|K(j\omega)|^2 H(j\omega)S_{xx}^T(\omega) + K(-j\omega)S_{xx}^T(\omega) + 2H(j\omega)S_{\delta\delta}^T(\omega) = 0,$$

откуда получаем КЧХ оптимального фильтра

$$H(j\omega) = \frac{K(-j\omega)}{|K(j\omega)|^2 + S_{\delta\delta}^T(\omega) / S_{xx}^T(\omega)}. \quad (8.25)$$

В этой интерпретации винеровского обратного фильтра была выполнена задача восстановления давления взрыва паров дизельного топлива в цилиндре дизельного двигателя трактора «Кировец». Измерения выполнялись с помощью пьезоэлектрического датчика фирмы Kistler, давление к которому из цилиндра передавалось по трубке длиной 13 см и диаметром 5 мм. Трубка была приблизительно представлена, как инерционное звено первого порядка, а КЧХ датчика определена экспериментально. Общая КЧХ их последовательного соединения, считающаяся номинальной, была приблизительно с точностью до масштабного множителя представлена в виде

$$K_{ном}(j\omega) = \frac{1}{1 + 6,5 \cdot 10^{-3} j\omega - 10^{-8} \omega^2}.$$

Из-за погрешностей вычислений были определены нижняя и верхняя предельные границы КЧХ:

$$\text{нижняя граница } K_n(j\omega) = \frac{1}{1 + 5,5j\omega - 10^{-8}\omega^2},$$

$$\text{верхняя граница } K_B(j\omega) = \frac{1}{1 + 6,5j\omega - 10^{-8}\omega^2}.$$

Время измерений $T = 2 \cdot 10^{-3} \text{ с}$, что соответствует верхней частоте $\omega_B = \pi / T = 1,57 \cdot 10^3 \text{ рад/с}$. Спектральная плотность погрешности регистрации на интервале частот $[0, \omega_B)$ считается постоянной, не превышающей $S_{\delta\delta}(\omega) \leq 8 \cdot 10^{-7} \text{ МПа}^2 \text{ с / рад}$, что соответствует дисперсии примерно $\sigma_{\delta}^2 = 10^{-2} \text{ МПа}^2$. Величина $S_{xx}^T(\omega)$ рассчитывалась по теоретической кривой.

Результаты представлены на рис. 8.7. Кривая 1 на этом рисунке – теоретическая кривая давления, рассчитанная дизелистами по размерам объема сгорания и качеству топлива. Кривая 2 – результат измерения давления, зарегистрированный на выходе датчика. Кривая 3 – результат численного апостериорного восстановления давления в соответствии с (8.25) при имеющейся информации о спектральной плотности погрешности $S_{\delta\delta}(\omega)$, о характеристике сигнала $S_{xx}^T(\omega)$ и при номинальном значении КЧХ. Кривая 4 – результат восстановления давления при неполных исходных данных по формуле

$$x_r(j\omega) = \frac{K_H(-j\omega)}{|K_B(j\omega)|^2 + S_{\delta\delta}(\omega) / S_{xx}^T(\omega)},$$

где для вычисления $S_{xx}^T(\omega)$ использовано преобразование Фурье от прямоугольного сигнала длительностью T и высотой 20 МПа.

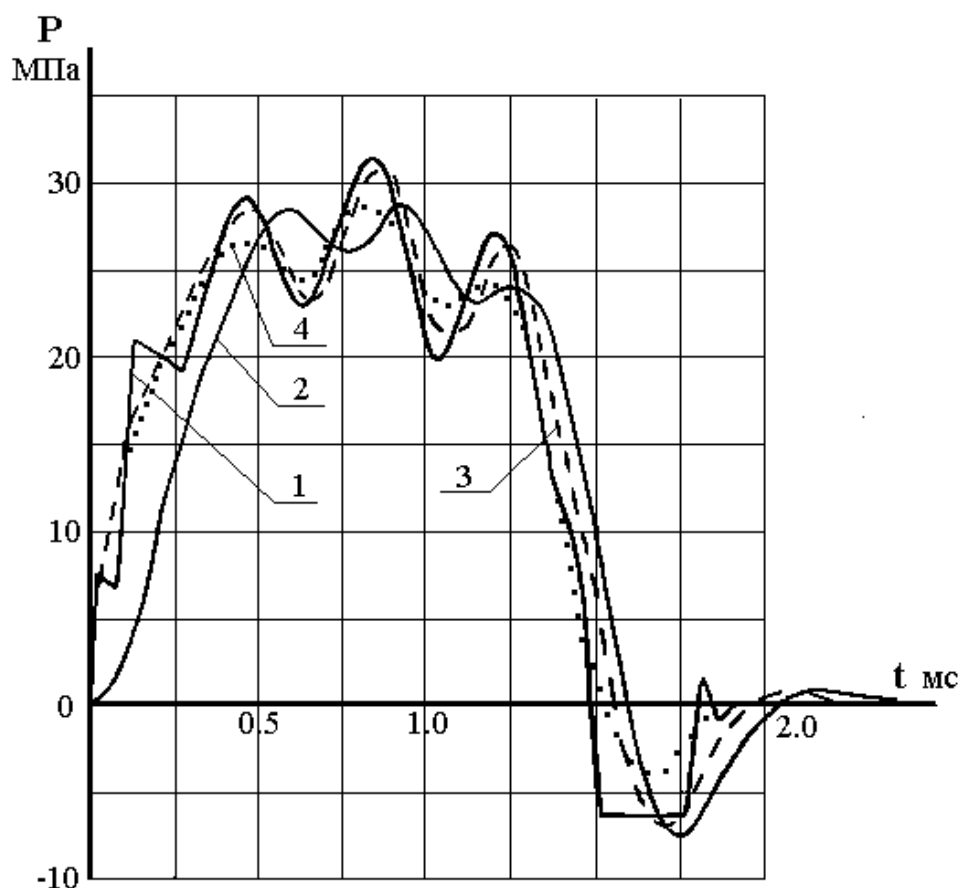


Рис. 8.7. Пример решения обратной задачи при измерении давления в цилиндре дизельного двигателя

Сопоставление кривых 3 и 4 говорит о том, что при неполных исходных данных совместное применение винеровского подхода с принципом минимального модуля приводит к тому, что решение обратной задачи остается гладким, хотя вследствие недостатка исходных данных решение обратной задачи теряет в своей точности.

Приведенный пример выполнен не в режиме реального времени, но апостериори, после регистрации выходного сигнала.

Понятно, что на практике, особенно, при реализации управляющих воздействий апостериорное решение обратных задач малоэффективно, за исключением некоторых особых случаев. По этой причине весьма актуальными представляются работы в области решения обратных задач в реальном времени с помощью физически реализуемых обратных фильтров.

8.6. ВОЗМОЖНОСТИ ПОСТРОЕНИЯ ФИЗИЧЕСКИ РЕАЛИЗУЕМОГО ОБРАТНОГО ФИЛЬТРА

Далее приводится формула для КЧХ оптимальной линейной динамической системы, позволяющей решить модифицированную задачу винеровской фильтрации, то есть выделить полезный сигнал $y(t)$ из смеси его с шумом $y_\delta(t) = y(t) + \delta(t)$ и обеспечить выполнение операции $z(t) = Ny(t)$ над ним по критерию минимума дисперсии погрешности результата. В наших обозначениях эта формула имеет вид:

$$H(j\omega) = \frac{I}{2\pi S_{yy}^-(\omega)} \int_{0+}^{\infty} e^{-j\omega t} \left\{ \int_{-\infty}^{\infty} \frac{S_{yz}(\omega')}{S_{yy}^+(\omega')} e^{j\omega' t} d\omega' \right\} dt, \quad (8.26)$$

где $S_{yy}(\omega')$ – спектральная плотность сигнала $y_\delta(t)$, $S_{yy}^+(\omega')$ и $S_{yy}^-(\omega')$ – результаты факторизации: $S_{yy}(\omega') = S_{yy}^+(\omega') \cdot S_{yy}^-(\omega')$, функция $S_{yy}^+(\omega')$ имеет все нули и полюса в нижней полуплоскости, нули и полюса функции $S_{yy}^-(\omega')$ лежат в верхней полуплоскости, $S_{yz}(\omega')$ – взаимная спектральная плотность сигналов $y_\delta(t)$ и $z(t)$.

Для решения операторного уравнения (8.2): $Ax(v) = y(v) + \delta(v)$ относительно $x(t)$, то есть обратной задачи необходимо, чтобы оператор N выполнял операцию $A^{-1}y(t)$, то есть был обратным оператору A , а это означает, что в этом случае $S_{yz}(\omega') = K(-j\omega') \cdot S_{xx}(\omega')$, где $S_{xx}(\omega')$ – спектральная плотность сигнала $x(t)$. Становится очевидным, что вследствие некорректности обратной задачи от отрицательного аргумента в КЧХ избавиться не удастся. Ситуация немного облегчается тем, что нули и полюса функции $S_{yy}^+(\omega')$ так же, как и у КЧХ, лежат в нижней полуплоскости, но утешение от этого незначительно.

В качестве выхода из этого положения предлагается допустить некоторую задержку выходного восстановленного сигнала $\tilde{z}(t)$ на

время τ . Кроме того, поскольку в наших задачах нереально всегда требовать стационарности сигналов, в формуле (8.26) сделаем такие же замены, как в разделе 8.6. В конечном итоге получим:

$$H(j\omega) = \frac{1}{2\pi S_{yy}^{T-}(\omega)} \int_{\tau+0}^{\infty} e^{-j\omega t} \left\{ \int_{-\infty}^{\infty} \frac{K(-j\omega') \cdot S_{xx}^T(\omega')}{S_{yy}^{T+}(\omega')} e^{j\omega' t} d\omega' \right\} dt, \quad (8.27)$$

где $S_{yy}^T(\omega') = S_{xx}^T(\omega') \cdot |K(j\omega')|^2 + S_{\delta\delta}^T(\omega')$, функции $S_{yy}^{T+}(\omega')$ и $S_{yy}^{T-}(\omega')$ – суть результаты факторизации спектральной плотности $S_{yy}^T(\omega')$.

Можно предположить, что вполне удовлетворительные результаты могут быть получены, если время задержки будет незначительно превышать постоянную времени корректируемого динамического звена.

Факторизация спектральных областей или заменяющих их функций сильно облегчается, если они выражены в дробно-рациональном виде. Программы, позволяющие выполнить дробно-рациональную аппроксимацию и последующую факторизацию экспериментальных значений вещественных функций, упоминаются в следующем разделе.

Можно предположить, что вторая возможность построить физически реализуемый обратный фильтр, восходит к работе Р. Р. Харченко, в которой он предложил способ "коррекции динамических погрешностей", создаваемых линейным динамическим звеном с передаточной функцией

$$K(p) = \frac{L_n(p)}{Q_m(p)},$$

где $L_n(p)$ и $Q_m(p)$ степенные полиномы p степеней n и m , соответственно, причем $m \geq n$.

Способ заключается в следующем. Последовательно с этим звеном включить звено с передаточной функцией

$$K_I(p) = \frac{Q_m(p)}{Q_k(p)},$$

такое, чтобы степень полинома числителя была выше степени полинома знаменателя, а полином числителя этого второго звена сокращался с полиномом знаменателя корректируемого. В этой ситуации передаточная функция этого последовательного соединений будет

$$K_2(p) = \frac{L_n(p)}{Q_k(p)},$$

и если полюсы этого соединения обеспечат устойчивость второго звена и меньшую постоянную времени, то тем самым быстродействие в результате увеличится, и погрешности в динамическом режиме измерений будут меньше.

Указанное утверждение не учитывает тот факт, что погрешности, возникающие в месте соединения двух динамических звеньев, могут быть усилены вторым звеном.

Здесь можно предложить следующий регуляризирующий прием. Полином знаменателя второго звена выбрать так, чтобы он обеспечивал очень маленькую постоянную времени. После второго звена, а лучше перед ним включить несколько (например, r) динамических инерционных звеньев первого порядка с передаточными функциями вида

$$W(p) = \frac{1}{1 + Tp}.$$

Если таких звеньев будет 2, то есть, если $r = 2$, то эти звенья будут играть роль ограничения первых производных сигнала, что ведет к ограничению его негладкости. С увеличением таких звеньев гладкость восстанавливаемого сигнала будет увеличиваться, что эквивалентно увеличению порядка регуляризации. Постоянная времени T будет играть роль параметра регуляризации. В соответствии с рекомендацией А. Н. Тихонова [13], при сильно сглаживающем операторе корректируемого звена рост порядка регуляризации выше двух играет незначительную роль. Поэтому представляется, что хватит двух таких звеньев.

По поводу изложенной второй возможности построения физически реализуемого обратного фильтра было высказано только предварительное соображение, не более того. Это предположение неплохо было бы исследовать хотя бы на моделях. При этом придерживаться следующей тактики. Вначале постараться искать решение при больших значениях постоянной времени T , что соответствует поиску решения в области очень гладких функций. Затем постепенно уменьшать это значение, ухудшая тем самым гладкость, и остановиться, когда результат будет удовлетворительным. Сведения об удовлетворительности результата могут быть получены при настройке такого обратного фильтра на ступенчатом входном сигнале.

8.7. ДРОБНО-РАЦИОНАЛЬНАЯ АППРОКСИМАЦИЯ И ФАКТОРИЗАЦИЯ

Первоначально дробно-рациональная аппроксимация вещественных функций использовалась для экспериментального определения КЧХ линейных динамических звеньев по дискретным экспериментальным значениям АЧХ [12]. В работе приведены формулы для оценки характеристик параметров КЧХ и АЧХ.

Такие ситуации возникают обычно при идентификации динамических характеристик датчиков неэлектрических величин, когда измерение фазового сдвига между гармоническим входным сигналом и гармоническим выходным сигналом невозможно. Подобные ситуации могут возникать и при идентификации динамических характеристик, например, усилителей с большим коэффициентом.

Полученные дискретные значения АЧХ возводятся в квадрат. Эти значения аппроксимируются дробно-рациональными выражениями, аргументами которых являлись квадраты круговой частоты, то есть ω^2 . При аппроксимации степени числителя и знаменателя увеличиваются на единицу, но каждый раз так, чтобы степень числителя не превышала степень знаменателя. Остановка

этого процесса выполняется после положительного исхода проверки гипотезы о том, что невязка не превышает характеристику погрешности регистрации экспериментальных значений АЧХ.

По окончании аппроксимации выполняется факторизация полученного выражения квадрата АЧХ. Для этого находятся корни полинома числителя и полинома знаменателя. Из каждой четверки корней выбираются те, которые располагаются в верхней полуплоскости комплексной плоскости, а для числителя – и на вещественной оси. Из этих корней формируются числитель и знаменатель дробно-рационального выражения КЧХ $K(j\omega)$ в каноническом виде: в виде произведения биномов и трехчленов.

8.8. ПРИМЕР ЦИФРОВОГО МОДЕЛИРОВАНИЯ ДЛЯ НОРМИРОВАНИЯ ПОГРЕШНОСТЕЙ

Результат любого измерения должен иметь корректное метрологическое обеспечение, конечной целью которого является определение гарантированного значения погрешности (неопределенности). Для успешного достижения этой цели в подавляющем большинстве реальных измерений используются метрологические характеристики, которые нормируются для средств измерений и контролируются при их метрологических испытаниях [8]. Проблема обеспечения единства измерений и взаимного доверия к результатам измерений нигде не стоит так остро, как в области хроматографического (и спектрометрического) количественного химического анализа сложных веществ.

Метрологические характеристики хроматографов нормируются для колонки, детектора и интегратора и контролируются с помощью эталонных проб веществ с хорошо разделенными пиками. Значение этой предельной погрешности составляет от 0.1 % до 1 %. Однако большое количество компонентов исследуемых веществ, пересечение хроматографических пиков, наличие пиков-всадников, разнообразие программ разделения пиков и вычисления их площадей, засорение колонки и другие факторы приводят к появлению значительных

погрешностей до 50% [14], несоизмеримых с нормированными погрешностями хроматографов.

Наибольший вклад в погрешность хроматографического анализа вносят программы обработки хроматограмм, которые реализуются в компьютерах, встроенных в современные хроматографы, и предназначены для вычисления площадей пиков, в том числе, пересекающихся и пиков-всадников. Можно считать, что погрешности, вызванные несовершенством колонок, детекторов и интегратора (АЦП), практически не зависят от формы реальной хроматограммы и соответствуют нормированным характеристикам.

Очевидным, но очень трудоемким способом комплектной оценки характеристик погрешностей результата хроматографического анализа сложного вещества представляется следующее действие.

После выполнения анализа подготавливается реальная проба исследованного вещества, концентрации компонентов которого равны полученным результатам. Эта образцовая проба вводится в хроматограф вновь, и погрешности анализа вычисляются как разности между вновь полученными результатами и концентрациями, которые теперь уже известны. Таким образом оценивается характеристика погрешности всего аналитического канала: колонки, детектора, интегратора и вычислений.

Вместо этого предлагается использовать поэлементную оценку погрешности хроматографического анализа (рис. 8.8), а именно, отдельно оценить погрешность, вносимую программой вычислений, и оценить суммарную погрешность с помощью расчета с учетом характеристики погрешности, нормированной для хроматографа.



Рис. 8.8. Метод определения погрешностей

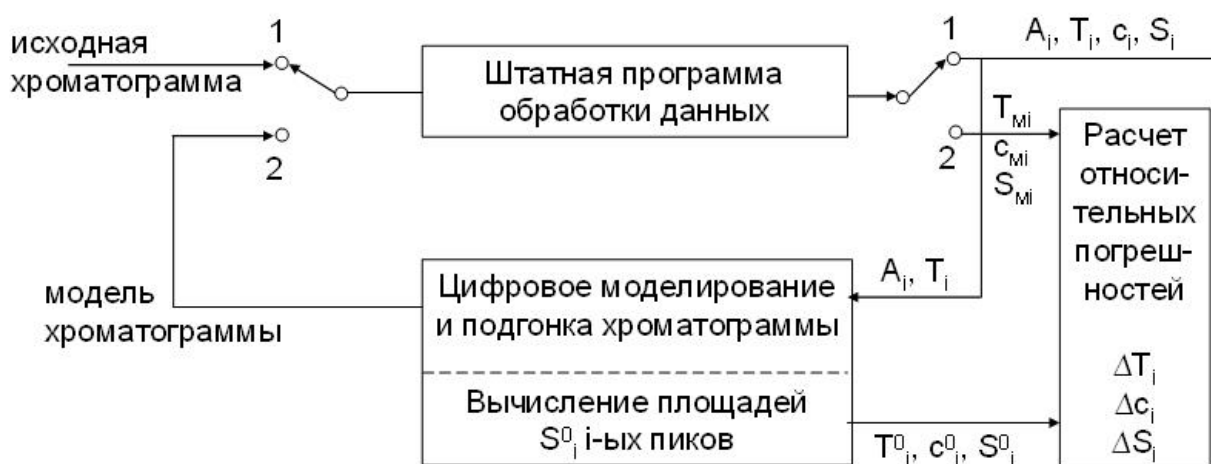


Рис. 8.9. Процедура определения характеристик погрешности

Цифровая модель хроматограммы исследуемого вещества строится по результатам выполненного анализа с очень узкими, хорошо разделенными пиками, площади которых известны и равны тем, которые получены при анализе. После этого путем свертывания с предполагаемой аппаратной функцией и путем пополнения модели другими факторами (моделирование пиков-всадников, базовой линии и др.) модель подгоняется к реальной хроматограмме визуалью на дисплее компьютера. После такой подгонки модель подается на вход встроенного в хроматограф компьютера и обрабатывается вновь с помощью штатного программного обеспечения (рис. 8.9).

Пусть, например, в результате хроматографического анализа многокомпонентного вещества получены значения положений пиков (времени удерживания) T_j , площадей S_j , амплитуд A_j (рис. 8.10).

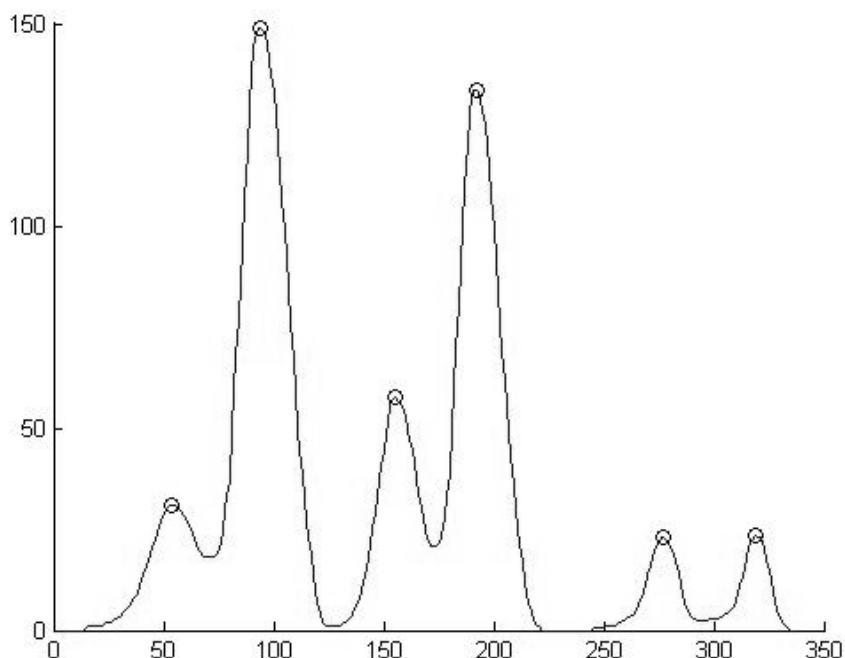


Рис. 8.10. Анализируемая хроматограмма. Время удерживания и амплитуды отмечены отдельными точками

Модель хроматограммы строится как сумма пиков, которые моделируются по одному, каждый из которых обладает параметрами T_j , S_j , A_j . Начальная ширина пиков не превосходит 1/20 от ширины пиков реальной хроматограммы (рис. 8.11).

Модель пика строится в виде:

$$x_i(t) = K \cdot A_i \exp\left(-\frac{(t - T_i)^2}{2d^2(t)}\right),$$

где K – коэффициент пропорциональности, A_i , T_i – амплитуда и время удерживания i -го пика, $d(t)$ – характеристика ширины пика, задается в виде $d(t) = \sigma(1 + bt)$.

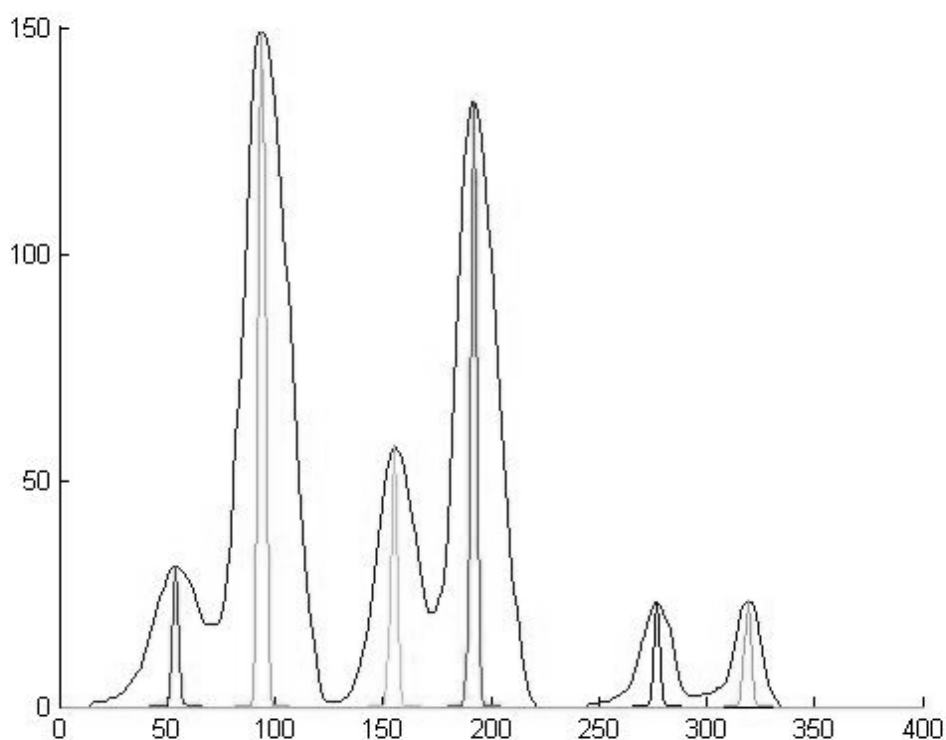


Рис. 8.11. Первоначальное моделирование пиков

Уширение каждого пика моделируется как свертка с гауссовским симметричным ядром единичной площади (рис. 8.12). При этом контролируются значения амплитуд, которые сохраняются равными A_j .

$$y_i(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x_i(t) \exp\left(-\frac{(t-\tau)^2}{2\sigma^2}\right) d\tau.$$

Подгонка пиков модели по амплитуде выполняется изменением параметра K . Для подгонки пиков по ширине регулируются параметры σ и b функции $d(t)$.

Моделируется также экспоненциальное затягивание заднего фронта пиков, пьедесталы под пиками-всадниками, базовые линии.

Модель пьедестала под пиком-всадником моделируется в виде кусочной функции:

$$z(t) = \begin{cases} C, & t \leq T_0 \\ C \cdot e^{-\alpha(t-T_0)}, & t > T_0 \end{cases}.$$

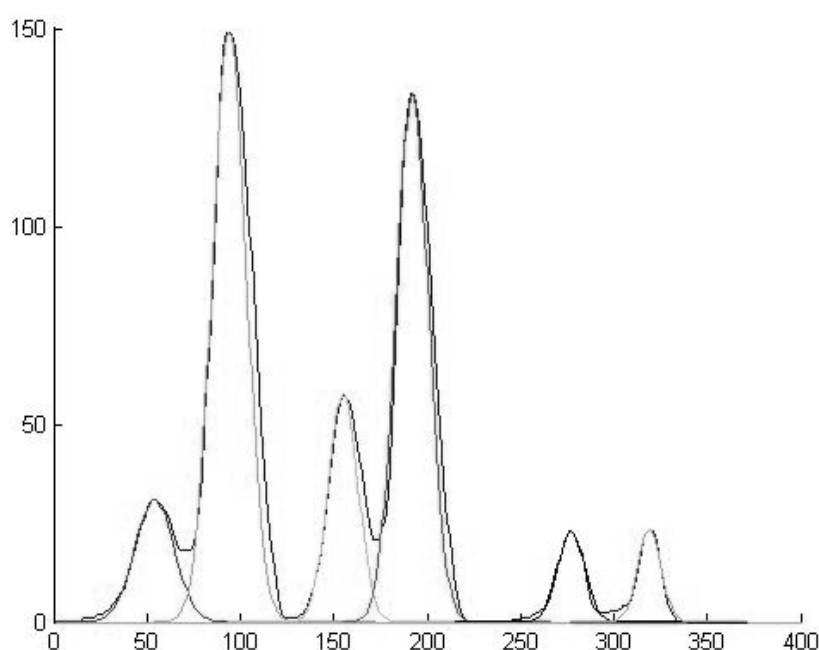


Рис. 8.12. Смоделировано уширение каждого пика

Экспоненциальное затягивание пиков моделируется следующим образом (рис. 8.12).

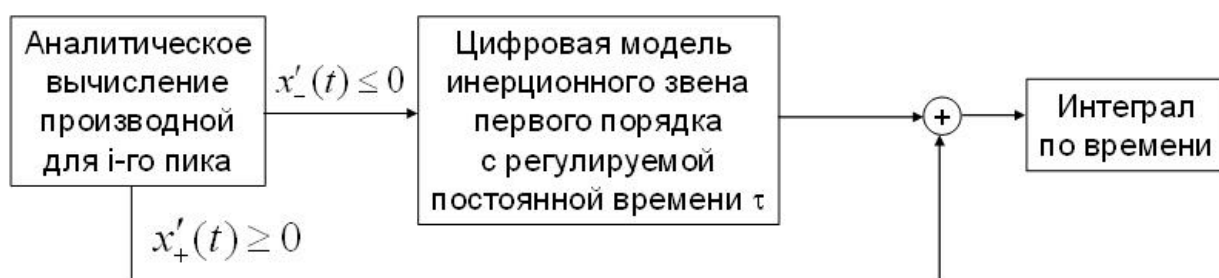


Рис. 8.12. Алгоритм экспоненциального затягивания пика

1. Модель каждого i – го пика дифференцируется (производная вычисляется аналитически от исходного сигнала $x(t)$ – (2.10)).

2. Отрицательная функция $x'_-(t)$ пропускается через цифровую модель инерционного звена первого порядка с регулируемой постоянной времени τ : $h(t) = \frac{1}{\tau} \exp\left(-\frac{t}{\tau}\right)$.

3. Сигнал на выходе инерционного звена является сверткой $x'_-(t)$ и $h(t)$ (см. (8.10), (2.1)):

$$y(t) = h(t) * x(t).$$

4. После чего функции $x'_+(t)$ и $y(t)$ объединяются вновь и интегрируются по времени (рис. 8.14).

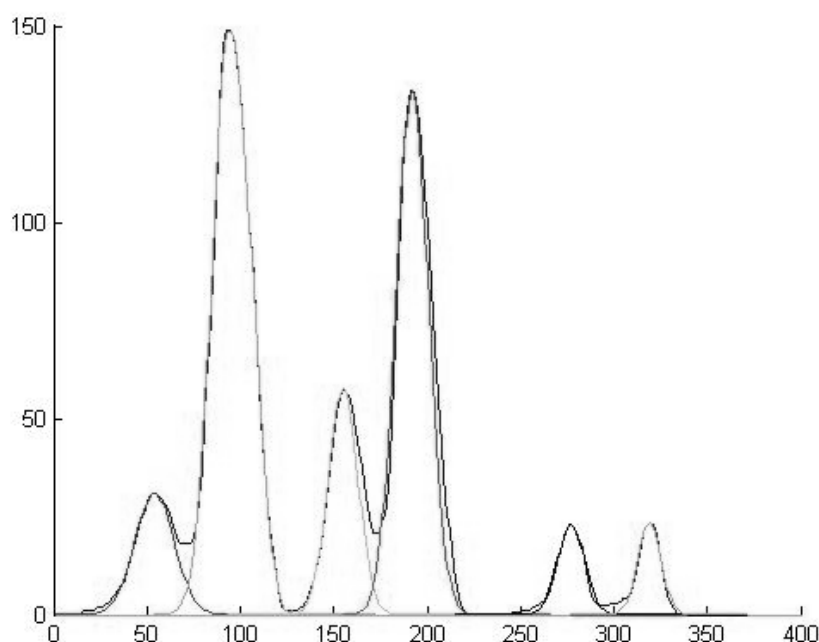


Рис. 8.14. Для второго пика выполнено экспоненциальное затягивание

В реализации данного алгоритма используется предварительная дискретизация во времени, выполняемая электронным преобразователем (рис. 8.8) в соответствии с выражением (1.2).

Алгоритм свертки при вычислении $y(t)$ реализован в соответствии со свойством 4 раздела 3.3.4 настоящего пособия путем последовательного применения ППФ и ОПФ.

Модель базовой линии строится для отдельных участков хроматограммы также в виде участков линейной или гармонической функций, вид и параметры которой подбираются оператором.

Затем все эти модели складываются, и полученная сумма сравнивается с исходной хроматограммой на экране дисплея (рис. 8.15).

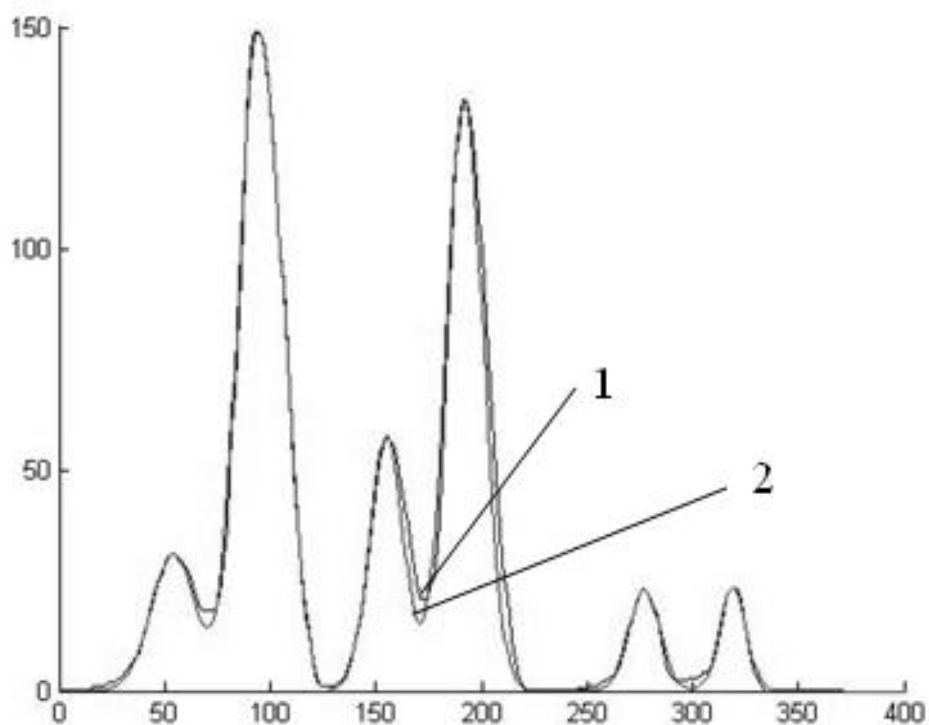


Рис. 8.15. Исходная (1) и смоделированная (2) хроматограммы

После удовлетворительного совпадения модели с реальной хроматограммой модель подается на вход штатной программы, в результате чего будут получены значения параметров T_{mj} , площадей S_{mj} , амплитуд A_j .

В качестве оценок погрешностей результатов анализа вещества используются разности: $\Delta T_j = T_{mj} - T_j$, $\Delta S_j = S_{mj} - S_j$.

Вычисляются оценки относительных погрешностей, и они складываются со значениями относительных погрешностей, нормированными для хроматографа.

Постановка этой задачи была предложена профессором Г. Н. Солопченко и является альтернативой традиционным методам нормирования погрешностей [14].

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Ашанин В. Н.* $\Sigma\Delta$ -АЦП: основы теории и проектирования / В. Н. Ашанин, Б. В. Чувькин, Э. К. Шахов – Пенза : ИИЦ ПГУ, 2009. – 188 с.
2. *Богач Н. В.* Автоматизация расчета вейвлет-функции с компактным носителем/ Н.В. Богач, Ш.Т. Ву // Научно-технические ведомости СПбГПУ. – 2006. – Том 1, № 6, с. 5–10.
3. *Варгаузин В. А.* Цифровая обработка сигналов: Минимаксные аппроксимации для задач цифровой фильтрации: Учебное пособие. / В.А. Варгаузин – СПб. : Изд-во СПбГПУ, 2004. – 87 с.
4. *Гутников В. С.* Фильтрация измерительных сигналов. / В.С. Гутников – Л. : Энергоатомиздат, 1990. – 192 с.
5. *Давыдов А. В.* Сигналы и линейные системы: Тематические лекции. / А. В. Давыдов – Екатеринбург : УГГУ, ИГиГ, кафедра геоинформатики. [Электронный ресурс], 2006.
6. *Добеши И.* Десять лекций по вейвлетам: [пер. с англ.] / И. Добеши – М. : Ижевск НИЦ «Регулярная и хаотическая динамика», 2006. – 464 с.
7. Измерительная информационная техника и метрология. Лаб. практикум Ч 4. / Лабораторные работы № 4, № 5; под общей ред. проф. Г. Н. Солопченко. – СПб. : Изд-во Политехн. ун-та, 2009. – 54 с.
8. *Куляшова З.В.* Моделирование хроматограмм, как средство оценки погрешности количественного хроматографического анализа / З. В. Куляшова // Сборник трудов международной научно-практической конференции “Измерения в современном мире-2009”, г. Санкт-Петербург, 2009.
9. *Олифер В. Г.* Компьютерные сети. Принципы, технологии, протоколы: учебник для ВУЗов. / В. Г Олифер, Н. А. Олифер – СПб. : ПИТЕР, 2007. – 544 с.
10. *Оппенгейм А.* Цифровая обработка сигналов. / А. Оппенгейм, Р. Шафер – М. : Технсфера, 2006. – 856 с.

11. *Сергиенко А. Б.* Цифровая обработка сигналов. / Сергиенко А. Б. – СПб. : ПИТЕР, 2006. – 752 с.
12. *Солопченко Г. Н.* Определение параметров дробно-рациональной передаточной функции по экспериментальным данным / Г. Н. Солопченко // Метрология. – 1978. – № 5, с. 25-35.
13. *Тихонов А.Н.* Методы решения некорректных задач. / А. Н. Тихонов, В. Я. Арсенин – М. : Наука, 1974. – 222 с.
14. *Фаткудинова Ш.Р.* Применение цифровых моделей реальных хроматограмм для оценки погрешности результатов химического анализа / Ш. Р. Фаткудинова, Г. Н. Солопченко // Журнал аналитической химии. – 1998. – № 11, с. 1158-1165.
15. *Хабурзания Т.З.* Нейросетевой алгоритм решения обратной задачи / Т. З. Хабурзания // Сборник трудов международной научно-практической конференции “Измерения в современном мире-2009”, г. Санкт-Петербург, 2009.

