



КАФЕДРА "РАДИОЭЛЕКТРОННЫХ СИСТЕМ"



С.А. Кудряков

Радиотехнические цепи и сигналы



Санкт-Петербург

2015

УДК 621.37

ББК 32.841

К88 Ку

К88 Ку

Кудряков С.А. Радиотехнические цепи и сигналы. (Учебное пособие).- С. Пб.: Изд-во «Свое Издательство», 2015.- 340 с.

Настоящее издание представляет собой печатную версию электронного учебного пособия по дисциплине "Теория радиотехнических цепей и сигналов". Пособие включает в себя краткий курс лекций, демонстрационные примеры и задания для компьютерного моделирования, используемые при подготовке бакалавров и дипломированных специалистов в Санкт-Петербургском государственном университете гражданской авиации по ФГОС ВПО 3-го поколения.

Материал учебного пособия излагается в соответствии с действующим учебным планом подготовки студентов специализации "Организация радиотехнического обеспечения полетов ВС" и профиля "Эксплуатация наземных средств радиотехнического обеспечения полетов ВС и авиационной электросвязи", предполагающим проведение теоретических лекций и практических занятий, а также лабораторных работ и выполнение курсовой работы.

Практикум включает в себя теоретическое решение задач по отдельным разделам курса и работу на компьютерах в среде *MatLab*, *MathCad*, и *Multisim (Electronics Workbench)* по моделированию радиотехнических сигналов и цепей.

Курс лекций построен с учетом знаний, полученных студентами при изучении предшествующих курсов по высшей математике, физике, информатике, электронике и электротехнике в объемах стандартных программ для технических вузов.

ISBN-978-

© С.А.Кудряков

СОДЕРЖАНИЕ

	Стр.
Содержание	3
0.ВВЕДЕНИЕ	8
0.1.Начальные сведения о радиотехнике	8
0.2.Основы научных парадигм детерминизма, стохастичности и хаоса. Классификация сигналов и цепей	13
0.3. Вопросы для самоконтроля и повторения	21
 ГЛАВА 1.Детерминированные сигналы	 22
1.1.Сигналы и информация	22
1.2.Пространства сигналов	26
1.3.Декомпозиция аналогового сигнала	30
1.4.Периодические сигналы. Ряд Фурье	36
1.5.Гармонический анализ непериодических сигналов. Преобразование Фурье	52
1.6. Основные свойства преобразования Фурье	55
1.7. Примеры расчета преобразования Фурье	60
1.8. Преобразование Фурье для некоторых сингулярных функций	64
1.9.Корреляционный анализ детерминированных сигналов	68
1.10. Вопросы для самоконтроля и повторения	72
 ГЛАВА 2.Модулированные сигналы	 74
2.1. Общие сведения о модуляции	74
2.2. Амплитудная и угловая модуляция	75
2.2.1. Амплитудная модуляция	75
2.2.2. Угловая модуляция	84
2.3.Аналитические сигналы. Преобразование Гильберта	88
2.4.Внутриимпульсная модуляция	93
2.5. Общие сведения о импульсно модулированных сигналах	95
2.6. Вопросы для самоконтроля и повторения	97
 ГЛАВА 3.Дискретные сигналы	 98
3.1.Дискретизация непрерывных сигналов	98
3.2. Дискретное преобразование Фурье и его свойства	104

3.3.Растекание спектра и эффект Гиббса	106
3.4.Быстрое преобразование Фурье. Дополнение нулями и частотное разрешение	109
3.5. Вопросы для самоконтроля и повторения	117
ГЛАВА 4. Общие сведения о случайных процессах	118
4.1.Случайные процессы непрерывного времени и их основные характеристики	118
4.2. Корреляционная функция	121
4.3. Понятия стационарности и эргодичности	123
4.4. Спектральная плотность мощности и ее связь с корреляционной функцией	124
4.5. Случайный узкополосный сигнал	128
4.6. Модели случайных сигналов в дискретном времени	132
4.7. Вопросы для самоконтроля и повторения	142
ГЛАВА 5. Линейные цепи с постоянными параметрами	144
5.1.Двухполюсники и четырехполюсники как базовые структурные элементы радиотехнических цепей	144
5.2. Преобразование Лапласа и его основные свойства	148
5.3.Радиотехнические цепи как линейные системы	152
5.3.1. Передаточные функции	152
5.3.2. Частотные характеристики линейных систем	154
5.3.3.Временные характеристики линейных систем	157
5.4.Формы представления передаточной функции	161
5.4.1.Нули и полюсы	161
5.4.2.Диаграмма полюсов и нулей	161
5.4.3.Полюсы и вычеты	162
5.5. Связь передаточной функции с коэффициентами четырехполюсника	164
5.6.Пространство состояний	165
5.7.Фундаментальная матрица системы и некоторые способы ее вычисления	169
5.8.Однонаправленные цепи и цепи с обратной связью	172
5.9. Общие сведения об устойчивости линейных цепей	175
5.10. Частотные и временные характеристики простейших RC и RL цепей.	177
Интегрирующие и дифференцирующие цепи	
5.11. Изолированный колебательный контур и его характеристики	184

5.11.1. Последовательный колебательный контур	184
5.11.2. Четырехполюсники на основе последовательного колебательного контура	187
5.11.3. Параллельный колебательный контур	188
5.11.4. Прохождение амплитудно-модулированного сигнала через колебательный контур	189
5.12. Электрические фильтры	191
5.12.1. Общие понятия и классификация	191
5.12.2. Реактивные фильтры	193
5.12.3. Пассивные RC- фильтры	197
5.12.4. Активные фильтры. Использование обратной связи для формирования желаемых частотных свойств фильтра	199
5.13. Прохождение случайного сигнала через линейную цепь	202
5.14. Вопросы для самоконтроля и повторения	203
ГЛАВА 6. Преобразование сигналов в нелинейных и параметрических цепях	205
6.1. Линейные и нелинейные элементы цепей	205
6.2. Аппроксимация характеристик безынерционных нелинейных элементов	206
6.3. Нелинейные безынерционные элементы при гармонических воздействиях	208
6.3.1. Полиномиальная аппроксимация ВАХ	211
6.3.2. Кусочно-линейная аппроксимация ВАХ	213
6.3.3. Воздействие на нелинейную цепь бигармонического колебания	216
6.3.4. Преобразование частоты сигнала (гетеродирование)	218
6.4. Нелинейное резонансное усиление и умножение частоты	218
6.5. Амплитудная модуляция	221
6.6. Амплитудное детектирование	225
6.6.1. Линейный диодный детектор	225
6.6.2. Коллекторный детектор	227
6.6.3. Квадратичный детектор	228
6.6.4. Синхронное детектирование	228
6.7. Частотное и фазовое детектирование сигналов	229
6.8. Воздействие случайных сигналов на нелинейный безынерционный элемент	232
6.8.1. Преобразования плотности вероятности входного сигнала	233

6.8.2. Преобразование корреляционной функции случайного процесса при прохождении нелинейной безынерционной цепи	236
6.8.3. Спектральная функция случайного процесса на выходе нелинейного безынерционного звена	237
6.9. Общие сведения о параметрических цепях	238
6.10. Вопросы для самоконтроля и повторения	241
ГЛАВА 7. Генерирование гармонических колебаний	242
7.1. Понятие автоколебательной системы	242
7.2. Баланс фаз и баланс амплитуд. Возникновение колебаний в автогенераторе.	243
7.3. Стационарный режим работы автогенератора	246
7.4. Мягкий и жесткий режимы самовозбуждения автогенератора.	248
7.5. Нелинейное уравнение автогенератора.	250
7.6. Основные схемы реализации LC-генераторов	252
7.7. RC-автогенераторы и автогенераторы с внутренней обратной связью	256
7.7.1. Автогенераторы на фазосдвигающих цепочках	256
7.7.2. RC-автогенератор с мостом Вина	259
7.7.3. Автогенераторы с внутренней обратной связью	260
7.8. Вопросы для самоконтроля и повторения	262
ГЛАВА 8. Выделение сигналов на фоне помех. Принципы оптимальной линейной фильтрации сигналов	263
8.1. Фильтрация и отношение сигнал-шум.	263
8.2. Согласованная фильтрация детерминированных сигналов	265
8.3. Оптимальная фильтрация при не белом шуме	268
8.4. Квазиоптимальная фильтрация детерминированных сигналов	271
8.5. Оптимальная фильтрация. Фильтр Винера.	274
8.6. Общие сведения о фильтре Калмана-Бьюси	278
8.7. Корреляторы сигналов.	281
8.8. Вопросы для самоконтроля и повторения	282
ГЛАВА 9. Основы цифровой обработки сигналов	283
9.1. Квантование аналогового сигнала.	283
9.2. Разностные уравнения цифрового фильтра. Понятие БИХ и КИХ фильтров.	285

9.3. Z-преобразование и его свойства	290
9.4. Передаточная функция цифрового фильтра и формы ее представления	292
9.5. Основы проектирование дискретных фильтров	296
9.5.1. Синтез рекурсивных фильтров по аналоговому прототипу. Метод билинейного Z-преобразования	297
9.5.2. Общие сведения о методе инвариантной импульсной характеристики и прямых методах синтеза цифровых фильтров	301
9.5.3. Частотные преобразования АЧХ фильтров нижних частот	302
9.6. Оценивание параметров сигналов по данным экспериментальных наблюдений	309
9.6.1. Принцип максимального правдоподобия и метод наименьших квадратов	309
9.6.2. Метод наименьших квадратов (блочный подход)	311
9.6.3. Рекурсивный метод наименьших квадратов	313
9.6.4. Оценивание параметров случайных сигналов	314
9.7. Основы адаптивной обработки сигналов	323
9.7.1. Общие сведения об адаптивных фильтрах	325
9.7.2. Дискретный фильтр Калмана	324
9.8. Области применения адаптивных фильтров.	329
9.9. Общие сведения о цифровом спектральном оценивании	334
9.10. Вопросы для самоконтроля и повторения	339
Рекомендуемая литература	340

ВВЕДЕНИЕ

0.1. Начальные сведения о радиотехнике

Термин *радиотехника*, происходящий от латинского radio (излучаю) самим своим названием определяет ту область науки и техники, которая занимается изучением и применением электромагнитных колебаний и волн соответствующего диапазона для передачи и приема информации на расстоянии. К радиоволнам принято относить электромагнитные волны с частотами от 3 до 3 ТГц, распространяющиеся в пространстве без искусственных направляющих линий.

Радиотехника является относительно молодой наукой, берущей свое начало с открытия явления электромагнитной индукции (1831 г. М.Фарадей), электромагнитного поля (1865 г. Дж. Максвелл) и практического получения электромагнитной волны (1887 г. Г.Герц).

Начало практического использования электромагнитных волн для передачи информации связано с именами российского профессора А.Попова, и итальянского инженера Гульемо Маркони. Г. Маркони проводил свои работы после опубликования Поповым своих результатов в 1895 году, но успел их запатентовать, поэтому за рубежом именно его часто считают изобретателем первого устройства для радиосвязи.

За прошедшее с представления первого устройства для радиосвязи время радиотехника сделала огромный шаг вперед и на сегодняшний день охватила практически все сферы деятельности человека. Постоянное развитие технологии, снижение относительной стоимости радиотехнических устройств и миниатюризация их размеров обеспечили широчайшее распространение средств радиотехники, как на производстве, так и в быту. К настоящему времени существенно расширился и диапазон используемых радиоволн. Современная классификация диапазонов радиоволн приведена в табл.1.1. В реальной практике авиационных радиотехнических систем используются радиоволны, диапазон которых простирается от ОНЧ до КВЧ.

Таблица 1.1.

Классификация диапазонов радиоволн

Наименование частотного диапазона	Границы диапазона (Гц)	Наименование волнового диапазона	Границы диапазона (м)
Крайне низкие, КНЧ	3-30 Гц	Декаметровые	100-10Мм
Сверхнизкие, СНЧ	30-300 Гц	Метровые	10-1 Мм
Инфранизкие, ИНЧ	0,3-3 кГц	Гектокилометровые	1000-100 км
Очень низкие, ОНЧ	3-30 кГц	Мириаметровые	100-10 км
Низкие частоты, НЧ	30-300 кГц	Километровые	10-1 км
Средние, СЧ	0,3-3 МГц	Гектометровые	1-0,1 км

Высокие частоты, ВЧ	3-30 МГц	Декаметровые	100-10 м
Очень высокие, ОВЧ	30-300 МГц	Метровые	10-1 м
Ультравысокие, УВЧ	0,3-3 ГГц	Дециметровые	1-0,1 м
Сверхвысокие, СВЧ	3-30 ГГц	Сантиметровые	10-1 см
Крайне высокие, КВЧ	30-300 ГГц	Миллиметровые	10-1 мм
Гипервысокие, ГВЧ	300-3000 ГГц	Децимиллиметровые	1-0,1 мм

Наиболее близкими к радиотехнике являются такие отрасли как *электротехника* и *электросвязь*. Используя во многом общий понятийный аппарат, методологию анализа и синтеза, эти отрасли, тем не менее, имеют различные главные задачи. Электросвязь ориентирована на задачи приема и передачи информации по проводным каналам, а электротехника сосредоточивается на вопросах передачи и преобразования электрической энергии.

Радиотехнические системы являются разновидностью информационно-управляющих систем и в общем случае осуществляют задачи *радиоуправления*, а так же *передачи, извлечения* или *разрушения* информации с помощью электромагнитных волн.

Системы передачи информации в свою очередь подразделяются на системы радиосвязи, телеметрии, передачи команд, радиовещания и телевидения.

К системам извлечения информации относятся радиолокационные и радионавигационные системы, а так же системы радиоастрономии и радиоразведки.

К системам разрушения информации относятся системы, предназначенные для целенаправленного выведения из строя радиосистем противника или создания условий, в которых работа радиосистем противника становится затрудненной или невозможной.

Системы радиоуправления применяются для управления работой различных объектов с помощью радиосигналов.

В общей задаче радиотехники принято выделять несколько основных направлений, а именно: радиосвязь, телевидение, радиолокацию, радионавигацию, радиотелеметрию и радиотелемеханику.

Радиосвязь представляет собой информационную связь между объектами посредством радиоволн. При этом радиосвязь может быть односторонней, когда потребитель имеет в своем распоряжении только приемник, а источник имеет только передатчик или двусторонней, когда в распоряжении каждого участника радиосвязи имеется и приемник и передатчик. В сеансах связи могут быть задействованы как только два участника, так и несколько участников.

К самым распространенным видам односторонней связи относится, например, классическое *радиовещание и телевидение*, при котором источники информации имеют только передатчики, а потребители обладают только приемниками.

Радиолокация представляет собой обнаружение объектов, определение их координат, а также параметров движения.

Радионавигация как способ определения собственных координат объекта с помощью радиотехнических средств нашла широчайшее применение в авиации, космонавтике и различных транспортных системах.

Радиотелеметрия предназначена для измерения физических величин на удаленных объектах и в труднодоступных местах, например, таких как поисковые зонды, спутники и т.п.

Радиоуправление обеспечивает решение задачи автоматического управления удаленными объектами посредством радиоволн.

По мере развития технического прогресса, совершенствования элементной базы, появлению возможностей миниатюризации исполнения различных блоков и систем наблюдается широкая интеграция радиотехники практически во все области науки и техники, включая биологию и медицину.

Обобщенно, произвольную радиотехническую систему, как систему информационного обмена можно представить с помощью структурной схемы, изображенной на рис.0.1.

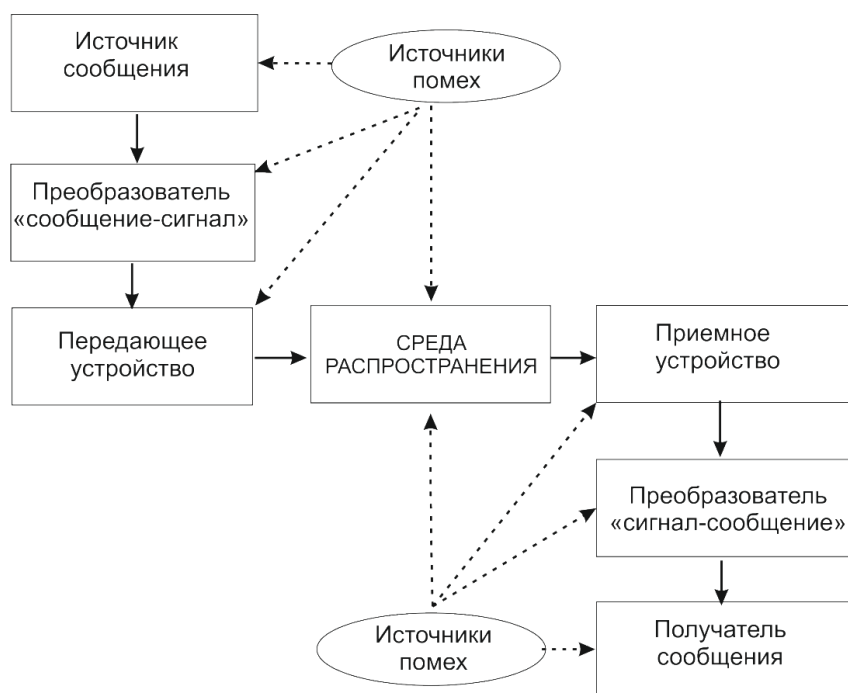


Рис.0.1. Обобщенная схема радиотехнической системы

Обязательными элементами обобщенной радиотехнической системы являются источник информации (сообщения), преобразователь исходной информации в первичный сигнал, передающее устройство (передатчик), среда распространения сигнала, приемное устройство (приемник), преобразователь первичного сигнала в сообщение и получатель информации (сообщения). Совокупность технических средств и среды распространения сигналов для передачи сообщений от источника к получателю называют каналом связи. Канал связи, понимаемый в узком смысле, представляет только физическую среду распространения сигналов.

Кроме того в любой реальной системе присутствуют источники помех и возмущений, которые воздействуют на все ранее перечисленные элементы радиосистемы. Воздействие источника помех на канал связи и другие элементы радиосистемы приводят к искажению передаваемого сигнала, что естественным образом может повлиять на правильность восприятия информации получателем. Поэтому одной из важнейших задач радиотехники является не просто передача и прием информации, а обеспечение процесса безошибочной передачи и восприятия информации в условиях воздействия естественных или искусственных помех.

При распространении сигналов, несущих полезную информацию, в дополнение к основному сигналу и одновременно с ним регистрируются и мешающие сигналы - шумы и помехи различной физической природы. К помехам относят не только искажения полезных сигналов, возникающие в канале связи, но и влияние различных дестабилизирующих факторов на работу отдельных элементов радиотехнической

системы. Выделение полезных составляющих из общей суммы принятых сигналов или максимальное подавление шумов и помех в информационном сигнале при сохранении его полезных составляющих является одной из основных задач первичной обработки сигналов или результатов наблюдений.

Типы помех могут быть классифицированы по источникам их возникновения, по энергетическому спектру, по характеру воздействия на сигнал, по вероятностным характеристикам и другим признакам. Источники помех могут быть внутренними и внешними.

Внутренние шумы объясняются физической природой источников сигналов (например, тепловые шумы электронных потоков в электрических цепях или дробовые эффекты в электронных приборах). Кроме того они возникают в измерительных устройствах и системах передачи и обработки сигналов из-за влияния различных дестабилизирующих факторов (изменение температуры, влажности, нестабильности источников питания, влияния механических вибраций на гальванические соединения, и т.п.)

Внешние источники шумов могут быть искусственного и естественного происхождения. Естественными источниками помех являются молнии, флуктуации магнитных полей, всплески солнечной энергии, и т.д. К искусственным источникам помех относятся промышленные или преднамеренные воздействия на радиотехнические системы с целью вывода их из строя (военное применение). Электрические и магнитные поля различных источников помех вследствие наличия индуктивных, емкостных и резистивных связей создают на различных участках и цепях сигнальных систем паразитные разности потенциалов и токи, которые добавляются к полезным сигналам и искажают их.

По характеру изменения помехи во времени принято подразделение на флуктуационные, импульсные и периодические помехи. Флуктуационные или шумовые помехи представляют хаотический и беспорядочный во времени процесс в виде нерегулярных случайных всплесков различной амплитуды. Как правило, при моделировании предполагают, что флуктуационные помехи распределены по нормальному закону с нулевым средним и оказывают существенное влияние только на сигналы низкого уровня.

Импульсные помехи проявляются как в виде отдельных импульсов, так и в виде последовательности импульсов, форма и параметры которых имеют случайный характер. Причинами импульсных помех являются резкие броски тока и напряжения в промышленных установках, транспортных средствах, а также природные электрические явления.

Периодические помехи вызываются ритмическими воздействиями на канал связи или части радиотехнических систем вызванные низкочастотными или высокочастотными изменениями электромагнитных полей линий электропередач, силовых электроустановок и другой техники. В том случае, когда основная мощность помех сосредоточена на отдельных определенных участках диапазона частот, то такие помехи называют сосредоточенными. Одним из примеров такой помехи являются, например, периодические помехи на частоте напряжения промышленной сети (или кратных частотах).

По характеру воздействия на полезный сигнал помехи принято подразделять на аддитивные и мультипликативные. Аддитивные помехи суммируются с сигналом (предполагается простое алгебраическое суммирование), не зависят от его значений и формы и не изменяют информативной составляющей самого сигнала. Мультипликативные помехи представляют собой случай нелинейного воздействия на полезный сигнал и могут изменять форму информационной части сигнала.

Следует заметить, что деление сигналов на полезные и мешающие (шумовые) является достаточно условным. Те из них, которые в некотором случае являются мешающими, в других оказываются носителями основной информации. Так, например, в диагностических системах возникающие шумовые искажения несут в себе основную информацию о характере имеющейся неисправности.

Каждый из элементов радиотехнической системы в той или иной степени предназначен для преобразования передаваемых или принимаемых сигналов. При этом простейшими "первокирпичиками" любой самой сложной радиотехнической системы являются элементы, из которых они изготавливаются. Эти элементы объединяются в звенья, каскады и блоки, на основе которых строятся функционально законченные устройства и комплексные системы. Эту логическую связь можно условно представить в виде вложенных друг в друга "матрешек", как показано на рис.0.2. Следует отметить, что приведенная классификация зависит от контекста рассматриваемой задачи и то, что в одних условиях рассматривается как система, в других обстоятельствах трактуется как элемент (или подсистема) более сложного образования.

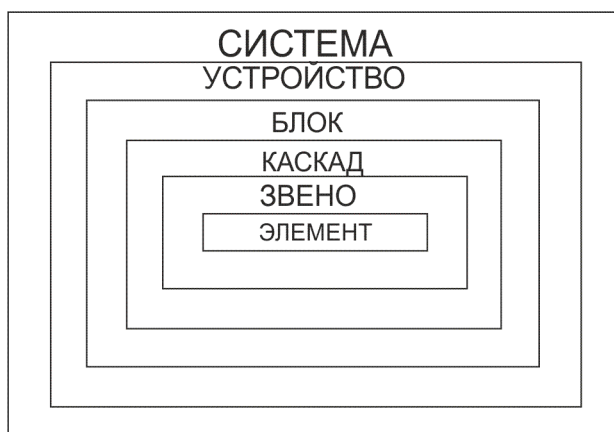


Рис.0.2. Иерархическая структура радиотехнических цепей

Строго говоря, подобная классификация характеризует не столько сами цепи, сколько принимаемую при их исследовании модель, описывающую реальное устройство. Выбор подходящей модели в любой научной дисциплине является весьма важной и непростой процедурой, во многом определяемой взглядами конкретного исследователя и сложившимися в конкретной области научных знаний традициями. В рамках настоящего курса лекций, исходя из методических целей, мы будем придерживаться именно такой классификации радиотехнических цепей.

Последующий раздел Введения мы посвятим вопросам краткого исторического обзора развития общенаучных парадигм моделирования окружающего мира, так как именно они, в конечном счете, формируют философскую основу инженерного мировоззрения.

В настоящем курсе лекций мы будем использовать понятие радиотехнической цепи, которое объединяет все радиотехнические устройства и в зависимости от контекста может обозначать и простейший элемент, и сложную систему.

Исторически принято подразделять радиотехнические цепи на три крупные категории в соответствии со схемой, приведенной на рис.0.3.

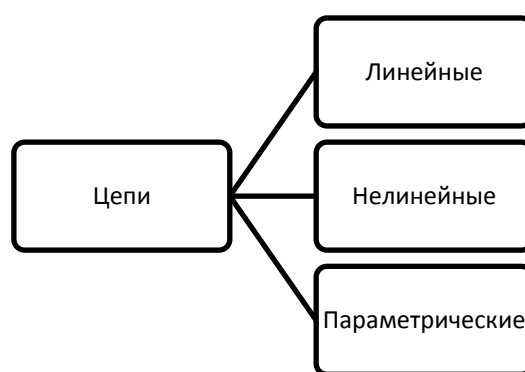


Рис.0.3 Простейшая классификация радиотехнических цепей

0.2. Основы научных парадигм детерминизма, стохастичности и хаоса. Классификация сигналов

Радиотехника как точная наука с ориентацией на конкретное практическое применение имеет свои методические особенности, но в то же время непосредственно связана с развитием общенаучных мировоззренческих парадигм, которые во много определяют направление ее развития.

Методология научного подхода к исследованию любых явлений окружающего нас мира состоит в формировании основанных на опыте абстрактных представлений, адекватность которых проверяется практикой. К сожалению, процессы исследования и познания действительности далеко не просты и основаны на методе «проб и ошибок», в результате которых расширяются границы нашего представления о природе, предлагаются новые концепции, а также отвергаются некоторые ошибочные представления, ранее считавшиеся очевидными и общепризнанными.

Изучая отдельные стороны реальных процессов, мы вынуждены «отделять» их от общей картины мироздания с целью детального изучения отдельных свойств и причинно-следственных связей. Данный процесс «отделения» базируется на сложившихся к текущему моменту системах абстрактных понятий, иногда носящих чисто умозрительный характер. Однако, по меткому изречению А.Эйнштейна «не существует эмпирического опыта без чисто умозрительных понятий и систем, и не существует систем чистого мышления, при более близком изучении которых, не обнаруживался бы эмпирический материал, на котором они строятся».

Процесс абстрагирования при изучении любого явления действительности состоит в выделении наиболее существенных свойств, признаков и связей объекта изучения и представления данных свойств, признаков и связей в форме, позволяющей производить дальнейшие теоретические и экспериментальные исследования. Такое упрощенное представление явления или объекта называют **моделью**. Практическая ценность выводов, получаемых при использовании построенной модели, служит критерием ее адекватности реальному явлению или объекту. В процессе дальнейших исследований и накопления опыта практического использования модель может уточняться, совершенствоваться, а иногда и отвергаться. При этом вопрос цели, ради которой построена та или иная модель, является определяющим при дифференциации существенных и несущественных свойств, признаков и связей. Любая модель является неполной по своему определению и, следовательно, не может описывать все стороны реального явления или объекта. Поэтому рассматривать степень адекватности модели целесообразно с позиций практических целей, для которых она построена. Кроме того следует помнить, что в зависимости от целей для одного и того же объекта могут быть построены разные модели.

В прикладных науках первый этап абстрагирования состоит в качественном описании изучаемого явления или объекта. В зависимости от конкретного типа изучаемого явления или объекта различают модели физических и химических процессов, биологические, экономические, психологические модели и целый ряд других.

Теоретическая дисциплина становится точной наукой, когда она оперирует количественными характеристиками изучаемых явлений и объектов. Придание модели количественных характеристик составляет второй этап абстрагирования при построении модели. Таким образом, получается **математическая модель**, которая является количественной формализацией абстрактных представлений об изучаемом явлении или объекте. Принципиальный шаг в математическом моделировании явлений природы был сделан И.Ньютоном более 300 лет тому назад. Главное достижение Ньютона состоит в предложении использовать для описания различных явлений дифференциальные уравнения. Это было совсем не тривиальным решением, хотя в наше время данный

подход и кажется очевидным. Подход Ньютона предполагает, что изучаемое явление или объект описываются конечным набором чисел, скорость изменения этих чисел зависит состояния объекта в тот же момент времени, а также предполагается возможность описать реальные зависимости гладкими дифференцируемыми функциями.

Согласно традициям, сложившимся в XVIII-XIX веках основным типом математической модели являлись системы дифференциальных уравнений, описывающих некоторый реальный объект или явление. Эти уравнения необходимо составить исходя из общих физических законов, а затем найти их решение.

Под решением понималась некоторая функция, зависящая от времени и начальных условий. В случае невозможности нахождения аналитического решения можно использовать приближенные численные методы.

Переход от первого этапа абстрагирования ко второму, то есть от качественного описания к математической модели зачастую освобождает модель от специфических особенностей, присущих конкретному изучаемому явлению или объекту. В результате, «очистившись» от предметных подробностей, математическая модель приобретает универсальность. Поэтому одна и та же математическая модель может количественно описывать различные по своей физической природе процессы. В этом свойстве математических моделей проявляется практическая ценность формализации предмета исследования, что позволяет во многих случаях при постановке новых практических задач использовать существующий математический аппарат с небольшой модификацией или новой интерпретацией.

Во многих случаях построение количественной модели неразрывно связано с построением математической модели, поэтому выделение двух этапов абстрагирования в процессе построения модели является относительно условным.

Математическая модель может быть представлена различными математическими средствами: дифференциальными или интегральными уравнениями, функциями и функционалами, неравенствами, множествами, функциями распределения вероятностей, логическими выражениями и т.д. Выбор конкретных средств определяется спецификой конкретной предметной области и целью, для которой строится модель.

В наиболее общем виде можно выделить четыре наиболее часто встречающиеся в практической деятельности области, в рамках которых модели обладают схожими свойствами:

1. Модели механических, электрических и радиотехнических процессов и систем,
2. Модели термодинамических и химических процессов и систем,
3. Модели биологических процессов и систем,
4. Модели экономических и социальных процессов и систем.

Условно данная классификация моделей систем представлена в виде соответствующих кластеров на рис.0.4.

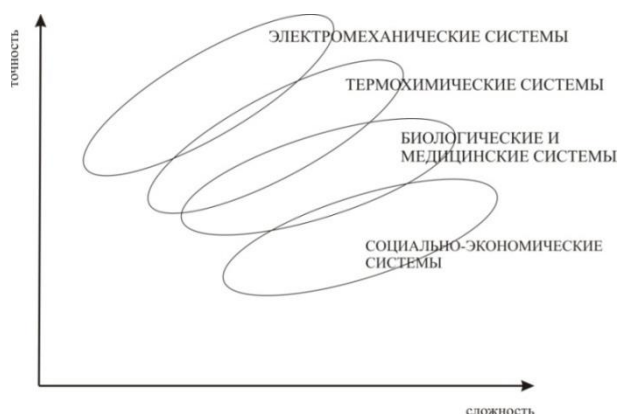


Рис. 0.4. Условная кластеризация моделей систем различной физической природы

Приведенные группы процессов и объектов отличаются по степени «количественности», и уровню потенциально достижимой точности описывающих их моделей (убывание от механических к социальным процессам). При этом размерность моделей, как правило, имеет обратную тенденцию, то есть число

моделируемых параметров увеличивается при движении от механики к социологии. Данные закономерности носят общий характер, и естественно, не отвергают возможности существования простых моделей социальных явлений и сложных моделей механических систем.

Как видно из приведенной классификации, радиотехника может быть отнесена к числу наиболее "количественных" научных дисциплин.

Практическое применение модели наиболее часто сводится к предсказанию состояния исследуемого явления или объекта в будущем (или прошлом), а также к управлению явлением или объектом с целью достижения желаемого результата.

Исторически сложилось так, что после построения Ньютоном классической механики основным принципом, описывающим причинно-следственные связи в физических системах, стал **принцип детерминизма**. Суть данного принципа состоит в том, знание состояния системы в данный момент времени позволяет определить ее состояние в любой другой момент времени. Принцип физического детерминизма оказал серьезное влияние не только на физические исследования, но и послужил поводом для широкого обсуждения общих философских проблем естествознания. Уверенность в возможностях использования принципа детерминизма очень ярко и красиво была выражена Пьером Лапласом. «Разум, который в каждый заданный момент времени знал бы все движущие силы природы, так же как и относительное положение всех составляющих ее элементов, и вдобавок был бы столь всеобъемлющим, чтобы подвергнуть анализу эти данные, мог бы выразить единым соотношением как движение величайших тел мира, так и движение мельчайших атомов: ничто не осталось бы для него неизвестным, и он мог бы обозревать одним взглядом как будущее, так и прошлое».

Описанный Лапласом «Разум», получивший в дальнейшем имя «лапласовского демона» связывал принцип детерминизма со всем мирозданием, которое рассматривалось как единая замкнутая система. На практике исследуемое явление или объект рассматривают как замкнутую систему, на которую внешняя среда действует известной совокупностью входных величин.

Принцип физического детерминизма позволил решить большое число прикладных задач классической механики и надолго занял лидирующее положение среди общенаучных концепций познания.

Дальнейшие исследования в квантовой механике показали ограниченность области применения детерминизма, что нашло свое выражение в принципе неопределенности Гейзенберга, согласно которому невозможно одновременно и с одинаковой точностью определить положение и скорость материальной частицы, так как

$$\Delta x \Delta v \geq h/m,$$

где Δx – ошибка измерения положения,
 Δv – ошибка измерения скорости,
 m – масса частицы,
 h – постоянная Планка.

Аналогичное соотношение имеет место в теории сигналов и систем: неопределенность в измерении частоты обратно пропорциональна времени ее измерения

$$\Delta t \Delta f \geq 1,$$

где Δt – длительность сигнала,
 Δf – ширина полосы частот сигнала.

Принцип неопределенности, в более широком виде иногда называемый принципом физического индетерминизма, является альтернативой принципу физического детерминизма. Концептуальное противостояние принципов детерминизма и индетерминизма затронуло не только область естественных наук, но нашло свой отклик во всех областях человеческой деятельности, включая философию и политику. Наиболее радикальные сторонники индетерминизма пытались трактовать его как эмпирическое доказательство философского индетерминизма, отрицающего возможность адекватного отображения объективной реальности в сознании человека. При таких трактовках принципа физического индетерминизма происходит неявная подмена понятий, так как физический индетерминизм не отрицает возможности процесса адекватного познания действительности, а лишь указывает на зону принципиальной неопределенности, которая реально существует из-за ограничений разрешающей способности имеющихся экспериментальных средств.

Принцип физического индетерминизма имеет негативное свойство (особенно тяжело воспринимаемое психикой человека), состоящее в том, что любое наблюдение за происходящим явлением или объектом трактуется как событие с непредсказуемым однозначно результатом. В отличие от принципа физического детерминизма, позволяющего при многократном повторении эксперимента в сохраняющихся условиях ожидать полного совпадения результатов, принцип индетерминизма указывает лишь на ситуацию неопределенности, которая возникает в результате эксперимента.

Понятно, что степень неопределенности результатов для каждого явления или объекта будет своя. Так, например, возвращаясь к приведенной выше классификации процессов и систем, заметим, что в механических и электрических системах неопределенность будет, как правило, самой незначительной, в термодинамических и химических системах более высокой, и, наконец, достигнет максимума в биологических, социальных и экономических процессах. Кроме того значимость степени неопределенности зависит от цели, которую ставит перед собой исследователь.

Неопределенность присуща всем наблюдениям за явлениями окружающей действительности, но в ряде практических применений этой неопределенностью можно пренебречь, считая, что при многократном повторении некоторого наблюдения за системой при неизменных условиях получается один и тот же результат. Понятно, что данное допущение является абстрактным упрощением ситуации, но в ряде случаев оно дает желаемый результат при существенном упрощении рассматриваемой модели.

Процесс, система или объект называются **детерминированными**, если априорно известно, что при многократном повторении эксперимента в идентичных условиях будет наблюдаться один и тот же однозначный результат.

Процесс, система или объект называются **стохастическими**, если априорно известно лишь множество возможных результатов, которые могут наблюдаться в результате эксперимента, многократно повторяемого в идентичных условиях. Таким образом, результат наблюдения за стохастическим явлением достоверно предсказать невозможно. Такое заключение может показаться смертным приговором для практического использования моделей стохастических систем. Однако все далеко не так уж страшно, причем, именно с точки зрения практического использования. Дело в том, что если наблюдать достаточно длинные серии экспериментов над стохастическими системами при сохранении условий эксперимента, то во многих случаях оказывается, что несмотря на разное поведение каждой отдельной реализации эксперимента в среднем проявляется некоторая устойчивость результатов. Данная устойчивость средних результатов является статистической закономерностью, позволяющей использовать для построения моделей таких систем математический аппарат теории вероятностей, основоположником которой является А.Н.Колмогоров.

В физике вероятностный подход впервые использовался в исследованиях по термодинамике Л.Больцмана и Дж.Гиббса.

Исследования в области стохастических систем происходили очень интенсивно и в различных областях науки, благодаря чему в нашем распоряжении имеется широкий арсенал методов их моделирования.

Успехи компьютерной техники создали мощные и быстрые вычислительные системы, которые, казалось бы, вот-вот смогут выполнить роль «лапласовского демона», но этого оказалось недостаточно. Спор между случайностью и детерминизмом продолжается.

Классическая теорема о непрерывности и дифференцируемости решений дифференциальных уравнений по начальным условиям вселяла оптимизм в вопросах прогнозирования поведения исследуемых явлений на бесконечно большие интервалы времени вперед. Действительно, если мы немного ошибемся в описании прошлого состояния, то и в будущем ошибка будет небольшой. Кроме того в большинстве практических применений временное поведение исследуемого объекта чаще всего тоже является простым – колебательным или стационарным.

К сожалению, все вышесказанное относится только к так называемым *линейным* явлениям и объектам, то есть объектам описываемым системами линейных уравнений, для которых справедлив принцип суперпозиции.

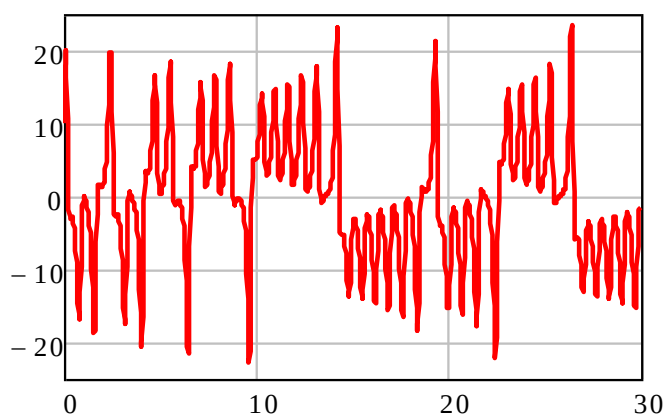


Рис.0.5. Хаотический сигнал модели Лоренца.

Исследования в области нелинейных систем показали, что ключевым моментом в точности прогноза (то есть соответствия модели и моделируемого явления на значительном временном интервале) является чувствительность к начальным условиям, вследствие эффекта разбегания двух бесконечно близких траекторий для класса так называемых хаотических аттракторов. В силу чувствительности к начальным условиям нельзя сравнивать траекторию объекта и модели

«поточечно», так как сколь угодно малая ошибка в начальный момент времени будет экспоненциально возрастать. Поэтому либо приходится ограничиваться краткосрочным прогнозированием, либо изыскивать адекватные способы сравнения поведения модели и объекта. После того как возможность подобной ситуации была осознана, начался новый виток в развитии взглядов на детерминированное и случайное. Возникли новые термины «динамический хаос» и «странный аттрактор». Оказалось, что хаотическое поведение может наблюдаться в относительно простых системах, состоящих всего из трех обыкновенных дифференциальных уравнений. Активное развитие данного направления началось с публикаций Э.Лоренса в 1963 году и продолжается до настоящего времени. Пример такого типа сигналов представлен на рис.0.5.

Поскольку радиотехника проникла практически во все области человеческой деятельности, то совершенно очевидно, что с ее помощью пытаются передовать информацию о самых разнообразных физических процессах. Поэтому классификация радиотехнических сигналов на сегодняшний день должна включать в себя как сигналы, свойственные собственно радиотехнике, так и сигналы, используемые в тех областях жизнедеятельности человека, которые радиотехника обслуживает. С этой точки зрения,

упрощенную классификацию сигналов можно предложить в виде, представленном на рис.0.6.

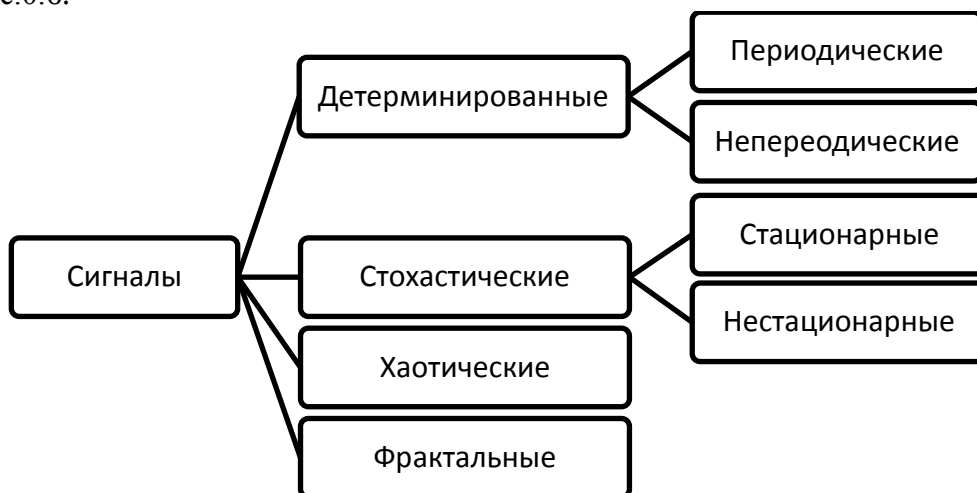


Рис.0.6. Классификация сигналов

Более подробно вопросы, касающиеся радиотехнических сигналов, будут рассмотрены в последующих разделах. Хаотические и фрактальные сигналы в данном курсе рассматриваться не будут.

В рамках настоящего курса для целей моделирования и анализа радиотехнических сигналов и систем мы будем использовать компьютерные программы Matlab, Simulink, MathCad и Multisim. Предполагается, что изучающие имеют первоначальные навыки работы в указанных программах, поэтому описание пользовательских интерфейсов этих программных продуктов приводиться не будет. Дополнительные навыки по работе с программами могут быть получены во время аудиторных практических занятий или самостоятельно.

Приводимые ниже примеры носят ознакомительный и иллюстративный характер, поэтому комментарии в них практически отсутствуют. В дальнейшем компьютерное моделирование будет проводиться после изучения соответствующего теоретического материала, и сопровождаться необходимыми пояснениями в ходе лекционных и практических занятий.

Пример 0.1.

Генерация и визуализация простейших тестовых сигналов в программе MATLAB. Проведем генерацию трех сигналов: синусоидальные колебания, полисинусоидальное колебание, полисинусоидальное колебание с шумовой помехой различной мощности. m-файл для выполнения данного упражнения приведен на рис.0.7

```

t=(0:0.01:5);
y1=sin(2*pi*1*t);
y2=sin(2*pi*1*t)+0.3*sin(2*pi*4*t);
y3=sin(2*pi*1*t)+0.3*sin(2*pi*4*t)+0.5*randn(size(t));
y4=sin(2*pi*1*t)+0.3*sin(2*pi*4*t)+1.5*randn(size(t));
subplot(2,2,1);plot(t,y1);grid on;
subplot(2,2,2);plot(t,y2);grid on;
subplot(2,2,3);plot(t,y3);grid on;
subplot(2,2,4);plot(t,y4);grid on;
  
```

Рис.0.7. Генерация и визуализация простых тестовых сигналов с применением MATLAB

Результаты моделирования приведены на рис.0.8.

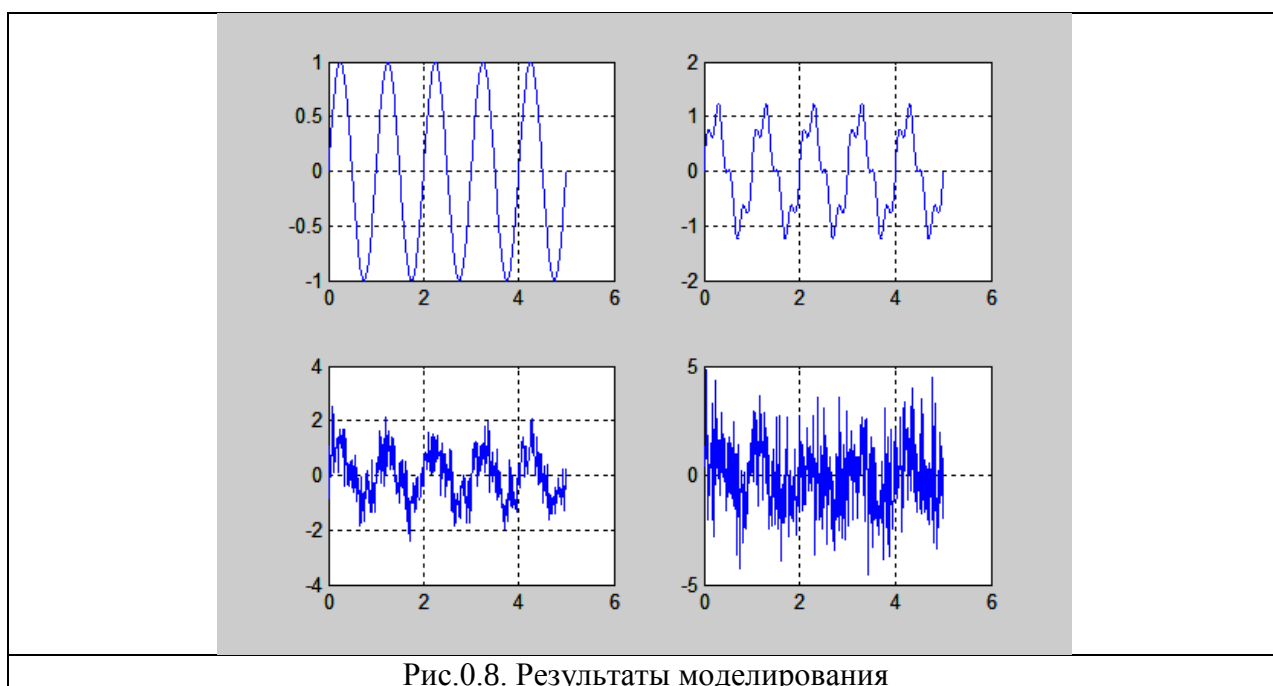


Рис.0.8. Результаты моделирования

Пример 0.2.

Генерация и визуализация функции единичного скачка с помощью программы MathCad. Листинг рабочей сессии приведен на рис.0.9.

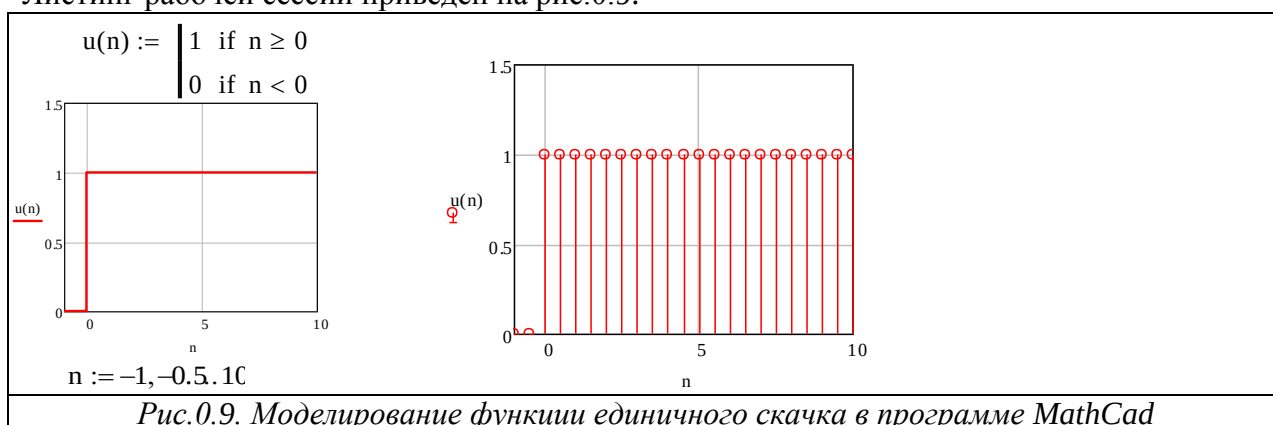
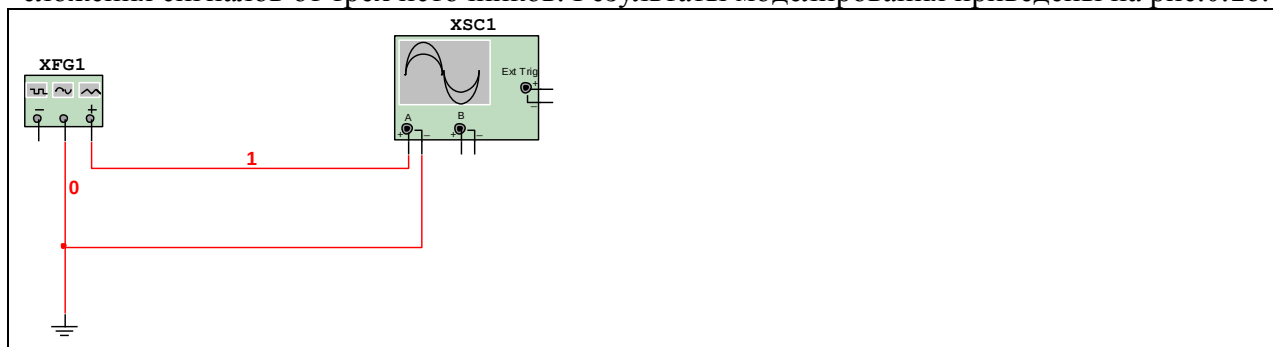


Рис.0.9. Моделирование функции единичного скачка в программе MathCad

Пример 0.3.

Генерация и визуализация сигналов с применением Multisim. В первой части данного примера продемонстрированы возможности генерации сигналов с использованием блока функционального генератора, а во второй – генерация полигармонического сигнала путем сложения сигналов от трех источников. Результаты моделирования приведены на рис.0.10.



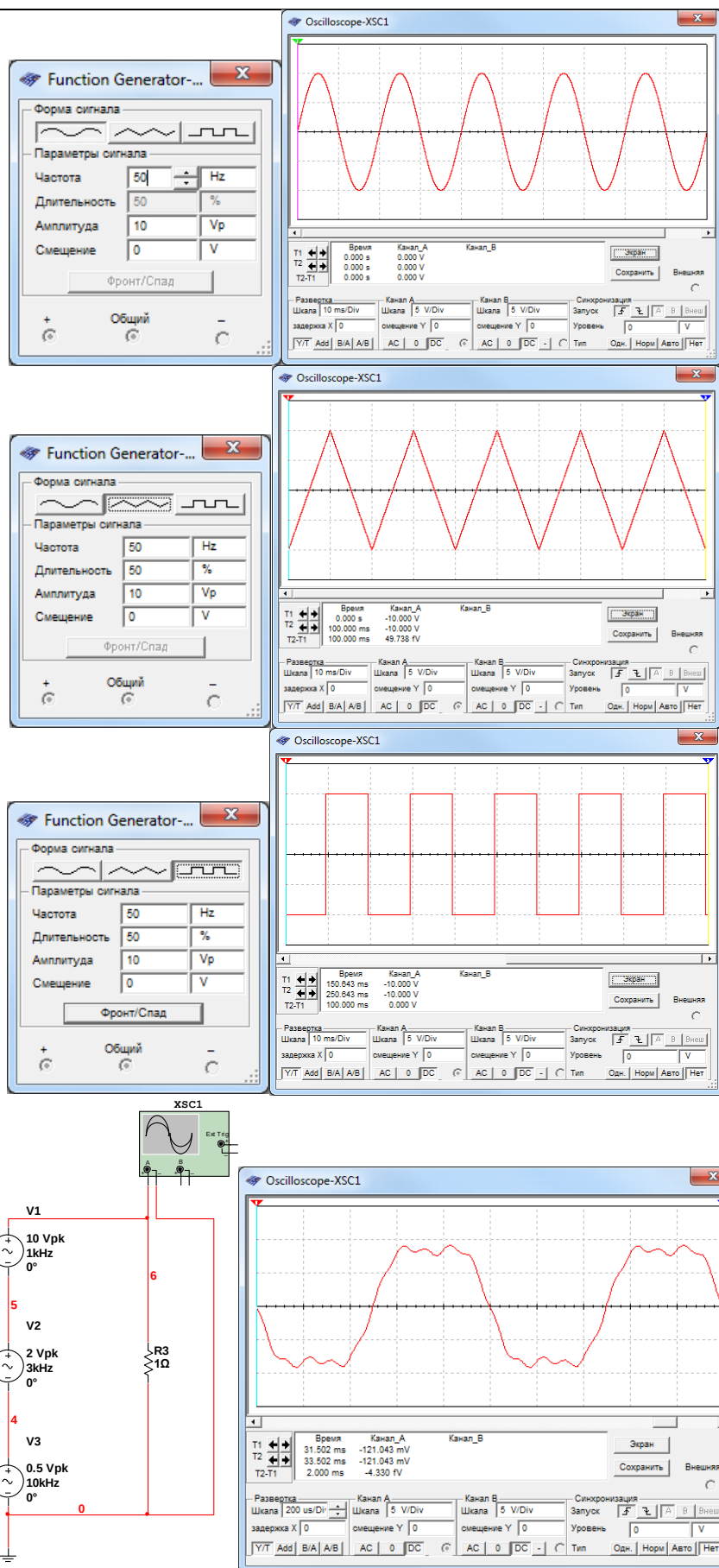


Рис.0.10. Генерация и визуализация сигналов с помощью программы Multisim.

Пример 0.4.

Генерация случайного сигнала с помощью программы Simulink (см. рис.0.11)

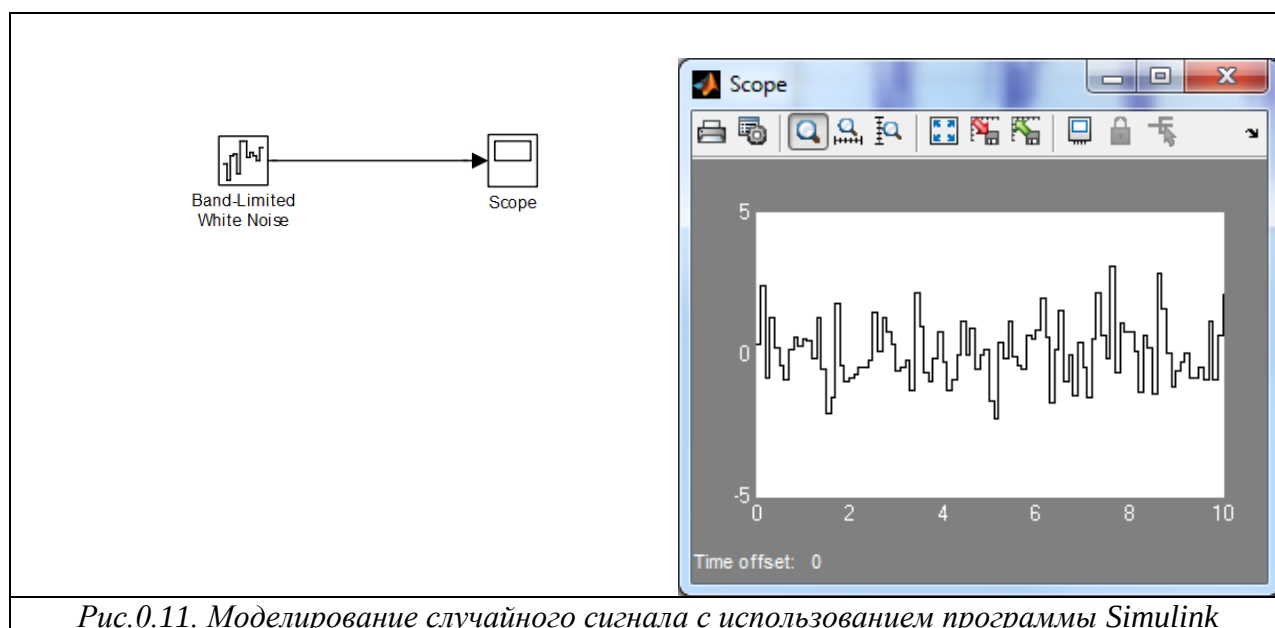


Рис.0.11. Моделирование случайного сигнала с использованием программы Simulink

0.3. Вопросы для самоконтроля и повторения

1. Какая область науки и техники относится к радиотехнике?
2. В чем состоит отличие радиотехники от электротехники и электросвязи?
3. Какие виды радиосвязи вам известны? В чем состоит их различие?
4. Перечислите основные задачи радиолокации, радионавигации, радиотелеметрии и радиоуправления.
5. Изобразите обобщенную схему радиотехнической системы.
6. Какие типы помех могут присутствовать в радиотехнической системе?
7. Приведите простейшую классификацию радиотехнических цепей.
8. Какие виды моделей используются при изучении явлений реального мира?
9. Для какой цели используются математические модели?
10. В чем состоит принцип детерминизма?
11. Какие сигналы называются детерминированными?
12. Какие сигналы называются стохастическими?

ГЛАВА 1. Детерминированные сигналы

1.1. Сигналы и информация

Слово «сигнал» является столь широко распространенным в настоящее время, что в большинстве случаев оно используется, основываясь на интуитивных понятиях о его значении. При таком бытовом использовании практически становятся синонимами и сливаются в одно понятие «сигнал», «сообщение», «данные» и «информация».

В рамках нашего курса мы не будем ставить задачу глобального философского осмысления указанных понятий, но введем некоторую ясность в разграничение значений этих терминов, по крайней мере для того, чтобы при дальнейшем изложении материала не возникало противоречий и непонимания.

Согласно толковому словарю русского языка *информация* – это сведения об окружающем мире и протекающих в нем процессах, воспринимаемые человеком или специальным устройством.

Закон РФ "Об информации, информационных технологиях и защите информации" слово *информация* определяет как сведения о людях, предметах, фактах, событиях и процессах, независимо от формы их представления.

В справочниках и словарях можно найти и другие определения:

-*информация* (лат. *informatio* — разъяснение, изложение, осведомлённость) — одно из наиболее общих понятий науки, обозначающее некоторые сведения, совокупность каких-либо данных, знаний и т. п.

-*информацией* (от лат. *informatio* — «научение», «сведение», «оповещение») называется продукт взаимодействия данных и методов, рассмотренный в контексте этого взаимодействия,

-*информация* — это универсальное свойство материи, представляющее собой распространение в пространстве и времени содержания объектов (явлений) действительности посредством объективно существующих носителей различной природы.

В технических дисциплинах с позиций детерминизма информация понимается как сведения об объекте, а с вероятностных позиций как снятие неопределенности. Подходы к определению количества информации также разделились на алгоритмический А.Н.Колмогорова и вероятностный К.Шеннона. Основоположник кибернетики Н.Винер подчеркивал, что передача информации возможна лишь как передача альтернатив.

Термин *данные* (от лат. *datum* – факт) - это совокупность фактов, результатов наблюдений, измерений о каких-либо объектах, явлениях или процессах материального мира, представленных в формализованном виде, количественном или качественном. Данные всего лишь атрибут информации - сырье для получения информации путем соответствующей обработки и интерпретации (истолкования).

Одной из важнейших форм представления информации являются знаки или символы. Под знаком обычно понимают чувственно-воспринимаемый предмет (или процесс), вводимый в процессе познания и общения, используемый в качестве заместителя другого предмета для хранения, передачи, получения и преобразования информации.

Совокупность знаков, представляющих определенную информацию, называют *сообщением*.

Системы, которые функционируют на основе использования информации о событиях, происходящих вне и внутри рассматриваемой системы, называются *информационными системами*.

Функции информационной системы могут быть различными, включая формирование, передачу, прием, хранение и обработку сообщений. Данные функции выполняются посредством технических средств, использующих подходящие физические носители информации. Примерами физических процессов, отображающих сообщения, могут быть электрический ток, акустические колебания, оптические волны и т.п.

Сигналами (от лат signal, то есть «знак») называются физические процессы или явления, несущие информацию (используемые для передачи сообщений на расстояние).

В кибернетике выделяют четыре компонента сигнала:

- физический носитель,
- форма выражения (синтаксический аспект информации),
- интерпретация смысла (семантический аспект информации),
- правила приписывания различного смысла одному и тому же сообщению (прагматический аспект информации).

Единицу количественной меры информации - БИТ (сокращение binary digit - двоичная цифра), впервые предложил Р. Хартли в 1928 году. 1 бит - это информация о двух возможных равновероятных состояниях объекта соответствующая неопределенности выбора из двух равновероятных событий. Математически это соответствует состояниям 1 или 0 одного разряда двоичной системы счисления. Количество информации H (в битах), необходимое и достаточное для полного снятия неопределенности о состоянии объекта, который имеет N равновозможных состояний, измеряется как логарифм по основанию 2 из числа возможных состояний:

$$H = \log_2 N$$

Соответственно, двоичный числовой информационный код одного из N возможных состояний объекта занимает H двоичных разрядов.

Основание логарифма не имеет принципиального значения и определяет только масштаб или единицу неопределенности. Двоичная мера информации получила общее признание в связи с простотой реализации информационной техники на элементах с двумя устойчивыми состояниями. В десятичном исчислении единицей информации является один десятичный разряд - ДИТ. Иногда при проведении математических преобразований оказывается более удобным пользоваться натуральными логарифмами. В этом случае информация измеряется в натуральных единицах или НАТах.

Информация, передаваемая по каналу связи или извлекаемая в результате измерения, заключена в сигнале. Для оценки **информационной емкости** сигнала должна быть установлена связь между параметрами сигнала и количеством информации, которое можно передать с помощью данного сигнала.

Прирост количества информации равен:

$$I = \log_2 \frac{p_1}{p_2} = \log_2 p_1 - \log_2 p_2$$

где p_1 – априорная вероятность события, p_2 – апостериорная вероятность.

В более общем случае степень неопределенности состояния объекта зависит не только от числа его возможных состояний, но и от вероятностей этих состояний. Если источник информации однозначно и полно характеризуется ансамблем состояний $U = \{u_1, u_2, \dots, u_N\}$ с вероятностями каждого из состояний $\{p(u_1), p(u_2), \dots, p(u_N)\}$ соответственно (при условии, что сумма вероятностей всех состояний равна 1), то в качестве меры количества информации, как неопределенности выбора дискретным источником состояния из ансамбля U , может использоваться предложенная К. Шенноном в 1946 году величина, которая получила название **энтропии дискретного источника информации**

или энтропии конечного ансамбля:

$$H(U) = - \sum_{n=1}^N p_n \log_2 p_n$$

Мера энтропии Шеннона является обобщением меры Хартли на случай ансамблей с **неравновероятными состояниями**, в чем нетрудно убедиться, если в приведенном выражении значение p_n заменить значением $p=1/N$ для ансамбля равновероятных состояний. **Энтропия конечного ансамбля $H(U)$ характеризует неопределенность, приходящуюся в среднем на одно состояние ансамбля.**

Отметим основные свойства энтропии:

1. Энтропия является вещественной и неотрицательной величиной. (Так как значения вероятностей p_n находятся в интервале 0-1, то значения $\log_2 p_n$ всегда отрицательны, и соответственно $-p_n \log_2 p_n$ всегда положительно).

2. Энтропия - величина ограниченная.

3. Если вероятность одного из состояний источника информации равна 1, то есть состояние источника полностью определено (вероятности остальных состояний источника равны нулю, т.к. сумма вероятностей должна быть равна 1), то энтропия равна 0.

4. Энтропия максимальна при равной вероятности всех состояний источника информации:

$$H_{max}(U) = - \sum_{n=1}^N \frac{1}{N} \log_2 \left(\frac{1}{N} \right) = \log_2 N$$

Например, энтропия источника с двумя состояниями u_1 и u_2 с вероятностями $p(u_1) = p$ и $p(u_2) = 1 - p$ определяется выражением:

$$H(U) = -[p \log_2 p + (1 - p) \log_2 (1 - p)]$$

и при изменении p в пределах от 0 до 1, достигает максимума при равенстве вероятностей.

5. Энтропия объединенных статистически независимых источников информации равна сумме их энтропий.

Поскольку сигналы предназначены для передачи информации, то всегда возникает естественный вопрос об их информационной емкости, которая существенно зависит от типа сигналов и определяет требования к каналам передачи данных (каналам связи). Одновременно с этим технические характеристики (или технические ограничения) каналов связи определяют требования к информационной емкости сигналов, передаваемых по этим каналам.

Для каналов передачи дискретных сигналов (дискретные канала связи) используют понятия технической и информационной скорости передачи данных.

Под технической скоростью передачи V_t подразумевают число элементарных сигналов (символов), передаваемых по каналу в единицу времени. Единицей измерения технической скорости служит БОД (один символ в секунду). Полоса пропускания канала связи обычно ограничивается определенной предельной частотой $F_{пред}$ по уровню затухания сигнала до уровня статистических помех, при этом значение технической скорости передачи данных, естественно, не может быть выше $F_{пред}$ без каких-либо специальных устройств выделения информационных сигналов.

При известной технической скорости V_t скорость передачи информации V_h измеряется в битах в секунду, и при уровне помех меньше амплитудных значений символьных импульсов задается соотношением:

$$V_h = V_t H(s),$$

где $H(s)$ – энтропия символа.

Для двоичных дискретных символов с двумя возможными состояниями $[0, 1]$ значение $H(s)$ равно 1. При увеличении числа возможных состояний до L значение $H(s)$ равно $\log_2 L$.

Информационная емкость сигнала или полное количество информации в сигнале S определяется полным количеством $N = t/T$ энтропии символов в битах на интервале задания сигнала t :

$$I_t(s) = N \log_2 L = \left(\frac{t}{T}\right) \log_2 L,$$

где T – интервал передачи одного символа.

Увеличение числа уровней L увеличивает пропускную способность каналов связи, но усложняет аппаратуру кодирования данных и снижает помехоустойчивость связи.

Для непрерывных сигналов передача по каналам связи возможна только при условии, что максимальная информационная частота в сигнале F_{max} не превышает предельной частоты $F_{пред}$ передачи сигналов каналом связи. Для оценки предельной информационной емкости непрерывного сигнала необходимо выполнить его дискретизацию с предельным интервалом $\tau = 1/2F_{max}$, допускающим восстановление сигнала без потери информации. Тогда если принять полную длительность сигнала за t_s , число отсчетов будет составлять

$$N = \frac{t_s}{\tau} = 2F_{max} t_s$$

Определим максимально возможное число выборок в каждом отсчете при наличии шума в канале со средней мощностью $P_{ш} = \delta^2$. При средней мощности сигнала $P_s = s^2$:

$$L = \sqrt{\frac{s^2 + \delta^2}{\delta^2}} = \sqrt{1 + \frac{s^2}{\delta^2}}$$

Тогда информационная емкость сигнала может быть определена как

$$I(s) = 2F_{max} t_s \log_2 L,$$

где $L = \sqrt{\frac{s^2 + \delta^2}{\delta^2}} = \sqrt{1 + \frac{s^2}{\delta^2}}$ определяется степенью превышения средней мощности сигнала s^2 над мощностью шумов δ^2 , имеющихся в канале связи.

Информационные возможности сигнала возрастают с расширением его спектра и превышением его уровня над уровнем помех.

Пример 1.1.

Для источника информации с двумя состояниями u_1 и u_2 , которым соответствуют вероятности появления $p(u_1)=p$ и $p(u_2)=1-p$ соответственно, построить график характеризующий изменение энтропии при изменении значения p в пределах от 0 до 1 включительно. Определить вероятности состояний $p(u_1)=p_{max}$ и $p(u_2)=1-p_{max}$, при которых достигается максимум энтропии.

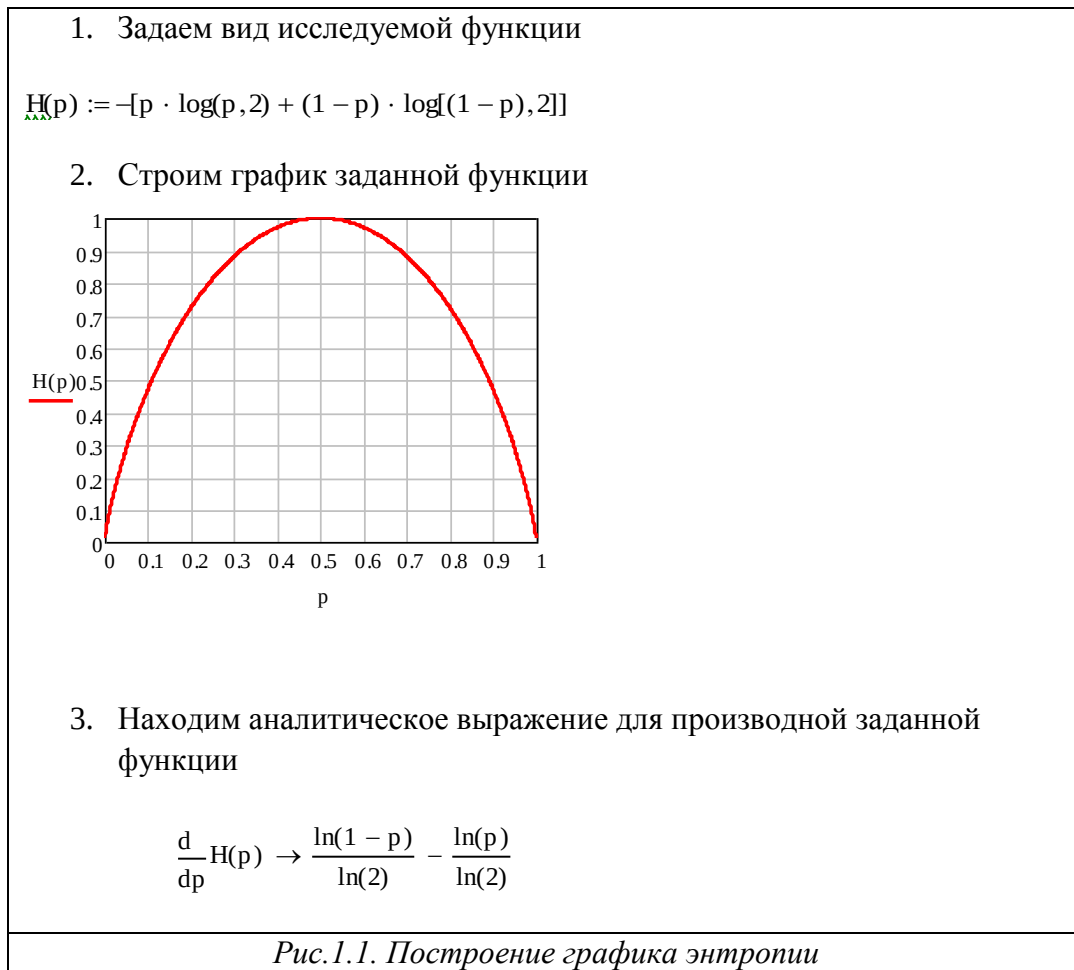
Решение.

Воспользуемся программой MathCad, зададим функцию

$$H(U) = -[p \log_2 p + (1 - p) \log_2 (1 - p)]$$

и построим ее график, задав пределы изменения аргумента от 0 до 1, как показано на

прилагаемом листинге рабочей сессии (рис.1.1.).



Из приведенного графика наглядно видно, что максимальное значение энтропии $H(p)=1$ достигается при значении $p_{max}=0.5$ то есть при условии $p=1-p$, которое соответствует случаю равной вероятности состояний u_1 и u_2 .

Для аналитического нахождения максимума функции $H(p)$ воспользуемся возможностями символьных вычислений и определим производную от $H(p)$, как показано в п.3. приведенного листинга. Дальнейшие действия по составлению уравнения $\frac{dH(p)}{dp} = 0$ и нахождению его решения тривиальны.

В качестве дополнительного задания рекомендуем провести аналитический расчет производной от $H(p)$ самостоятельно, сравнить свой результат с формулой, полученной выше, а также объяснить возникшую разницу в форме представления результата при вычислении производной.

1.2. Пространства сигналов

Математическая модель сигнала, в самом общем смысле, это зависимость одной величины y (именуемой зависимой переменной) от другой величины x (именуемой независимой переменной) и обозначаемой в виде

$$y=y(x) \quad \text{или} \quad y=f(x),$$

где $f(.)$ - общее обозначение функциональной зависимости.

Понятие функции является очень удобной математической моделью сигнала. При этом чаще всего рассматривается случай, когда независимая переменная – это время t , хотя строго говоря, это не обязательное условие. Независимая переменная вполне может быть, например, пространственной координатой, как это бывает в оптических или локационных системах.

Таким образом, математическая модель сигнала устанавливает соответствие между любым моментом времени и величиной сигнала

$$y=y(t), \quad (1.1)$$

где $t \in T$, T - область определения сигнала и величиной сигнала,

$y \in Y$, Y - множество возможных значений сигнал.

Функция, входящая в выражение (1.1) может быть как скалярной $y(t)$, так и векторной $y(t)$.

В зависимости от вида области определения сигнала T сигналы подразделяются на **аналоговые** (для которых T – континуум) и **дискретные** (для которых T – конечное или счетное множество).

Если конечным или счетным являются оба множества T и Y , то такой сигнал называется **цифровым**. Иными словами, цифровой сигнал – это квантованный по амплитуде дискретный сигнал.

В радиотехнических системах часто используются сигналы, зависящие от другого сигнала $s(t)$, называемого **модулирующим** сигналом

$$y=y(t,s(t)). \quad (1.2)$$

При такой зависимости если $s(t)$ является аналоговым сигналом, то $y(t,s(t))$ называют **модулированным** сигналом, если $s(t)$ – дискретный сигнал, то $y(t,s(t))$ называют **манипулированным** сигналом. Более подробно эти виды сигналов и способы технической реализации зависимостей вида (1.2) будут рассмотрены в следующих разделах курса.

В теории информационных систем используются не только свойства отдельных сигналов, но и целых классов сигналов. Такой подход заимствован из функционального анализа с заменой понятий пространства функций понятием пространства сигналов, отличие которых лишь терминологическое. Понятие пространства, то есть множества, элементами которого являются отдельные сигналы, требует от элементов (сигналов) некоторых дополнительных свойств.

Таким образом, определяются **метрические** пространства, **линейные** (или векторные) пространства, **нормированные** пространства, **унитарные** пространства (пространства со скалярным произведением).

Метрическое пространство – это пространство, в котором каждой паре сигналов $x(t) \in X$ и $y(t) \in X$ поставлено вещественное число $\rho(x,y)$, называемое метрикой и обладающее следующими свойствами:

1. $\rho(x, y) \geq 0$, $\rho(x, y)=0$ тогда и только тогда, когда $x=y$;
2. $\rho(x, y)=\rho(y, x)$ – (условие симметрии);
3. $\rho(x,z) \leq \rho(x,y)+\rho(y,z)$ – для $\forall z \in X$ – (неравенство треугольника).

Метрика по своей сути является расстоянием между двумя элементами множества X и может быть определена разными способами.

Для кусочно-непрерывных на интервале времени $(0,T)$ сигналов $x(t) \in X$ и $y(t) \in X$ наибольшее распространение получила метрика вида

$$\rho(x, y) = \sqrt{\int_0^T (x(t) - y(t))^2 dt},$$

которая является частным случаем более общего варианта определения метрики

$$\rho(x, y) = \sqrt[p]{\int_0^T |x(t) - y(t)|^p dt} \quad (1.3)$$

при значении $p=2$. Иногда также используются метрики вида (1.3) со значениями $p=1$ и $p=\infty$, что приводит следующим видам метрик

$$\rho(x, y) = \int_0^T (x(t) - y(t))^2 dt,$$

$$\rho(x, y) = \max_{t \in T} |x(t) - y(t)|.$$

Линейное (или векторное) пространство – это пространство сигналов, в котором определены операции сложения и умножения на скаляр, то есть пространство сигналов наделенных свойствами векторов.

В линейном пространстве для любых сигналов $x(t) \in X$ и $y(t) \in X$ определена сумма сигналов $(x(t)+y(t)) \in X$ и для любого действительного числа α определено произведение $\alpha x(t) \in X$, при этом выполняются следующие свойства:

1. $x(t)+y(t) = y(t)+x(t)$ – коммутативность,
2. $(x(t)+y(t))+z(t) = x(t)+(y(t)+z(t))$ – для $\forall z(t) \in X$ – ассоциативность,
3. Существует единственный нулевой элемент $O \in X$, для которого $O+x(t)=x(t)$ и $O.x(t)=O$ для $\forall x(t) \in X$,
4. для $\forall x(t) \in X$ существует единственный $-x(t)$, такой что $x(t)+(-x(t))=O$,
5. $\alpha(\beta x(t))=(\alpha\beta)x(t)$ – для любых скаляров α, β – ассоциативность,
6. $\alpha(x(t)+y(t)) = \alpha x(t)+\alpha y(t)$ – дистрибутивность векторная,
7. $(\alpha+\beta)x(t) = \alpha x(t)+\beta x(t)$ – дистрибутивность скалярная.

Нормированное линейное пространство – это линейное пространство, в котором определена норма (размер) каждого элемента. Норма сигнала $x(t)$ обозначается символом $\|x(t)\|$ и ставит в соответствие всем элементам пространства соответствующие вещественные числа, такие, что выполняются следующие условия:

1. $\|x(t)\| \geq 0$ для $\forall x(t)$, $\|x(t)\|=0$ тогда и только тогда, когда $x(t)=O$,
2. $\|x(t)+y(t)\| \leq \|x(t)\| + \|y(t)\|$ для любых $x(t) \in X$ и $y(t) \in X$ – (неравенство треугольника),
3. $\|\alpha x(t)\| = |\alpha| \cdot \|x(t)\|$ для любого $x(t) \in X$ и скаляра α .

Норма аналогично метрике может быть определена различными способами, например, норма может быть определена в виде

$$\|x(t)\| = \sqrt[p]{\int_0^T (x(t))^p dt}, \quad (1.4)$$

что для случая $p=2$ выражение (1.4.) дает

$$||x(t)|| = \sqrt{\int_0^T (x(t))^2 dt}$$

для действительной функции $x(t)$ и

$$||x(t)|| = \sqrt{\int_0^T |x(t)|^2 dt} = \sqrt{\int_0^T x(t) x^*(t) dt},$$

где $x^*(t)$ – функция комплексно сопряженная $x(t)$.

Унитарное пространство (или пространство со скалярным произведением) – это линейное пространство, для упорядоченных пар которого определена операция скалярного произведения. Эта операция для любой упорядоченной пары сигналов $x(t) \in X$ и $y(t) \in X$ обозначается как $(x(t), y(t))$ или $\langle x(t), y(t) \rangle$ и отображает указанную упорядоченную пару сигналов в комплексное число, так, чтобы выполнялись следующие условия:

1. $\langle x(t), y(t) \rangle = \langle y(t), x(t) \rangle^*$, где знак $*$ обозначает комплексное сопряжение,
2. $\langle \alpha x(t) + \beta y(t), z(t) \rangle = \alpha \langle x(t), z(t) \rangle + \beta \langle y(t), z(t) \rangle$ для любых скаляров α, β и сигналов $x(t) \in X, y(t) \in X, z(t) \in X$,
3. $\langle x(t), x(t) \rangle \geq 0$, причем $\langle x(t), x(t) \rangle = 0$ тогда и только тогда, когда $x(t) = 0$.

Для кусочно-непрерывных на интервале времени $(0, T)$ сигналов $x(t) \in X$ и $y(t) \in X$ скалярное произведение вводится следующим образом:

$$\langle x(t), y(t) \rangle = \int_0^T x(t) y^*(t) dt$$

Из приведенных выше соотношений видно, что на основе скалярного произведения можно ввести норму сигнала в виде

$$||x(t)|| = \sqrt{\langle x(t), x(t) \rangle},$$

а введенная таким способом норма в свою очередь порождает метрику вида

$$\rho(x(t), y(t)) = ||x(t) - y(t)|| = \sqrt{\langle x(t) - y(t), x(t) - y(t) \rangle}.$$

Скалярное произведение позволяет аналогично векторам ввести понятие **ортгональных** сигналов, скалярное произведение которых равно нулю

$$\langle x(t), y(t) \rangle = 0.$$

С понятием скалярного произведения также связано важное неравенство, чаще всего называемое неравенством Шварца

$$|\langle x(t), y(t) \rangle|^2 \leq \langle x(t), x(t) \rangle \langle y(t), y(t) \rangle.$$

Полное нормированное пространство, т.е. такое, в котором всякая фундаментальная последовательность имеет предел, называется **банаховым**. Банахово пространство, в котором выполняются свойства скалярного произведения, называется вещественным **гильбертовым** пространством.

На практике приходится иметь дело с сигналами энергия, которых на наблюдаемом отрезке времени $(0, T)$ ограничена. Пространство таких сигналов обозначают

символом $L^2(T)$ и часто называют пространством функций суммируемых с квадратом. Если для такого пространства определено показанным выше способом скалярное произведение, то пространство становится *гильбертовым*.

1.3.Декомпозиция аналогового сигнала

Многие методы анализа сигналов построены на замене исходного сигнала суммой других сигналов, математическая обработка которых более проста или представление которых более полно раскрывает какой-либо информационный аспект исходного сигнала.

Данная операция особенно эффективна при анализе прохождения сигнала через линейные цепи и системы, так как реакция этих систем на сумму воздействий равна сумме реакций на каждое из воздействий в отдельности. Данное свойство носит название линейной суперпозиции.

В предыдущем разделе были рассмотрены функциональные пространства аналоговых сигналов, для которых может быть построен базис пространства в виде бесконечной последовательности (системы) функций

$$\varphi_0(t), \varphi_1(t), \dots, \varphi_n(t), \dots \quad (1.5)$$

Данная система будет являться ортогональной на отрезке $[t_1, t_2]$, если

$$\int_{t_1}^{t_2} \varphi_n(t) \varphi_m(t) dt = 0 \text{ при } n \neq m, \quad (1.6)$$

для случая действительных функций $\varphi_n(t)$ и

$$\int_{t_1}^{t_2} \varphi_n(t) \varphi_m^*(t) dt = 0 \text{ при } n \neq m,$$

для случая комплексных функций $\varphi_n(t)$.

Приведенные условия показывают попарную ортогональность функций системы. Кроме того предполагается, что для действительных функций помимо условия (1.6)

$$\int_{t_1}^{t_2} \varphi_n^2(t) dt \neq 0 \text{ при } \forall n$$

и для комплексных функций

$$\int_{t_1}^{t_2} \varphi_n(t) \varphi_n^*(t) dt \neq 0 \text{ при } \forall n,$$

то есть функции $\varphi_n(t)$, $n=0, \infty$ не являются тождественными нулю на рассматриваемом временном интервале.

Если функции $\varphi_n(t)$, удовлетворяющие (1.6), выбраны нормированными, то есть такими, что (для действительных функций)

$$\|\varphi_n(t)\|^2 = \int_{t_1}^{t_2} \varphi_n^2(t) dt = 1 \quad (1.7)$$

или (для комплексных функций)

$$||\varphi_n(t)||^2 = \int_{t_1}^{t_2} \varphi_n(t) \varphi_n^*(t) dt = \int_{t_1}^{t_2} |\varphi_n(t)|^2 dt = 1,$$

то система функций (1.5), удовлетворяющих условиям (1.6) и (1.7) называется ортонормированной.

Из математического анализа известно, что произвольная кусочно-непрерывная функция $x(t)$, для которой выполняется условие

$$\int |x(t)|^2 dt < \infty$$

где интеграл берется по всей области определения $t \in T$ сигнала $x(t)$, может быть представлена в виде суммы ряда

$$x(t) = c_0 \varphi_0(t) + c_1 \varphi_1(t) + \dots + c_n \varphi_n(t) + \dots$$

Если левую и правую части данного равенства умножить на $\varphi_n(t)$, затем проинтегрировать в указанных ранее пределах и учесть свойство попарной ортогональности, то для действительных функций легко получится, что

$$\int_{t_1}^{t_2} x(t) \varphi_n(t) dt = c_n ||\varphi_n(t)||^2$$

откуда

$$c_n = \frac{1}{||\varphi_n(t)||^2} \int_{t_1}^{t_2} x(t) \varphi_n(t) dt \quad (1.8)$$

Аналогично для комплексных функций

$$c_n = \frac{1}{||\varphi_n(t)||^2} \int_{t_1}^{t_2} x(t) \varphi_n^*(t) dt. \quad (1.8c)$$

Точность разложения заданной функции $x(t)$ по системе функций $\varphi_n(t)$ в ряд конечной длины N

$$x(t) = c_0 \varphi_0(t) + c_1 \varphi_1(t) + \dots + c_N \varphi_N(t)$$

можно охарактеризовать среднеквадратической ошибкой вида

$$E = \int_{t_1}^{t_2} [x(t) - \sum_{n=0}^N c_n \varphi_n(t)]^2 dt. \quad (1.9)$$

Если *при увеличении* числа членов разложения N среднеквадратическую ошибку E из (1.9) можно сделать сколь угодно малой величиной, то система функций $\varphi_n(t)$, $n=0, \infty$ называется *полной*.

Условие полноты записывается в следующем виде

$$\sum_{n=0}^{\infty} c_n^2 ||\varphi_n(t)||^2 = ||x(t)||^2$$

Выбор наиболее подходящей системы функций $\varphi_n(t)$, $n=0,\infty$ зависит от конкретной задачи, для решения которой производится разложение исходного сигнала в ряд.

На практике наиболее часто разложение сигнала проводится для решения двух задач:

1. максимально точное разложение сигнала на удобные для дальнейшего анализа функции,
2. аппроксимация сигнала минимальным числом членов ряда при удовлетворении заданной степени точности.

При изучении темы декомпозиции сигналов по системе заданных функций часто возникает неправильное отождествление этой процедуры с процедурой разложения заданной функции в ряд.

В последующих примерах мы продемонстрируем эту разницу.

Пример 1.2.

Рассмотрим процедуру разложения заданной функции $x(t)$ в ряд Тейлора (при разложении вблизи точки $t=0$, этот ряд часто также называют рядом Маклорена).

Общий вид ряда Маклорена формально совпадает с формулой декомпозиции

$$x(t) = c_0\varphi_0(t) + c_1\varphi_1(t) + \dots + c_n\varphi_n(t) + \dots,$$

если принять

$$\varphi_0(t) = 1, \varphi_1(t) = t, \dots, \varphi_n(t) = t^n, \dots,$$

то есть представляет собой степенной ряд. Однако коэффициенты данного ряда вычисляются не по приведенным выше формулам декомпозиции, а как значения производных (частных производных) исходной функции в точке разложения.

Разложение в ряд Тейлора (Маклорена) позволяет аппроксимировать исходную функцию только в малой окрестности точки $t=0$, в то время как процедура декомпозиции призвана определить соответствующие коэффициенты разложения таким образом, чтобы минимизировать суммарную (интегральную) ошибку аппроксимации на всем заданном интервале $t_1 < t < t_2$.

Зададим исходную функцию в виде

$$x(t) = \sin(t)$$

и найдем ее разложение в ряд Тейлора в окрестности точки $t=0$. Листинг рабочей сессии представлен на рис.1.2.

```

1. Задаем исходную функцию
x(t) := sin(t)

2. Проводим разложение исходной функции в ряд Маклорена при n=3
x(t) series ,t,3 -> t - t^3/6

3. Проводим разложение исходной функции в ряд Маклорена при n=5
    
```

$$x(t) \text{ series, } t, 5 \rightarrow t - \frac{t^3}{6} + \frac{t^5}{120}$$

4. Строим график исходной функции и двух полученных аппроксимаций в окрестности точки $t=0$

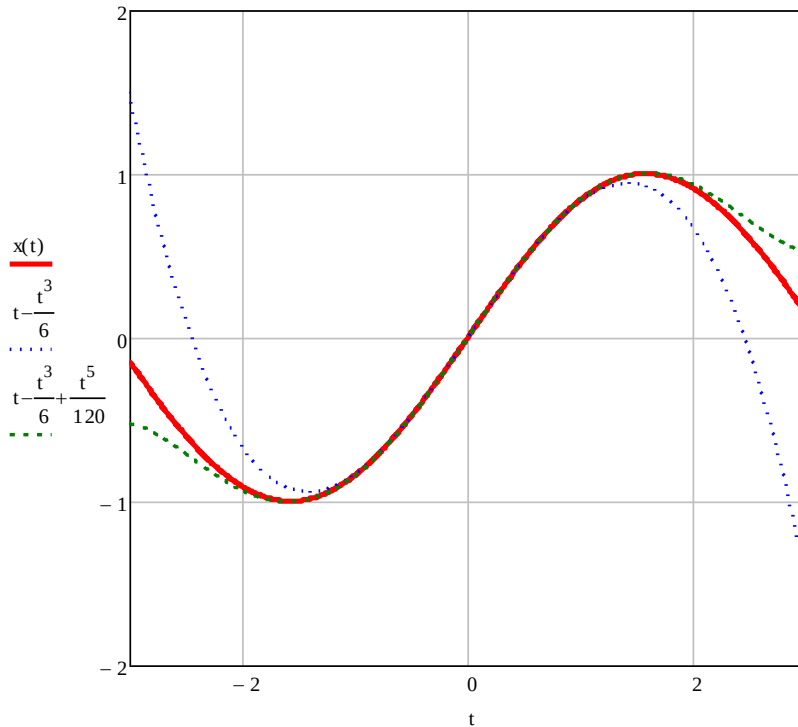


Рис.1.2. Разложение функции $\sin(t)$ в ряд Маклорена

Из приведенных графиков видно, что вблизи точки разложения достигнута высокая точность, но при удалении от этой точки ошибки аппроксимации возрастают.

Пример1.3.

Рассмотрим процедуру аппроксимации функции $x(t)=\sin(t)$ на интервале $-3 < t < 3$ суммой степенных функций $x(t) = c_0\varphi_0(t) + c_1\varphi_1(t) + \dots + c_n\varphi_n(t) + \dots$, где $\varphi_0(t) = 1, \varphi_1(t) = t, \dots, \varphi_n(t) = t^n$, при условии минимизации ошибки $E = \int_{t_1}^{t_2} [x(t) - \sum_{n=0}^N c_n \varphi_n(t)]^2 dt$, где $t_1 = -3, t_2 = 3$. (Для простоты и наглядности решение данной задачи мы проведем не аналитически, а по конечному набору значений исследуемой функции в заданном диапазоне. Соответственно при выборе значений искомых коэффициентов разложения будет минимизироваться не интеграл (1.9) от квадрата ошибки, а сумма квадратов ошибок аппроксимации в каждой из заданных точек).

Для решения задачи воспользуемся программой MATLAB.

m-файл соответствующий рассматриваемому примеру приведен на рис. 1.3.

```
t=(-3:0.1:3); % задаем диапазон изменения аргумента
x=sin(t); % задаем вид исследуемой функции
p3=polyfit(t,x,3); % находим коэффициенты для полинома 3-го порядка
```

```
p5=polyfit(t,x,5); % 5-го
p7=polyfit(t,x,7); % 7-го
f3=polyval(p3,t); % строим аппроксимирующую функцию 3-го порядка
f5=polyval(p5,t); % 5-го
f7=polyval(p7,t); % 7-го
subplot(2,2,1); plot(t,x,'-r',t,f3);grid on; % строим график исходной функции и аппр.3-го порядка
subplot(2,2,2); plot(t,x,'-r',t,f5);grid on;
subplot(2,2,3); plot(t,x,'-r',t,f7);grid on;
subplot(2,2,4); plot(t,f3,'-r',t,f7);grid on;
```

Рис.1.3. Полиномиальная аппроксимация заданной функции

Полученные графики (см. рис. 1.4.) показывают значительно лучший результат аппроксимации исходной функции по сравнению с результатами предыдущего примера.

Однако теперь обратим внимание на один очень важный аспект полученных результатов. Представим себе, что подбор порядка подходящего для целей аппроксимации ряда будет проводиться путем последовательного увеличения числа его членов, пока остаточная ошибка не снизится до приемлемого значения. Обратим внимание на значения получаемых при этом коэффициентов c_n .

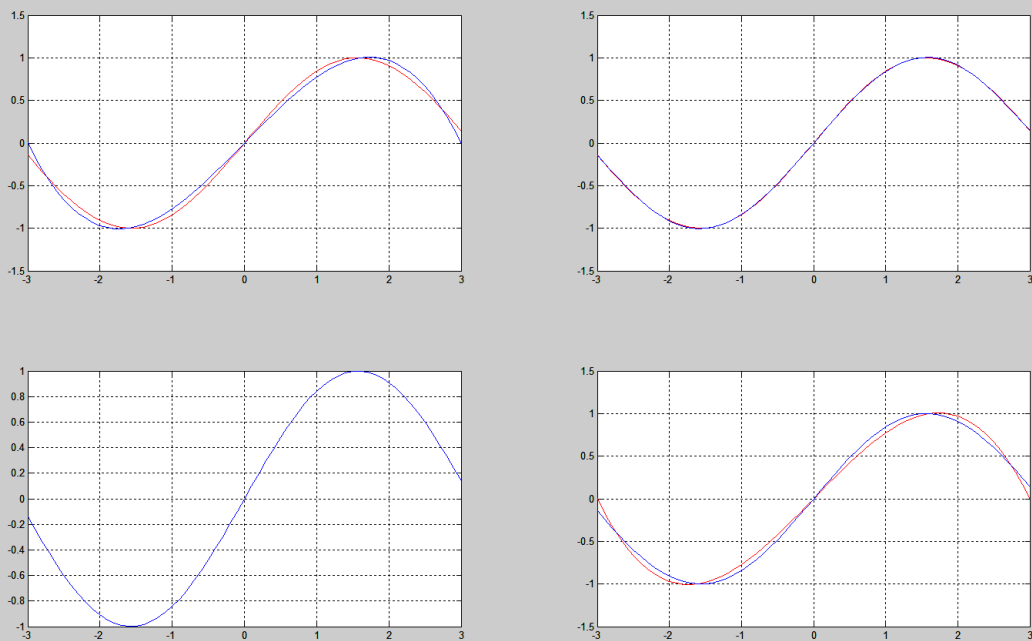


Рис.1.4. Графические результаты моделирования

Для этого распечатаем соответствующие значения коэффициентов для ранее смоделированных полиномов 3-го, 5-го и 7-го порядков (см.рис.1.5.)

```
>> p3
p3 =
-0.0968  0.0000  0.8712 -0.0000
>> p5
p5 =
0.0058  0.0000 -0.1565 -0.0000  0.9898  0.0000
>> p7
p7 =
-0.0002  0.0000  0.0080 -0.0000 -0.1660  0.0000  0.9996 -0.0000
```

Рис.1.5. Коэффициенты полиномиальной аппроксимации различных порядков

Обратим внимание на то, что при увеличении порядка модели изменяются все коэффициенты, то есть, например, коэффициенты, рассчитанные для модели 3-го порядка, не равны первым трем коэффициентам модели 5-го порядка и т.д. Данный факт вызван тем, что система выбранных нами степенных полиномов **не является ортогональной**.

На практике в большинстве случаев было бы крайне удобно иметь такие наборы аппроксимирующих функций, которые при увеличении порядка модели сохраняли бы значения коэффициентов, полученных для моделей меньшего порядка и требовали бы только расчета коэффициентов, появление которых связано с увеличением порядка модели.

К числу подобных полиномов, например, относятся полиномы Лежандра (первого рода), которые определяются в соответствии с формулой

$$P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n} (x^2 - 1),$$

которые ортогональны на интервале $-1 < x < 1$ и при целых $n \geq 0$ содержат конечное число членов.

В соответствии с приведенной формулой первые пять полиномов Лежандра выглядят следующим образом:

$$\begin{aligned} P_0(x) &= 1, \\ P_1(x) &= x, \\ P_2(x) &= \frac{1}{2}(3x^2 - 1), \\ P_3(x) &= \frac{1}{2}(5x^3 - 3x), \\ P_4(x) &= \frac{1}{8}(35x^4 - 30x^2 + 3). \end{aligned}$$

Пример 1.4.

Произвести разложение заданной функции $x(t) = \sin(t)$ на интервале $-1 < t < 1$ с помощью полиномов Лежандра 3-го и 5-го порядка. Сравнить величины получаемых значений коэффициентов разложения.

Для решения поставленной задачи воспользуемся программой MathCad.

Листинг результатов моделирования приведен на рис.1.6.

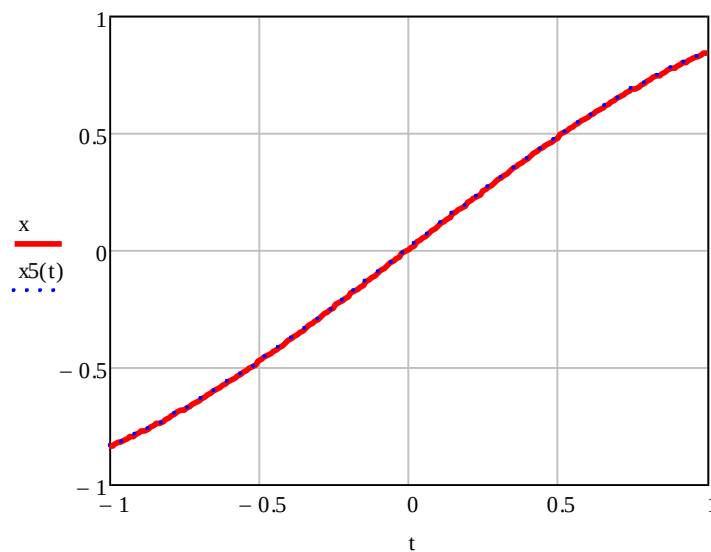
$k := 0..20$	задание количества отсчетов, используемых при моделировании
$t_k := -1 + 0.01 \cdot k$	задание интервала области изменения аргумента
$x_k := \sin(t_k)$	задание исследуемой функции
$F3(t) := \begin{bmatrix} 1 \\ t \\ \frac{1}{2}(3 \cdot t^2 - 1) \end{bmatrix}$	задание общей функции регрессии в виде 3-х полиномов Лежандра
$F5(t) := \begin{bmatrix} 1 \\ t \\ \frac{1}{2}(3 \cdot t^2 - 1) \\ \frac{1}{2}(5 \cdot t^3 - 3 \cdot t) \\ \frac{1}{8}(35 \cdot t^4 - 30 \cdot t^2 + 3) \end{bmatrix}$	задание ... 5-ти полиномов Лежандра
$C5 := \text{linfit}(t, x, F5)$	вычисление коэффициентов разложения
$C3 := \text{linfit}(t, x, F3)$	

$$C3 = \begin{pmatrix} 0 \\ 0.903 \\ 0 \\ 0 \end{pmatrix} \text{ вывод значений коэффициентов разложения (для 3-полиномов)}$$

$$C5 = \begin{pmatrix} 0 \\ 0.904 \\ 0 \\ -0.063 \\ 0 \end{pmatrix} \text{ вывод значений коэффициентов разложения (для 5-ти полиномов)}$$

Формируем аппроксимирующую функцию:

$$x5(t) := 0 \cdot 1 + t \cdot 0.904 + 0 \cdot \left[\frac{1}{2}(3 \cdot t^2 - 1) \right] - 0.063 \cdot \left[\frac{1}{2}(5 \cdot t^3 - 3 \cdot t) \right] + 0 \cdot \left[\frac{1}{8}(35 \cdot t^4 - 30 \cdot t^2 + 3) \right]$$



Графики исходной и аппроксимирующей функций на анализируемом интервале практически совпали.

Рис.1.6. Разложение заданной функции с помощью полиномов Лежандра

При моделировании было выбрано относительно небольшое количество значений исследуемой функции, поэтому в результатах обнаруживается некоторое отклонение от теоретических ожиданий. Тем не менее данный пример наглядно демонстрирует, что применение для целей декомпозиции ортогональной системы функций позволяет при последовательном увеличении порядка аппроксимирующей модели сохранять значения коэффициентов разложения, полученные ранее с помощью модели меньшего порядка.

1.4.Периодические сигналы. Ряд Фурье

Среди многообразия детерминированных сигналов принято выделять две большие группы: **периодические** сигналы и **непериодические** (апериодические) сигналы.

Периодическим называется любой сигнал, для которого выполняется следующее равенство

$$x(t + kT_0) = x(t) \text{ для } \forall t \in T, \quad (1.10)$$

где k – произвольное целое число,

T_0 – **период** сигнала $x(t)$.

Под периодом понимается минимальное действительное число, удовлетворяющее приведенному равенству (1.10).

Величина, обратная периоду, называется **частотой** повторения сигнала и соответственно равна

$$f_0 = \frac{1}{T_0}.$$

В приведенных выражениях индекс «0» использован, чтобы отличить обозначение периода (T_0) от обозначения области определения сигнала (T). В дальнейшем, где из контекста ясно, о чем идет речь, мы будем опускать этот индекс для простоты вида формул.

Очень важное место в теории обработки сигналов занимают **гармонические** сигналы, которые являются частным случаем периодических сигналов, и описываются выражением

$$x(t) = A_m \cos(\omega t + \psi) = A_m \sin(\omega t + \psi') \quad (1.11)$$

где $\psi = \psi' + \pi/2$ – начальная фаза колебаний,

A_m – амплитуда колебаний,

$\theta = \omega t + \psi$ – мгновенная фаза,

$\omega = 2\pi f$ – угловая частота колебаний (рад/с),

f_0 – частота колебаний;

$T_0 = 1/f_0$ – период колебаний.

Начальная фаза колебаний может быть задана в радианах и угловых градусах. Связь между этими величинами определяется простым пропорциональным выражением

$$\psi_{\text{рад}} = \frac{\pi}{180} \psi_{\text{град}}$$

График гармонического сигнала представлен на рис.1.7

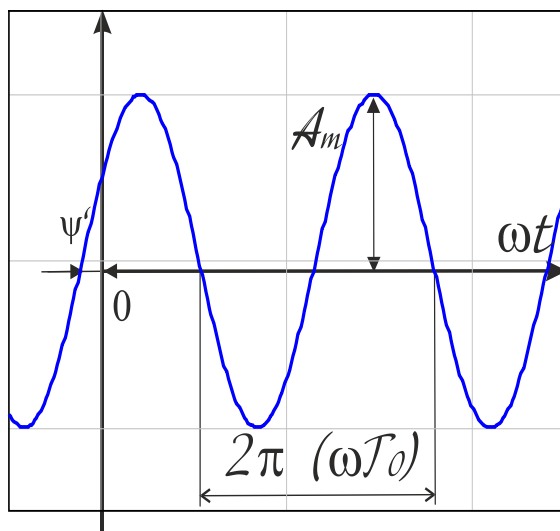


Рис.1.7. График гармонического сигнала

Традиционно в электротехнической литературе используют синусную форму записи, а в радиотехнической – косинусную. Обе формы совершенно эквивалентны.

Среднее значение периодической функции $x(t)$ за период определяется выражением

$$A_{cp} = \frac{1}{T} \int_{t_0}^{t_0+T} x(t) dt$$

Очевидно, что среднее значение гармонической функции за период равно нулю.

Средневыпрямленным значением периодического тока или напряжения называется среднее значение модуля соответствующей периодической функции за период

$$A_{cpB} = \frac{1}{T} \int_{t_0}^{t_0+T} |x(t)| dt$$

После интегрирования для гармонической функции получим

$$A_{cpB} = \frac{2}{\pi} A_m \approx 0,637 A_m$$

Действующим значением периодической функции $x(t)$ называют среднеквадратическое значение этой функции за период

$$A = \sqrt{\frac{1}{T} \int_{t_0}^{t_0+T} (x(t))^2 dt}$$

Действующее значение I переменного периодического тока $i(t)$ численно равно значению постоянного тока I , при протекании которого за время T выделяется такое же количество энергии, как и протекании тока $i(t)$.

Для гармонической функции $x(t)$

$$A = \frac{A_m}{\sqrt{2}} = 0,707 A_m$$

При описании гармонических токов и напряжений принято указывать их действующие, а не амплитудные значения, поэтому помимо варианта записи (1.11) распространена еще одна форма представления гармонической функции

$$x(t) = \sqrt{2} A \cos(\omega t + \psi) = \sqrt{2} A \sin(\omega t + \psi')$$

Гармонический сигнал широко используется в качестве тестового сигнала в радио и электромеханических системах. Этот же сигнал является основным для построения ортонормированного базиса при разложении произвольного периодического сигнала в так называемый ряд Фурье.

В соответствии с теоремой Фурье сложный периодический сигнал, можно представить в виде суммы гармонических колебаний с частотами кратными основной частоте исходного сигнала $\omega_0 = 2\pi/T_0$. Фурье был в большей степени инженером или физиком, нежели математиком, и поэтому первое строгое доказательство теоремы было проведено позже в работах Дирихле. Условия, которым должен соответствовать фрагмент сигнала длительностью в один период, для того чтобы существовало разложение в ряд Фурье получили название условий Дирихле.

Условия Дирихле требуют, чтобы

- число разрывов первого рода было конечным,
- не существовало разрывов второго рода,
- число экстремумов должно быть конечным.

При выполнении условий Дирихле, если исходный сигнал задан в виде (1.11), для

его разложения в ряд Фурье

$$x(t) = c_0\varphi_0(t) + c_1\varphi_1(t) + \dots + c_n\varphi_n(t) + \dots \quad (1.12)$$

следует выбрать ортогональную систему функций вида

$$\begin{aligned} \varphi_0(t) &= 1, \\ \varphi_1(t) &= \sin(\omega_0 t), \quad \varphi_2(t) = \cos(\omega_0 t), \\ \varphi_3(t) &= \sin(2\omega_0 t), \quad \varphi_4(t) = \cos(2\omega_0 t), \\ &\dots\dots\dots \\ \varphi_{2n-1}(t) &= \sin((2n-1)\omega_0 t), \quad \varphi_{2n}(t) = \cos((2n-1)\omega_0 t), \end{aligned}$$

и так далее для $n=0, 1, 2, 3, \dots \infty$.

При таком выборе базиса получается тригонометрическая форма ряда Фурье, используемая для действительных исходных функций сигнала $x(t)$

$$x(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos(n\omega_0 t) + b_n \sin(n\omega_0 t)) \quad (1.13)$$

или

$$x(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} A_n \cos(n\omega_0 t + \theta_n) \quad (1.14)$$

где

$$\begin{aligned} a_n &= \frac{2}{T} \int_{-T/2}^{T/2} x(t) \cos(n\omega_0 t) dt, \\ b_n &= \frac{2}{T} \int_{-T/2}^{T/2} x(t) \sin(n\omega_0 t) dt. \end{aligned}$$

Для комплексного представления разложения используется, так называемая экспоненциальная форма ряда Фурье,

$$x(t) = \dots + c_{-n}\varphi_{-n}(t) + \dots + c_{-1}\varphi_{-1}(t) + c_0\varphi_0(t) + c_1\varphi_1(t) + \dots + c_n\varphi_n(t) + \dots \quad (1.15)$$

для которой базисные функции выбираются как

$$\varphi_0(t) = 1,$$

$$\varphi_1(t) = e^{j\omega_0 t}, \quad \varphi_{-1}(t) = e^{-j\omega_0 t},$$

$$\varphi_2(t) = e^{j2\omega_0 t}, \quad \varphi_{-2}(t) = e^{-j2\omega_0 t},$$

$$\varphi_{-n}(t) = e^{j(2n-1)\omega_0 t}, \quad \varphi_n(t) = e^{j(2n-1)\omega_0 t},$$

где $j=\sqrt{-1}$ - мнимая единица,

$$n = -\infty, \dots, -2, -1, 0, 1, 2, \dots, \infty$$

Тогда исходный сигнал может быть представлен рядом вида

$$x(t) = \sum_{n=-\infty}^{\infty} c_n e^{jn\omega_0 t},$$

где c_n – коэффициенты разложения, определяемые по выражению (1.8.с).

Поскольку понятно, что интервал ортогональности для периодической функции равняется периоду T , то пределы интегрирования в выражениях (1.8) и (1.8.с) можно изменить следующим образом

$$\|\varphi_n(t)\|^2 = \int_{-T/2}^{T/2} e^{jn\omega_0 t} e^{-jn\omega_0 t} dt = T \text{ для } \forall n.$$

Следовательно, коэффициенты разложения могут быть вычислены по формуле

$$c_n = \frac{1}{T} \int_{-T/2}^{T/2} x(t) e^{-jn\omega_0 t} dt. \quad (1.16)$$

Полученное выражение позволяет легко определить связь между тригонометрическим и экспоненциальным представлением ряда Фурье. Действительно, приняв во внимание, что

$$e^{-jn\omega_0 t} = \cos(n\omega_0 t) - j\sin(n\omega_0 t),$$

получим

$$c_n = \frac{1}{T} \int_{-T/2}^{T/2} x(t) \cos(n\omega_0 t) dt - j \frac{1}{T} \int_{-T/2}^{T/2} x(t) \sin(n\omega_0 t) dt = c_{na} - jc_{nb}.$$

Теперь записав коэффициенты разложения как

$$c_n = |c_n| e^{j\theta},$$

$$\text{где } |c_n| = \sqrt{c_{na}^2 + c_{nb}^2},$$

$$\theta_n = -\arctg(c_{nb}/c_{na}),$$

общее выражения для разложения в ряд приводится к виду

$$x(t) = \sum_{n=-\infty}^{\infty} |c_n| e^{j(n\omega_0 t + \theta)} \quad (1.17)$$

Учитывая известное соотношение

$$e^{j\omega_0 t} + e^{-j\omega_0 t} = 2\cos(\omega_0 t),$$

из (1.17) получаем

$$x(t) = c_0 + \sum_{n=1}^{\infty} 2|c_n| \cos(n\omega_0 t + \theta_n) \quad (1.18)$$

Вместо полученного выражения (1.18) на практике часто используется другая форма записи ряда Фурье, получаемая путем раскрытия значения аргумента при косинусе, как было показано в уравнениях (1.13), (1.14).

Связь между коэффициентами разложения в различных формах определяется из очевидных соотношений

$$A_n = 2c_n = \sqrt{a_n^2 + b_n^2}, \quad a_n = 2c_{na}, \quad b_n = 2c_{nb}. \quad (1.19)$$

Если функция $x(t)$ четная относительно времени t , то в тригонометрическом разложении остаются только косинусные члены, если – нечетная, то остаются только синусные члены (см. уравнения (1.13), (1.14), (1.19)).

Зависимость коэффициентов A_n или $|c_n|$ от частоты называют **амплитудным спектром сигнала** $x(t)$. Зависимость θ_n от частоты называется **фазовым спектром**. Для периодических сигналов спектр всегда **линейчатый** (дискретный), то есть состоящий из отдельных линий, соответствующих дискретным значениям частот, кратных частоте повторения сигнала.

Рассмотрим несколько примеров простейших периодических сигналов и их представление рядами Фурье.

На рис. 1.8. изображен сигнал в виде прямоугольного колебания, называемый меандром.

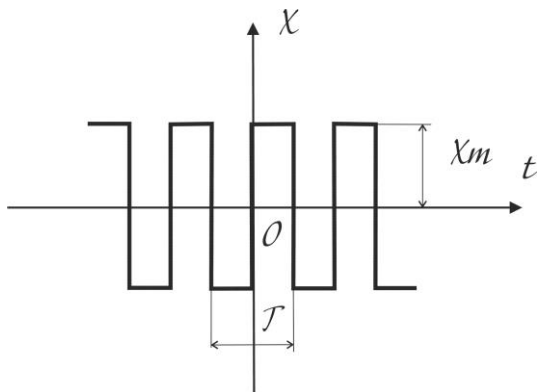


Рис.1.8. График прямоугольного колебания (меандра)

Вычислим коэффициенты ряда Фурье по полученным выше формулам (1.14)

$$a_n = \frac{2}{T} \int_{-T/2}^{T/2} x(t) \cos(n\omega_0 t) dt = 0,$$

$$b_n = \frac{2}{T} \int_{-T/2}^{T/2} x(t) \sin(n\omega_0 t) dt = \frac{2X_m}{T} \left[\int_{-T/2}^0 -\sin(n\omega_0 t) dt + \int_0^{T/2} \sin(n\omega_0 t) dt \right] =$$

$$= \frac{4X_m}{T} \int_0^{T/2} \sin(n\omega_0 t) dt = -\frac{4X_m}{Tn\omega_0} \cos(n\omega_0 t) \Big|_0^{T/2} = \frac{4X_m}{Tn\omega_0} \left(1 - \cos\left(\frac{n\omega_0 T}{2}\right) \right)$$

так как

$$T\omega_0 = 2\pi,$$

то

$$b_n = \begin{cases} 0, & \text{если } n = 0, 2, 4, \dots \\ \frac{4X_m}{n\pi}, & \text{если } n = 1, 3, 5, \dots \end{cases}$$

Разложение в ряд Фурье в тригонометрической форме примет вид

$$x(t) = \frac{4X_m}{\pi} \left(\sin(\omega_0 t) + \frac{1}{3} \sin(3\omega_0 t) + \frac{1}{5} \sin(5\omega_0 t) + \dots \right).$$

Амплитудные спектры меандра представлены на рис.1.9.

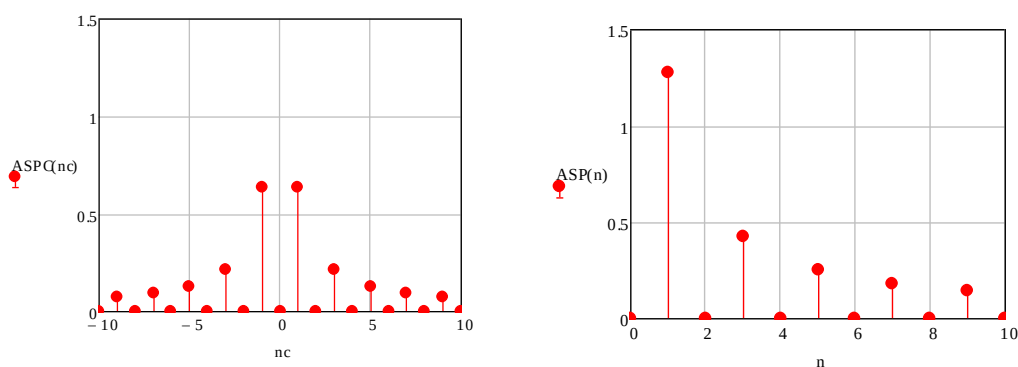


Рис.1.9. Комплексный и вещественный спектры меандра

Теперь посмотрим, как изменится разложение исходного сигнала в ряд Фурье, если произвести временной или амплитудный сдвиг исходного сигнала. На рис 1.10. исходный сигнал смещен половину периода ($T/2$) по оси времени, на рис. 1.11. сигнал смещен вверх на величину амплитуды колебаний X_m (Для простоты расчетов спектра положим $X_m=1$).

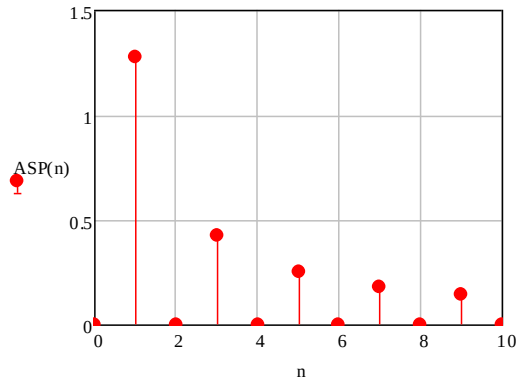
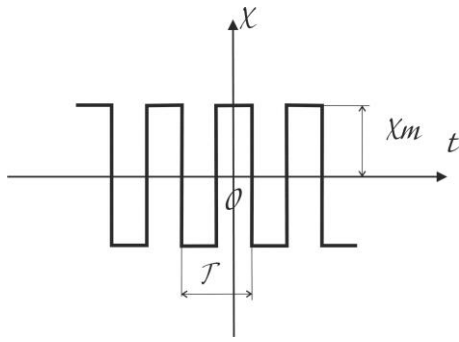


Рис.1.10. Сдвигение сигнала по оси времени

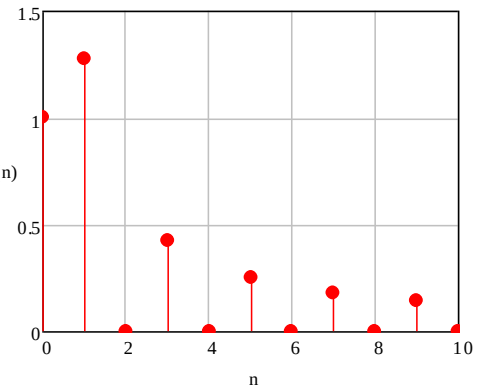
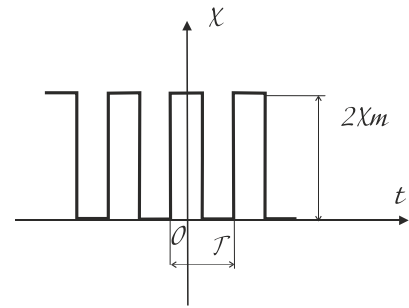


Рис.1.11. Сдвигение сигнала по оси амплитуд

Прделав аналогичные вышеприведенным вычисления для сигнала, изображенного на рис. 1.10 получим

$$x(t) = \frac{4X_m}{\pi} \left(\cos(\omega_0 t) - \frac{1}{3} \cos(3\omega_0 t) + \frac{1}{5} \cos(5\omega_0 t) - \dots \right).$$

Для сигнала, изображенного на рис.1.11. имеем

$$x(t) = X_m + \frac{4X_m}{\pi} \left(\cos(\omega_0 t) - \frac{1}{3} \cos(3\omega_0 t) + \frac{1}{5} \cos(5\omega_0 t) - \dots \right).$$

Сравним полученные выражения для разложения исходного сигнала и его сдвинутых по времени и амплитуде аналогов.

При сдвиге исходного сигнала по оси времени изменений в частотном составе сигнала не произошло (что и следовало ожидать). На рис.1.10. изображен амплитудный спектр исходного сигнала и сдвинутого по времени аналога. Для обоих сигналов он одинаков, что очевидно, так как мы определили амплитудный спектр A_n в виде

$$A_n = \sqrt{a_n^2 + b_n^2}.$$

Таким образом, в спектре присутствуют только линии соответствующие значениям частот $\omega_0, 3\omega_0, 5\omega_0$ и так далее. При этом значения A_n для обоих сигналов совпадают

$$A_n = \frac{4X_m}{n\pi}, \text{ для } n = 1, 3, 5, \dots$$

$$A_n = 0, \text{ для } n = 0, 2, 4, \dots$$

Приведенный пример показывает, что для полного представления исходного сигнала одного амплитудного спектра недостаточно. Требуется учитывать и фазовый спектр, который для рассмотренных сигналов будет различен.

Мы определили фазовый спектр в виде

$$\theta_n = -\arctg\left(\frac{b_n}{a_n}\right).$$

Теперь обратим внимание на изменения, возникающие в спектре сигнала при его амплитудном сдвиге. Из приведенных соотношений видно, что сдвиг по оси ординат привел к появлению дополнительного члена разложения соответствующего нулевой частоте, то есть постоянной составляющей сигнала:

$$\frac{a_0}{2} = X_m,$$

что отразилось на изменении амплитудного спектра. Фазовый спектр остался без изменений.

Рассмотренные примеры показали, что смещение сигнала по оси ординат изменяет амплитудный спектр (составляющую, соответствующую нулевой частоте), а сдвиг сигнала во времени - изменяет фазовый спектр сигнала. Ни в том, ни в другом случае дополнительных частотных составляющих в разложении сигнала не появляется.

Важный с практической точки зрения периодический сигнал представляет собой последовательность унipoлярных импульсов равной длительности τ , повторяющихся с периодичностью T . Этот тип сигнала подобен сигналу, изображенному на рис.1.11. с той разницей, что длительность импульсов τ может принимать любое значение в интервале $0 < \tau < T$, а не только $\tau = 0.5T$, как показано на рис.1.11.

Поскольку временная функция, описывающая данный сигнал является четной, то все коэффициенты ряда Фурье $b_n = 0$. Для коэффициентов a_n в данном случае аналогично рассмотренному выше примеру рис.1.11 нетрудно получить

$$a_n = \frac{2}{T} \int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} x(t) \cos(n\omega_1 t) dt = \frac{4X_m}{\pi n} \sin \frac{n\omega_1 \tau}{2}$$

что приводит к разложению вида

$$x(t) = 2X_m \left(\frac{\tau}{T} + \frac{2}{\pi} \sum_{n=1}^{\infty} \frac{1}{n} \sin \frac{n\omega_1 \tau}{2} \cos(n\omega_1 t) \right).$$

Отношение $\frac{T}{\tau}$ называется **скважностью** импульсной последовательности. При изменении величины скважности импульсов существенное изменение претерпевает и ее

амплитудный спектр, что продемонстрировано на рис.1.12 на примере нахождения разложения сигнала в ряд Фурье в программе Multisim.

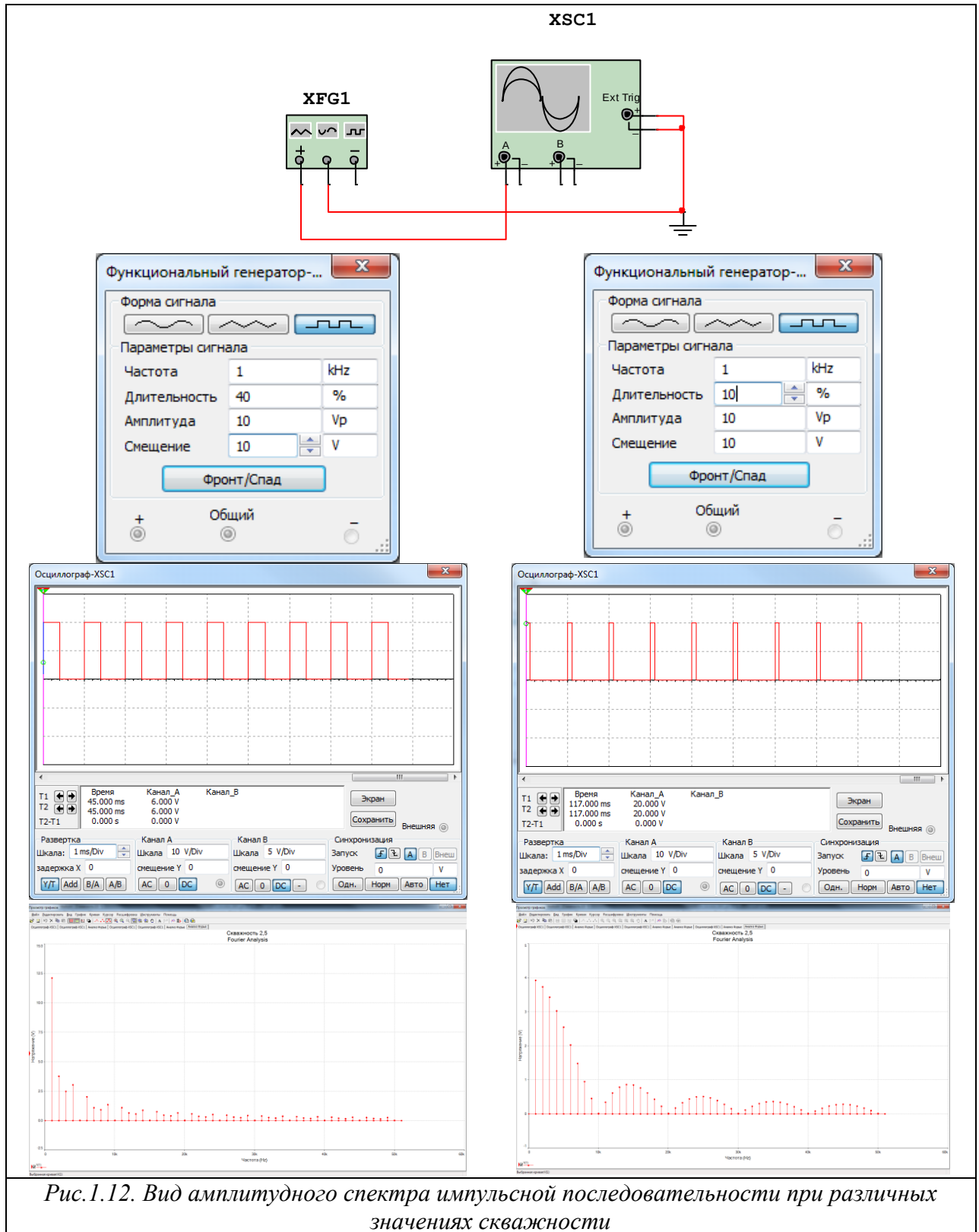


Рис.1.12. Вид амплитудного спектра импульсной последовательности при различных значениях скважности

Пример 1.5.

Для заданного исходного сигнала в виде меандра произвести численный расчет коэффициентов разложения в ряд Фурье.

Задание выполним с помощью программ MathCad (Листинг приведен на рис.1.13.a) и MATLAB ((Листинг приведен на рис.1.13.b).

$T := 1$ Период анализируемой функции

$N := 10$ Максимальное число учитываемых членов ряда Фурье

$x(t) := \begin{cases} 1 & \text{if } 0 < t \leq \frac{T}{2} \\ (-1) & \text{if } -\frac{T}{2} \leq t \leq 0 \end{cases}$ формирование одного периода исследуемого сигнала

график одного периода исследуемого сигнала

$CF(x, T, N) := \begin{pmatrix} CR^{(0)} \leftarrow \left(\frac{2}{T} \cdot \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t) dt \right) \\ \text{for } n \in 1..N \\ CR^{(n)} \leftarrow \begin{pmatrix} \frac{2}{T} \cdot \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t) \cdot \cos\left(n \cdot 2 \frac{\pi}{T} \cdot t\right) dt \\ \frac{2}{T} \cdot \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t) \cdot \sin\left(n \cdot 2 \frac{\pi}{T} \cdot t\right) dt \end{pmatrix} \end{pmatrix}$

CR^T

$CQ(N) := CF(x, T, N)$

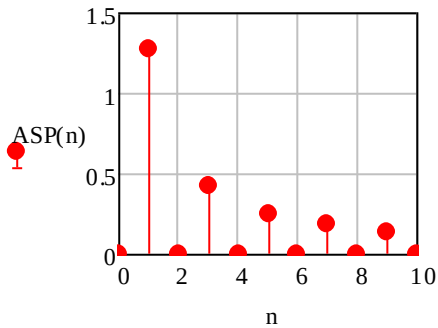
$A := CQ(N)^{(0)}$

$B := CQ(N)^{(1)}$

$n := 0..N$

$$ASP(n) := \begin{cases} \sqrt{(A_n)^2 + (B_n)^2} & \text{if } n > 0 \\ \frac{A_0}{2} & \text{if } n = 0 \end{cases}$$

Амплитудный спектр анализируемого сигнала



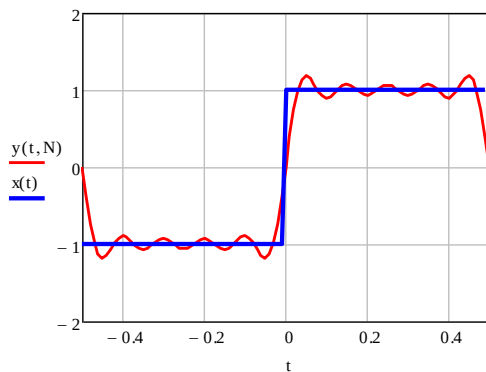
$$y(t, N) := \frac{A_0}{2} + \sum_{n=1}^N \left(A_n \cdot \cos\left(2 \cdot \frac{\pi}{T} \cdot n \cdot t\right) + B_n \cdot \sin\left(2 \cdot \frac{\pi}{T} \cdot n \cdot t\right) \right)$$

Задание аппроксимирующей функции

функции

$$t := \frac{-T}{2}, \frac{-T}{2} + \frac{T}{100} .. \frac{T}{2}$$

Задание интервала для численного моделирования



Графики исходного сигнала и его аппроксимации

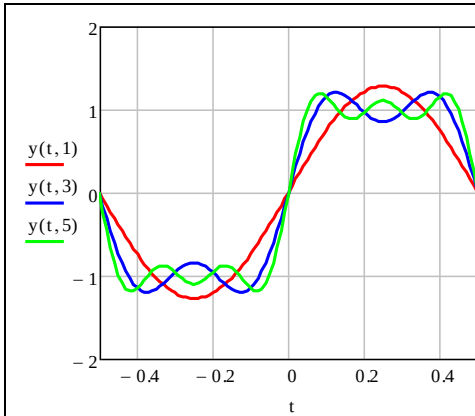
$$A =$$

	0
0	0
1	0
2	0
3	0
4	0
5	0
6	0
7	0
8	0
9	0
10	0

$$B =$$

	0
0	0
1	1.273
2	0
3	0.424
4	0
5	0.255
6	0
7	0.182
8	0
9	0.141
10	0

Коэффициенты ряда Фурье



Вид аппроксимирующих функций для $n=1,3,5$

Рис. 1.13а. Численное определение коэффициентов ряда Фурье (MathCad)

```

T=1;
N=10;
M=3;
MM=1000;
t=-T/2:T/1000:T/2;
fun=@(x)0.*(x>=-T/2 &x<=0)+1.*(x>=0 & x<=T/2);%вид исследуемого сигнала
funsin=@(x,n) fun(x).*sin(2*pi/T*n.*x);
funcos=@(x,n) fun(x).*cos(2*pi/T*n.*x);
subplot(2,2,1); plot(t,fun(t),'LineWidth',3); ylim([-0.5,1.5]);grid on;
a0=2/T*integral(fun,-T/2,T/2);
for k=1:N
a(k)=2/T.*integral(@(x) funcos(x,k),-T/2,T/2);
b(k)=2/T.*integral(@(x) funsin(x,k),-T/2,T/2);
end;
As=[a0 a]
Bs=[0 b]
Aspectr=sqrt(As.^2+Bs.^2);
om=0:N;
SP=Aspectr(om+1);
subplot(1,2,2);
stem(om,SP,'LineWidth',2);
grid on;%xlim([0,9]);
for k=1:MM+1
y(k)=0;
for kf=1:M
y(k)=y(k)+a(kf).*cos(2*pi/T*kf.*(-
T/2+T/MM*k))+b(kf).*sin(2*pi/T*kf.*(-T/2+T/MM*k));
end;
y(k)=y(k)+a0/2;
end;
subplot(2,2,3); plot(t,y,'r','LineWidth',3);grid on

```

As =

Columns 1 through 8

1.0000	-0.0000	-0.0000	0.0000	-0.0000	0.0000	-0.0000
0.0000						

Columns 9 through 11

-0.0000	0.0000	-0.0000
---------	--------	---------

```

Bs =

Columns 1 through 8
      0      0.6366     -0.0000      0.2122      0.0000      0.1273      0.0000
0.0909

Columns 9 through 11
     -0.0000      0.0707      0.0000
    
```

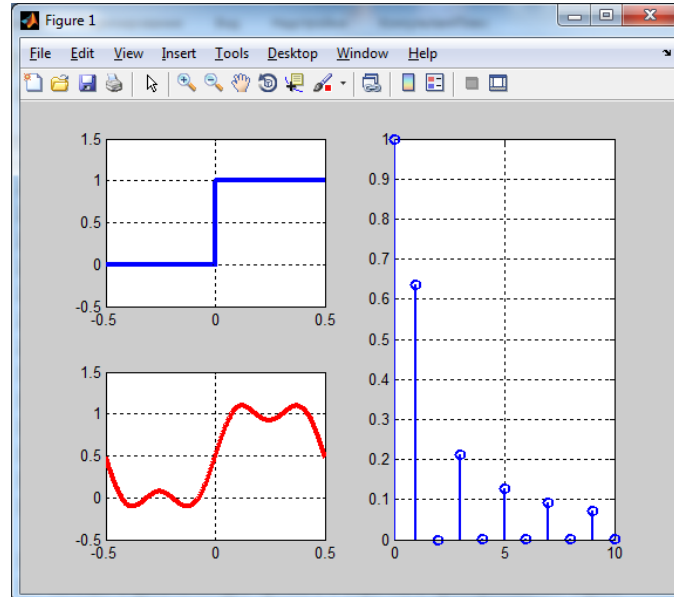
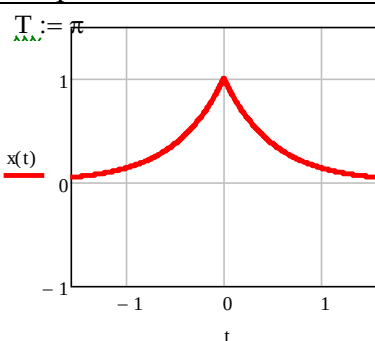


Рис. 1.13 б. Численное определение коэффициентов ряда Фурье (MATLAB)

Пример 1.6.

Произвести определение коэффициентов ряда Фурье и построить аппроксимирующие функции для сигнала, представляющего собой периодическое продолжение функции $x(t) = (\exp(-|t|))^2$, заданной на интервале $-\pi/2 < t < \pi/2$.

Для решения этой задачи частично воспользуемся материалами, полученными в предыдущем примере, заменив толь вид анализируемого сигнала. Листинг расчетов (с сокращениями идентичных предыдущему примеру блоков) приведен на рис.1.14.



Период анализируемой функции

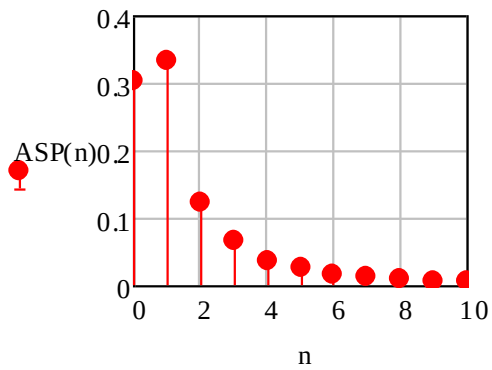
$N_x := 10$

Максимальное число учитываемых членов ряда Фурье

$$x(t) := (e^{-|t|})^2$$

график одного периода исследуемого сигнала

Спектр анализируемого сигнала



Графики исходного сигнала и его аппроксимации

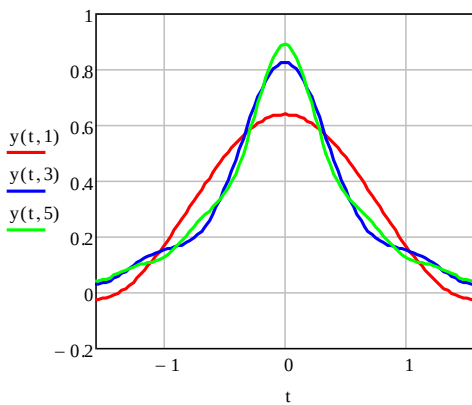
$A =$

	0
0	0.609
1	0.332
2	0.122
3	0.066
4	0.036
5	0.026
6	0.016
7	0.013
8	$9.371 \cdot 10^{-3}$
9	$8.099 \cdot 10^{-3}$
10	$6.031 \cdot 10^{-3}$

$B =$

	0
0	0
1	0
2	0
3	0
4	0
5	0
6	0
7	0
8	0
9	0
10	0

Коэффициенты ряда Фурье



Вид аппроксимирующих функций для $n=1,3,5$

Рис.1.14. Расчет коэффициентов ряда Фурье для заданного сигнала

Пример 1.7.

Произвести определение коэффициентов ряда Фурье и построить аппроксимирующие функции для сигнала, представляющего собой периодическое продолжение линейной функции $x(t)=t$, заданной на интервале $-2 < t < 2$. Листинг (с сокращениями идентичных предыдущим примерам блоков) приведен на рис. 1.15.

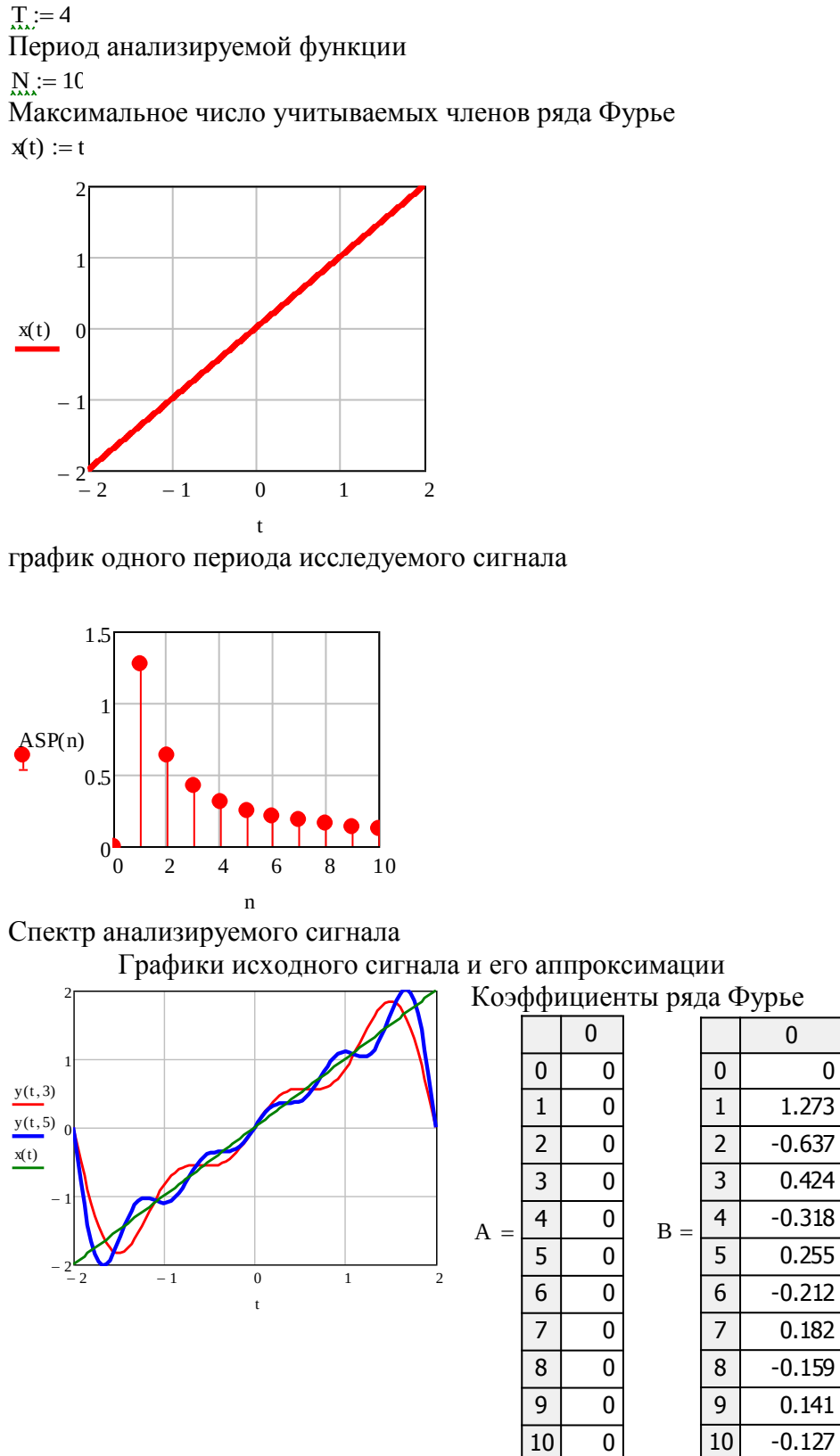


Рис. 1.15. Расчет коэффициентов ряда Фурье для пилообразного сигнала

1.5.Гармонический анализ непериодических сигналов. Интеграл Фурье

Рассмотренный ранее способ разложения периодического сигнала в ряд Фурье предполагает, что исходный сигнал $x(t)$ определен на всей временной оси и неограниченно распространен в прошлом и будущем. Ни один реальный сигнал такими свойствами не обладает. Тем не менее, идеи разложения исходного сигнала на составляющие в базисе из гармонических функций могут быть обобщены и на непериодические сигналы.

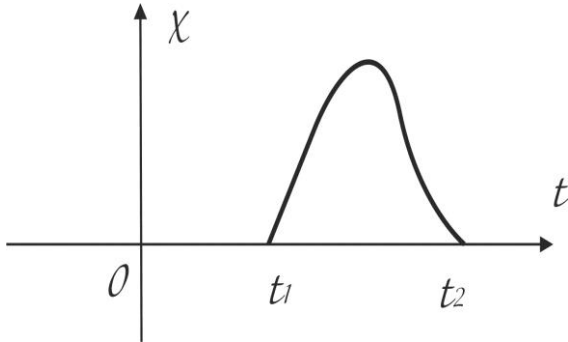


Рис.1.16 График одиночного импульса

Допустим, что исходный сигнал $x(t)$ задан в виде известной функции времени на интервале времени (t_1, t_2) , как показано на рис 1.16.

Предполагаем, что функция $x(t)$ удовлетворяет условиям Дирихле и является абсолютно интегрируемой (то есть интеграл от модуля функции является конечным числом)

$$\int_{-\infty}^{\infty} |x(t)| dt < \infty.$$

Выберем интервал времени $(0, T)$, таким образом, что $0 < t < T$ для $t \in (t_1, t_2)$.

Представим исходный сигнал в виде ряда экспоненциального Фурье

$$x(t) = \sum_{n=-\infty}^{\infty} c_n e^{jn\omega_0 t},$$

где $\omega_0 = \frac{2\pi}{T}$,
 $c_n = \frac{1}{T} \int_{t_1}^{t_2} x(t) e^{-jn\omega_0 t} dt.$

Таким образом,

$$x(t) = \frac{\omega_0}{2\pi} \sum_{n=-\infty}^{\infty} \left[\int_{t_1}^{t_2} x(\tau) e^{-jn\omega_0 \tau} d\tau \right] e^{jn\omega_0 t}, \quad (1.20)$$

Как мы уже отмечали, ряд Фурье описывает периодическую функцию, неограниченно распространяющуюся по оси времени влево и вправо. Поэтому ряд (1.20) периодически продолжит за границы периода $(0, T)$ рассматриваемый исходный сигнал $x(t)$. Для устранения этого эффекта необходимо устремить T к бесконечности, что в свою очередь приведет к

$$\omega_0 = \frac{2\pi}{T} \xrightarrow{T \rightarrow \infty} 0.$$

В предельном переходе от суммирования в выражении (1.20) к интегрированию с заменой ω_0 на $d\omega$ и переходе от линейчатой частоты $n\omega_0$ к текущему значению частоты ω получаем

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{j\omega t} \left[\int_{t_1}^{t_2} x(\tau) e^{-j\omega \tau} d\tau \right] d\omega .$$

Если интервал (t_1, t_2) не уточнен, то пределы интегрирования в целях общности можно раздвинуть до бесконечности.

Так вводится понятие **спектральной плотности** сигнала $x(t)$

$$S(\omega) = \int_{-\infty}^{\infty} x(t) e^{-j\omega t} dt , \quad (1.21)$$

которое позволяет представить сигнал $x(t)$ в виде

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega) e^{j\omega t} d\omega . \quad (1.22)$$

Последняя пара формул (1.21) и (1.22) определяет **прямое** и **обратное преобразование Фурье** соответственно.

Сравнивая выражения для спектральной плотности $S(\omega) = \int_{-\infty}^{\infty} x(t) e^{-j\omega t} dt$, и выражение для комплексного коэффициента ряда Фурье $c_n = \frac{1}{T} \int_{-T/2}^{T/2} x(t) e^{-jn\omega_0 t} dt$, не трудно заметить, что они структурно очень похожи и отличаются только наличием множителя $1/T$. Поэтому неудивительно, что спектральная плотность обладает многими свойствами коэффициентов экспоненциального ряда Фурье.

Поскольку $S(\omega)$ является комплексной функцией частоты ω , то вводя обозначения

$$A(\omega) = \operatorname{Re}(S(\omega)), \quad B(\omega) = -\operatorname{Im}(S(\omega)), \quad \theta(\omega) = -\arctg\left(\frac{B(\omega)}{A(\omega)}\right),$$

можно записать

$$S(\omega) = A(\omega) - j B(\omega)$$

или

$$S(\omega) = \sqrt{A^2(\omega) + B^2(\omega)} e^{j\theta(\omega)} = |S(\omega)| e^{j\theta(\omega)}. \quad (1.23)$$

При этом

$$A(\omega) = \int_{-\infty}^{\infty} x(t) \cos \omega t dt, \quad (1.24)$$

$$B(\omega) = \int_{-\infty}^{\infty} x(t) \sin \omega t dt . \quad (1.25)$$

Данное представление позволяет определить $|S(\omega)|$ как амплитудно-частотную характеристику (или амплитудный спектр), а $\theta(\omega)$ как фазо-частотную характеристику (или фазовый спектр) сплошного спектра сигнала $x(t)$.

Если анализируемый сигнал $x(t)$ является вещественным, то соответствующая спектральная плотность $S(\omega)$ в соответствии с (1.23) - (1.25) обладает свойством сопряженной симметрии

$$S(-\omega) = S^*(\omega).$$

Если $x(t)$ обладает свойством четности, то спектральная плотность будет чисто вещественной. В случае нечетности сигнала $x(t)$, его спектр будет чисто мнимым.

Принимая во внимание, что функция $S(\omega)$ является четной, а функция $\theta(\omega)$ – нечетной, при переходе от экспоненциальной формы записи спектральной плотности в тригонометрической можно записать следующее:

$$\begin{aligned} x(t) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} |S(\omega)| e^{j\theta(\omega)} e^{j\omega t} d\omega = \\ &= \int_{-\infty}^{\infty} \frac{|S(\omega)|}{2\pi} \cos(\omega t + \theta(\omega)) d\omega + j \int_{-\infty}^{\infty} \frac{|S(\omega)|}{2\pi} \sin(\omega t + \theta(\omega)) d\omega = \\ &= \int_0^{\infty} \frac{|S(\omega)|}{\pi} \cos(\omega t + \theta(\omega)) d\omega \end{aligned} \quad (1.26)$$

Полученное выражение (1.26) позволяет наглядно представить "физическую" интерпретацию понятия спектральной плотности. Сравнивая выражение для спектральной плотности и выражение для ряда Фурье легко заметить, что множитель $\frac{|S(\omega)|}{\pi} d\omega$ имеет смысл амплитуды A_n бесконечно малой гармонической составляющей частоты ω .

Выясним связь спектральной плотности одиночного импульса и коэффициентов ряда Фурье для сигнала, полученного периодическим повторением того же импульса с периодом T .

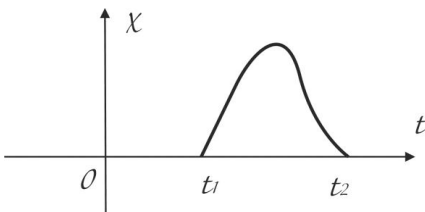


Рис.1.17. Одиночный импульс

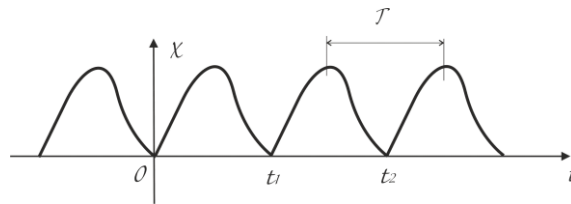


Рис.1.18. Периодическое продолжение импульса.

Допустим, что спектр одиночного импульса задается в виде $S_1(\omega)$. Тогда для коэффициентов ряда Фурье функции, полученной путем T-периодического продолжения, показанного на рис. 1.17 и рис.1.18 можно записать

$$c_n = \frac{S_1(n\omega_1)}{T}$$

или

$$A_n = 2c_n = \frac{2S_1(n\omega_1)}{T}$$

1.6. Основные свойства преобразования Фурье

Преобразование Фурье, описанное в предыдущем параграфе, определяет взаимно однозначную связь между сигналом $x(t)$ и его спектром $S(\omega)$. Отметим некоторые свойства соотношений сигнала со спектром.

1. Свойство сопряженной симметрии

Если анализируемый сигнал $x(t)$ является вещественным (то есть $\text{Im}[x(t)] = 0$), то соответствующая спектральная плотность $S(\omega)$ обладает свойством сопряженной симметрии

$$S(\omega) = S^*(-\omega). \quad (1.27)$$

Из чего следует, что амплитудный спектр $|S(\omega)|$ является четной функцией, а фазовый спектр $\theta(\omega)$ - нечетной.

2. Свойство четной и нечетной симметрии

2.1. Если исходный сигнал $x(t)$ является четной функцией времени, то есть

$$x(t) = x(-t), \quad (1.28 \text{ a})$$

то и спектр будет четной функцией

$$S(\omega) = S(-\omega). \quad (1.28 \text{ b})$$

Если исходный сигнал является и четным и действительным, то спектр также будет четным и действительным.

2.2. Если исходный сигнал $x(t)$ является нечетной функцией, то есть

$$x(t) = -x(-t), \quad (1.29 \text{ a})$$

то и спектр будет нечетной функцией

$$S(\omega) = -S(-\omega). \quad (1.29 \text{ b})$$

Если исходный сигнал одновременно действительный и нечетный, то спектр будет нечетным и чисто мнимым.

3. Свойство дуальности времени и частоты.

Из формул, определяющих прямое и обратное преобразование Фурье, непосредственно следует, что если функция $S(\omega)$ является преобразованием Фурье для функции $x(t)$, то функция $x(\omega)$ в свою очередь является преобразованием Фурье для функции $S(t)$ с учетом множителя $1/\pi$.

4. Линейность

Преобразование Фурье является линейным интегральным преобразованием, то есть для него выполняется принцип суперпозиции: преобразование Фурье от линейной комбинации сигналов равно линейной комбинации (с теми же коэффициентами) преобразований Фурье от каждого из сигналов. Или говоря более простыми словами, преобразование Фурье от суммы сигналов равно сумме преобразований Фурье от каждого сигнала.

Если исходный сигнал имеет вид

$$x(t) = \alpha x_1(t) + \beta x_2(t), \quad (1.30a)$$

то преобразование Фурье этого сигнала будет иметь вид

$$S(\omega) = \alpha S_1(\omega) + \beta S_2(\omega), \quad (1.30b)$$

где α, β – коэффициенты,

$S_1(\omega), S_2(\omega)$ - преобразования Фурье сигналов $x_1(t)$ и $x_2(t)$ соответственно.

5.Сдвиг сигнала во времени.

Если задан сигнал $x(t)$ и ему соответствует спектр $S(\omega)$, то для сигнала, получаемого из исходного путем сдвига во времени на интервал τ

$$x_1(t) = x(t - \tau), \quad (1.31a)$$

легко получить

$$\begin{aligned} S_1(\omega) &= \int_{-\infty}^{\infty} x_1(t) e^{j\omega t} dt = \int_{-\infty}^{\infty} x(t - \tau) e^{j\omega t} dt \\ &= \int_{-\infty}^{\infty} x(t - \tau) e^{j\omega(t - \tau)} d(t - \tau) e^{-j\omega\tau} = S(\omega) e^{-j\omega\tau} \end{aligned} \quad (1.31b)$$

Таким образом, спектр сдвинутого во времени сигнала $S_1(\omega)$ отличается от спектра исходного сигнала $S(\omega)$ только множителем $e^{-j\omega\tau}$. Следовательно, сдвиг сигнала во времени не изменяет амплитудный спектр сигнала, а фазовый спектр приобретает сдвиг $-\omega\tau$, линейно зависящий от частоты.

6. Изменение масштаба времени.

Пусть задан сигнал $x(t)$ и ему соответствует спектр $S(\omega)$. Если исходный сигнал $x(t)$ сжать или растянуть во времени, не меняя его формы, то получившийся сигнал можно записать в виде

$$x_1(t) = x(\alpha t). \quad (1.32a)$$

Понятно, что при положительных значениях α происходит сжатие ($\alpha > 1$) или растяжение ($\alpha < 1$) исходного сигнала.

Теперь определим, какие изменения происходят в спектре сигнала

$$S_1(\omega) = \int_{-\infty}^{\infty} x_1(t) e^{j\omega t} dt = \int_{-\infty}^{\infty} x(\alpha t) e^{j\omega t} dt = \frac{1}{\alpha} \int_{-\infty}^{\infty} x(\alpha t) e^{j\omega \frac{t}{\alpha}} d(\alpha t) = \frac{1}{\alpha} S\left(\frac{\omega}{\alpha}\right). \quad (1.32b)$$

Из полученного выражения видно, что при сжатии (растяжении) сигнала по оси времени во столько же раз растягивается (сжимается) его спектр по оси частот. Модуль спектральной плотности при этом уменьшается (увеличивается) в такое же количество раз.

Если в выражении (1.32a) допустить возможность отрицательных значений коэффициента α , то исходный сигнал кроме растяжения (сжатия) будет зеркально отражаться относительно оси ординат.

Не трудно показать, что для любых значений $\alpha \neq 0$ формула спектральной плотности сжатого (растянутого) сигнала в обобщенном виде принимает вид

$$S_1(\omega) = \frac{1}{|\alpha|} S\left(\frac{\omega}{\alpha}\right).$$

В частном случае $\alpha = -1$, получаем

$$S_1(\omega) = S(-\omega),$$

что соответствует зеркальному отражению спектра относительно нулевой частоты, а для действительного сигнала $x(t)$

$$S_1(\omega) = S(-\omega) = S^*(\omega)$$

комплексному сопряжению спектра.

7. Дифференцирование и интегрирование исходного сигнала

Пусть задан сигнал $x(t)$ и ему соответствует спектр $S(\omega)$. Если дифференцированный и интегрированный сигналы определить в виде

$$x_1(t) = \frac{dx(t)}{dt} \quad (1.33a)$$

$$x_2(t) = \int_{-\infty}^t x(t) dt, \quad (1.34a)$$

то

$$S_1(\omega) = j\omega S(\omega) \quad (1.33b)$$

и

$$S_2(\omega) = \frac{S(\omega)}{j\omega} + \pi S(0)\delta(\omega), \quad (1.34b)$$

где $S(0) = \int_{-\infty}^{\infty} x(t) dt$.

Данные свойства примем без доказательства.

Для сигналов, не содержащих постоянную составляющую, выражение (1.34b) для $S_2(\omega)$ принимает более простой вид

$$S_2(\omega) = \frac{S(\omega)}{j\omega}. \quad (1.34c)$$

8. Спектр свертки двух сигналов

Понятие свертки двух сигналов широко используется в теории обработки сигналов. Подробно об этом речь пойдет в разделе, посвященном прохождению сигнала через линейные системы. Пока же ограничимся простым определением. Сигнал $x(t)$ будем называть сверткой сигналов $x_1(t)$ и $x_2(t)$, если связь между ними задается уравнением следующего вида

$$x(t) = \int_{-\infty}^{\infty} x_1(\tau) x_2(t - \tau) d\tau. \quad (1.35a)$$

Найдем связь между спектром $S(\omega)$, свертки и спектрами исходных сигналов $S_1(\omega)$ и $S_2(\omega)$.

Подставляя в уравнение прямого преобразования Фурье выражение для свертки, получаем

$$S(\omega) = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} x_1(\tau) x_2(t - \tau) d\tau \right] e^{-j\omega t} dt =$$

$$= \int_{-\infty}^{\infty} x_1(\tau) e^{-j\omega\tau} \int_{-\infty}^{\infty} x_2(t-\tau) e^{-j\omega(t-\tau)} d(t-\tau) d\tau = S_1(\omega) S_2(\omega). \quad (1.35b)$$

Таким образом, спектр свертки сигналов равен произведению спектров сигналов в нее входящих.

9.Спектр сигнала, умноженного на гармоническую функцию

Пусть задан сигнал $x(t)$ и ему соответствует спектр $S(\omega)$. Рассмотрим сигнал, получаемый из исходного путем умножения на гармоническую функцию времени (то есть произведение сигналов, один из которых является гармоническим)

$$x_1(t) = x(t) \cos(\omega_0 t + \theta_0) \quad (1.36a)$$

Воспользовавшись формулой Эйлера, заменим косинус суммой экспонент и запишем уравнение прямого преобразования Фурье

$$\begin{aligned} S_1(\omega) &= \frac{1}{2} \int_{-\infty}^{\infty} x(t) [e^{j(\omega_0 t + \theta_0)} + e^{-j(\omega_0 t + \theta_0)}] e^{-j\omega t} dt = \\ &= \frac{1}{2} \int_{-\infty}^{\infty} x(t) e^{j\theta_0} e^{-j(\omega - \omega_0)t} dt + \frac{1}{2} \int_{-\infty}^{\infty} x(t) e^{-j\theta_0} e^{-j(\omega + \omega_0)t} dt = \\ &= \frac{1}{2} e^{j\theta_0} S(\omega - \omega_0) + \frac{1}{2} e^{-j\theta_0} S(\omega + \omega_0). \end{aligned} \quad (1.36b)$$

Как видно из полученного уравнения (1.36b) спектр исходного сигнала «расщепился» на две составляющие половинного значения модуля и сдвинутые по оси частот на величину круговой частоты гармонического сигнала ω_0 влево и вправо от нулевого значения. Кроме того в каждом из слагаемых появился множитель, зависящий от начальной фазы гармонического сигнала.

10.Спектр произведения сигналов. Пусть сигнал $x(t)$ является произведением сигналов $x_1(t)$ и $x_2(t)$, имеющих соответствующие спектры $S_1(\omega)$ и $S_2(\omega)$

$$x(t) = x_1(t) x_2(t). \quad (1.37a)$$

Тогда воспользовавшись свойством дуальности времени и частоты (см. *свойство 3*) легко получить

$$S(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_1(\vartheta) S_2(\omega - \vartheta) d\vartheta, \quad (1.37b)$$

то есть спектр произведения сигналов равен (с поправкой на множитель $1/2\pi$) свертке спектров исходных сигналов.

11.Связь преобразования Фурье и коэффициентов ряда Фурье

Допустим, что исходный сигнал $x(t)$ имеет конечную длительность и ему соответствует спектр $S(\omega)$. Рассмотрим сигнал $x_1(t)$, получаемый из исходного путем периодического повторения с периодом T_0 , превышающим длительность исходного сигнала, то есть

$$x_1(t) = \sum_{k=-\infty}^{\infty} x(t - kT_0). \quad (1.38a)$$

Графики функций $x(t)$ и $x_1(t)$ приведены на рис.1.16, 1.17. Поскольку функция $x_1(t)$ является периодической и для нее выполняются условия Дирихле (иначе бы не был определен спектр $S(\omega)$), то она может быть разложена в ряд Фурье с комплексными коэффициентами c_k .

Сравнивая уравнения для преобразования Фурье и коэффициентов ряда Фурье, легко заметить, что их получение основано на взятии интегралов идентичного вида. Разница состоит только в том, что при расчете коэффициентов ряда Фурье в подынтегральном выражении используются только дискретные значения частоты $\omega_k = k\omega_0$, $k = 0, \pm 1, \pm 2$, и т. д, кратные основной частоте периодического сигнала $\omega_0 = 2\pi T_0$. Кроме того результат интегрирования делится на период сигнала T_0 .

Таким образом, связь между спектральной функцией $S(\omega)$, исходного сигнала $x(t)$ и коэффициентами c_k ряда Фурье его периодически продолженного варианта $x_1(t)$ описывается следующим уравнением

$$c_k = \frac{1}{T_0} S(2\pi k/T_0). \quad (1.38b)$$

12. Распределение энергии в спектре непериодического колебания

Если сигналы $x_1(t)$ и $x_2(t)$ представляют собой один и тот же сигнал

$$x(t) = x_1(t) = x_2(t),$$

то интеграл от их произведения представляет собой полную энергию данного сигнала

$$\int_{-\infty}^{\infty} x_1(t)x_2(t)dt = \int_{-\infty}^{\infty} x^2(t)dt = \mathcal{E}.$$

С другой стороны произведение соответствующих спектральных плотностей согласно ранее приведенным свойствам приводится к следующему значению

$$X_1(\omega)X_2^*(\omega) = S(\omega)S^*(\omega) = |S(\omega)|^2 = S^2(\omega),$$

где $S(\omega)$ – спектр сигнала $x(t)$.

Тогда окончательно можно записать следующее равенство, называемое равенством Парсеваля

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} |S(\omega)|^2 d\omega = \frac{1}{\pi} \int_0^{\infty} S^2(\omega) d\omega = \int_{-\infty}^{\infty} x^2(t)dt = \mathcal{E} \quad (1.39)$$

Полученное соотношение показывает важную связь между представлениями энергии сигнала во временной и частотной областях. Отсюда также видно, что величина $|S(\omega)|^2$ может рассматриваться как спектральная плотность энергии, то есть имеет смысл энергии приходящейся на бесконечно малую полосу частот.

Выражения (1.27) – (1.39), приведенные в данном разделе, будут в дальнейшем использоваться при рассмотрении конкретных примеров сигналов и цепей.

1.7.Примеры расчета преобразования Фурье

1.Прямоугольный импульс

Начнем рассмотрение с прямоугольного импульса, который является основой для построения периодической функции меандра, рассмотренной нами ранее. Математически прямоугольный импульс определяется в виде

$$x(t) = \begin{cases} X_m, & \text{если } t \in [-\frac{T}{2}, \frac{T}{2}] \\ 0, & \text{если } |t| > \frac{T}{2}, \end{cases}$$

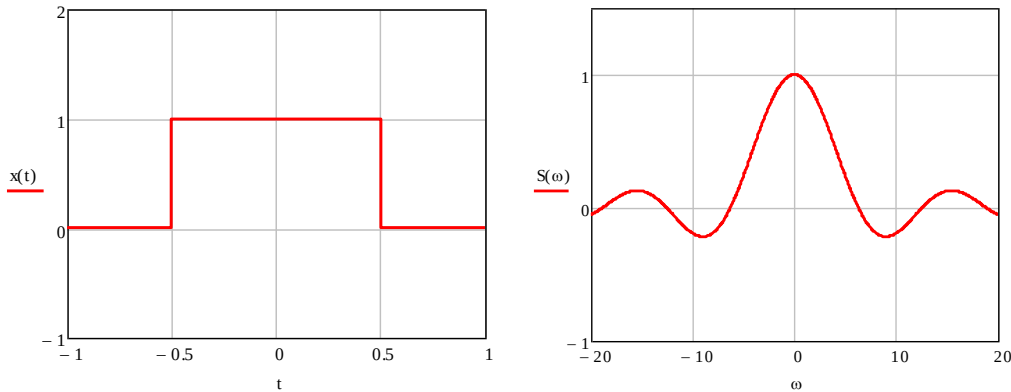


Рис.1.19. Спектральная функция прямоугольного импульса

Вычисляем спектральную функцию сигнала по формуле

$$S(\omega) = \int_{-T/2}^{T/2} X_m e^{-j\omega t} dt = \frac{X_m}{-j\omega} e^{-j\omega t} \Big|_{-T/2}^{T/2} = X_m T \frac{\sin(\omega T/2)}{\omega T/2}.$$

Функцию вида $(\sin(x))/x$ часто обозначают как $\text{sinc}(x)$. С учетом такого обозначения выражение для спектральной функции примет вид

$$S(\omega) = X_m T \text{sinc}(\omega T/2).$$

Заметим, что произведение $X_m T$, входящее в состав полученного уравнения, равно площади под импульсом и определяет значение спектральной функции при нулевой частоте

$$S(0) = X_m T.$$

Данный вывод справедлив не только для прямоугольного импульса, но и для импульса произвольной формы

$$S(0) = \int_{-T/2}^{T/2} x(t) e^{-j0t} dt = \int_{-T/2}^{T/2} x(t) dt.$$

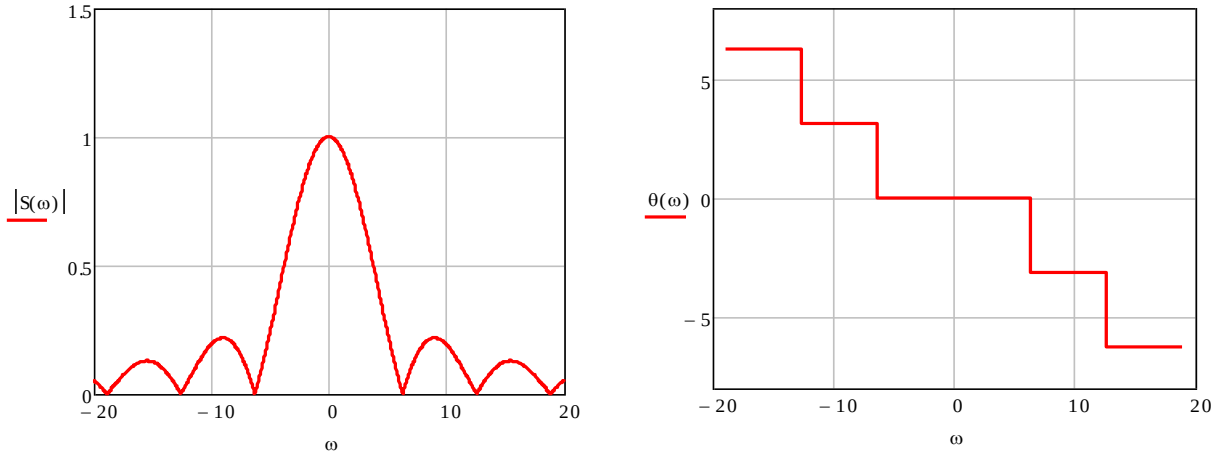
В результате проведенных преобразований для прямоугольного импульса получена спектральная плотность, мнимая часть которой тождественно равна нулю. Поэтому при определении фазового спектра в данном случае поступают следующим образом.

Запишем выражение для полученной спектральной плотности в экспоненциальном виде

$$S(\omega) = |S(\omega)|e^{j\theta(\omega)}$$

очевидно, что $|S(\omega)| = X_m T \text{sinc}(\omega T/2)$ является всюду неотрицательной величиной, поэтому фазовый спектр $\theta(\omega)$ в данном случае определяют таким образом, чтобы обеспечить необходимое чередование положительных и отрицательных значений функции $S(\omega)$.

Амплитудный и фазовый спектр для прямоугольного импульса изображены на рис. 1.20.



1.20. Амплитудный и фазовый спектры прямоугольного импульса

Полученная спектральная функция является вещественной, поэтому фазовый спектр принимает только два значения 0 и π . Поскольку значения фазы π и $-\pi$ неразличимы, то знаки для фазового спектра использованы лишь для удовлетворения свойства нечетности получаемой функции.

Посмотрим теперь, что будет происходить со спектральной функцией, если исходный прямоугольный импульс сдвинуть по вертикальной или горизонтальной оси.

В случае сдвига импульса по вертикальной оси ($x_1(t) = \alpha x(t)$, где α -скаляр) произойдет лишь добавление α в качестве дополнительного масштабирующего множителя

$$S_1(\omega) = \alpha S(\omega).$$

В случае сдвига исходного импульса по оси времени

$$x_1(t) = \begin{cases} X_m, & \text{если } t \in [0, T] \\ 0, & \text{если } t \notin [0, T] \end{cases}$$

при вычислении преобразования Фурье

$$S_1(\omega) = \int_0^T X_m e^{-j\omega t} dt = \frac{X_m}{-j\omega} e^{-j\omega t} \Big|_0^T = X_m T \frac{\sin(\omega T/2)}{\omega T/2} e^{-j\omega T/2} = S(\omega) e^{-j\omega T/2}$$

появляется дополнительный множитель $e^{-j\omega T/2}$, который как и в случае временного сдвига периодического сигнала, приведет к изменению фазового спектра. Амплитудный спектр в данном случае останется без изменений.

2. Импульс вида $\text{sinc}(t)$

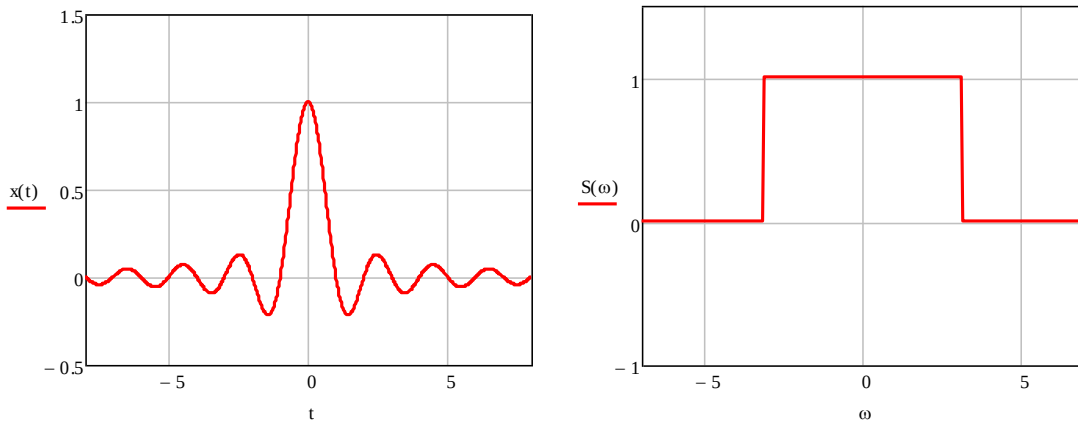
Полученные выражения позволяют легко определить спектральную функцию для сигнала

$$x(t) = \text{sinc}(\omega_0 t) = \frac{\sin(\omega_0 t)}{\omega_0 t}.$$

Воспользовавшись принципом дуальности времени и частоты в преобразовании Фурье получаем

$$S(\omega) = \begin{cases} \frac{\pi}{\omega_0}, & \text{если } |\omega| \leq \omega_0 \\ 0, & \text{если } |\omega| > \omega_0. \end{cases}$$

График данной спектральной функции приведен на рис.1.21.



1.21.Импульс $\text{sinc}(t)$ и его спектральная функция

3. Гауссов импульс

Импульс, задаваемый выражением

$$x(t) = X_m e^{-\frac{t^2}{2\alpha}},$$

где α - скаляр,

$$-\infty < t < \infty,$$

называют гауссовым (или колоколообразным) импульсом. Этот импульс изображен на рис.1.22.

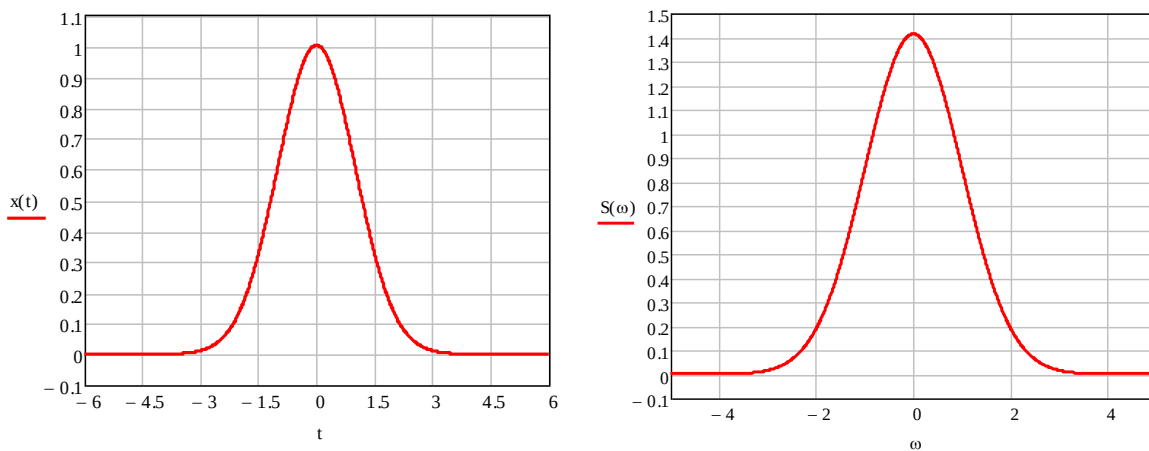


Рис.1.22. Гауссов импульс и его спектральная функция

Применяя к этому выражению преобразование Фурье, получаем

$$S(\omega) = \int_{-\infty}^{\infty} e^{-\frac{t^2}{2\alpha}} e^{-j\omega t} dt,$$

что после несложных преобразований дает

$$S(\omega) = X_m \alpha \sqrt{2\pi} e^{-\alpha^2 \omega^2 / 2}.$$

График этой функции изображен на рис. 3. Сравнение уравнений () и () показывает, что гауссов импульс и его спектр описываются одинаковыми функциями и обладают свойствами симметрии.

4. Прямоугольный радиоимпульс (радиосигнал)

Данный сигнал отличается от рассмотренного выше видеоимпульса тем, что интервал его длительности заполнен высокочастотным гармоническим колебанием, как показано на рис. 1.22.

Для такого сигнала спектральная функция может быть найдена следующим образом

$$\begin{aligned} S(\omega) &= \int_{-T/2}^{T/2} X_m \cos \omega_0 t e^{-j\omega t} dt = \int_{-T/2}^{T/2} X_m \frac{e^{j\omega_0 t} + e^{-j\omega_0 t}}{2} e^{-j\omega t} dt = \\ &= \frac{X_m}{2} \left\{ \int_{-T/2}^{T/2} e^{-j(\omega - \omega_0)t} dt + \int_{-T/2}^{T/2} e^{-j(\omega + \omega_0)t} dt \right\} = \\ &= \frac{X_m T}{2} \left\{ \frac{\sin\left(\frac{(\omega - \omega_0)T}{2}\right)}{\frac{(\omega - \omega_0)T}{2}} + \frac{\sin\left(\frac{(\omega + \omega_0)T}{2}\right)}{\frac{(\omega + \omega_0)T}{2}} \right\} \end{aligned}$$

График полученной спектральной характеристики приведен на рис.1.23

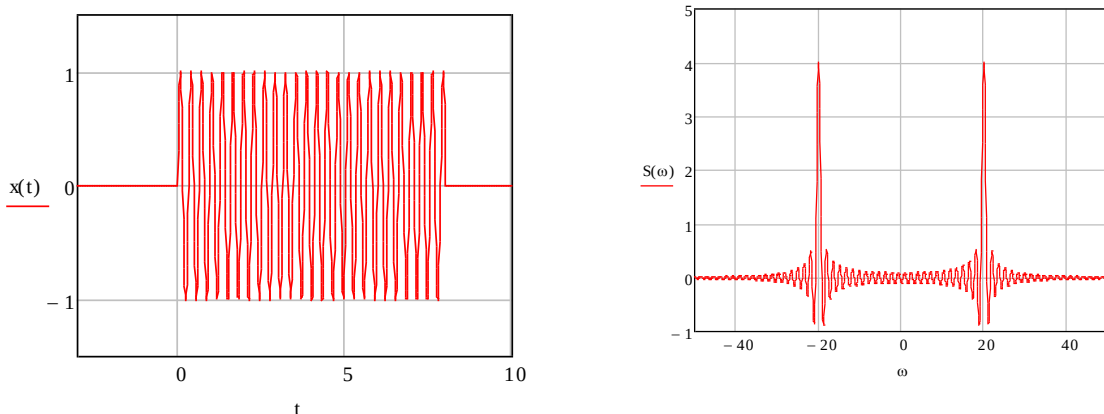


Рис.1.23. Прямоугольный радиоимпульс и его спектральная характеристика

Сравнивая полученное уравнение с выражением для прямоугольного видеоимпульса, легко заметить связь между ними

$$S(\omega) = \frac{1}{2}(S_v(\omega + \omega_0) + S_v(\omega - \omega_0)),$$

где $S_v(\omega)$ – спектральная характеристика прямоугольного видеоимпульса.

Поскольку видеоимпульс является огибающей функцией для прямоугольного радиоимпульса, то его спектр по отношению к радиоимпульсу называют спектром огибающей.

Из графика рис.1.23 хорошо видно, что хотя спектральные составляющие сконцентрированы в окрестностях точек $\omega_0, -\omega_0$, они, тем не менее, оказывают определенное влияние друг на друга. Степень их взаимного влияния зависит от соотношения частоты заполнения импульса ω_0 и его длительности T .

1.8.Преобразование Фурье для некоторых сингулярных функций

Как было указано ранее, для применения к сигналу преобразования Фурье должны выполняться условия Дирихле и абсолютная интегрируемость. Однако существуют сигналы, не удовлетворяющие этим требованиям, к которым можно применить преобразование Фурье и получить практически полезный результат.

1.Дельта-функция

Формальное определение дельта-функции можно дать в следующем виде

$$\delta(t) = \begin{cases} \infty, & \text{при } t = 0 \\ 0, & \text{при } t \neq 0 \end{cases} \quad (1.40)$$

и одновременно

$$\int_{-\infty}^{\infty} \delta(t) dt = 1$$

Таким образом, дельта-функция представляет собой идеализацию бесконечно короткого импульса с бесконечно большой амплитудой, площадь которого равна единице.

Дельта-функция (2.40) часто также именуется единичным импульсом, импульсной функцией или функцией Дирака.

На рис.1.24. изображены реальные симметричные импульсы, которые при $\tau \rightarrow 0$, могут рассматриваться как один из вариантов предельного перехода к единичному импульсу, изображение которого приведено на рис. 1.254.

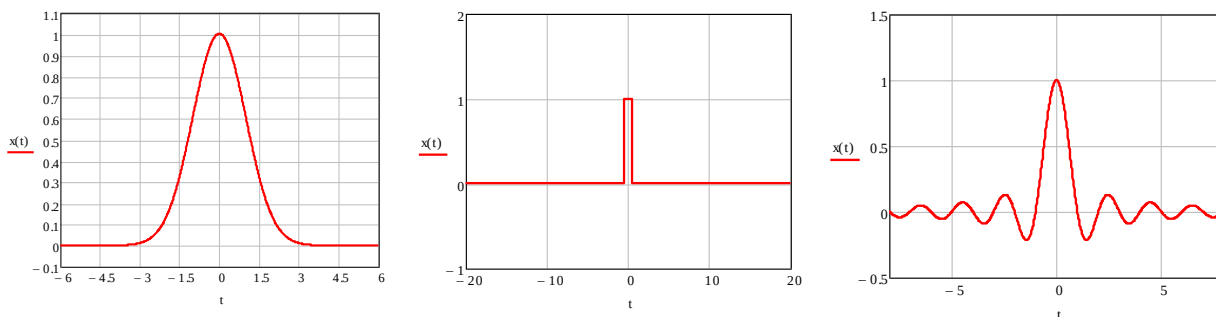


Рис.1.24. Реальные импульсы

Дельта функция обладает важным свойством, благодаря которому она получила широкое распространение в математике, физике и прикладных науках, а именно:

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

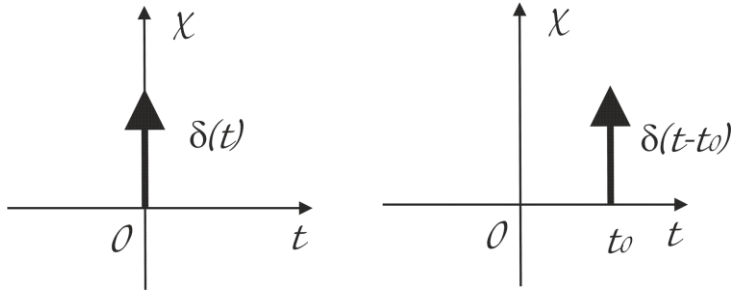


Рис.1.25. Идеальный импульс

Данное свойство получило наименование фильтрующего свойства, которое состоит в возможности выделять значение функции $x(t)$ в произвольный момент времени t_0 .

Действительно, заменив, пределы интегрирования в предыдущем уравнении получим

$$\int_{t_1}^{t_2} \delta(t - t_0)x(t)dt = x(t_0) \int_{t_1}^{t_2} \delta(t - t_0)dt = \begin{cases} x(t_0), & \text{если } t_0 \in]t_1, t_2[\\ 0, & \text{если } t_0 \notin]t_1, t_2[\end{cases}.$$

Все приведенные выше уравнения для дельта-функции использовали в качестве аргумента время t , при этом очевидно, что свойства дельта-функции не изменятся при замене аргумента, например, на частоту ω .

Теперь, применяя к дельта-функции формулу преобразования Фурье, запишем

$$S(\omega) = \int_{-\infty}^{\infty} \delta(t)e^{-j\omega t}dt = 1. \quad (1.41)$$

Спектральная функция единичного импульса является константой, то есть спектр этого сигнала равномерно распределен в бесконечной полосе частот. Полученный результат (1.41) полностью согласуется с общим соотношением между длительностью сигнала и его спектром: чем короче сигнал, тем шире его спектр. Поэтому при бесконечно короткой длительности сигнала следует ожидать бесконечную протяженность его спектра.

Взяв обратное преобразование Фурье от постоянной и равной единице спектральной функции, для единичного импульса получаем

$$\delta(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-j\omega t} d\omega$$

Понятие единичного импульса особенно широко используется при исследовании прохождения коротких сигналов через линейные системы, рассмотрение которых будет проведено ниже.

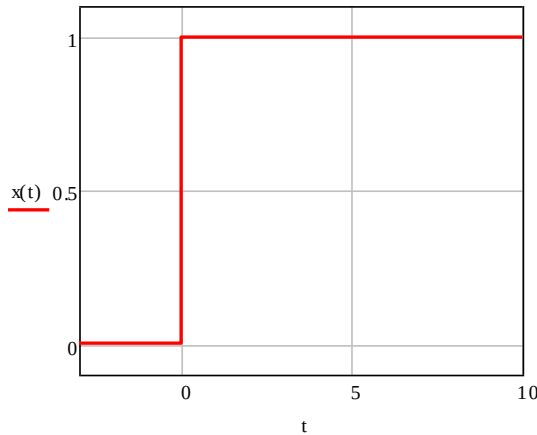
2. Единичный скачок

Другим важным сингулярным сигналом является, так называемый единичный скачок, математическое описание которого дается следующим образом

$$\mathbf{1}(t) = \begin{cases} 0, & \text{если } t < 0 \\ \frac{1}{2}, & \text{если } t = 0 \\ 1, & \text{если } t > 0. \end{cases} \quad (1.42)$$

Значение функции в нулевой момент времени иногда считают неопределенным или приравнивают единице.

Функцию единичного скачка также называют единичной функцией или функцией Хевисайда. График единичной функции приведен на рис.1.26.



Единичный скачок удобно использовать при создании математического описания сигналов конечной длительности, так например, прямоугольный импульс величиной X_m , длящийся от момента времени t_1 до момента времени t_2 , легко представить как разность двух единичных скачков

$$x(t) = X_m[\mathbf{1}(t - t_1) - \mathbf{1}(t - t_2)].$$

Связь между функциями единичного импульса и единичного скачка

Рис.1.26. График функции Хевисайда

задается в следующем виде

$$\delta(t - t_0) = \frac{d}{dt} \mathbf{1}(t - t_0) \quad (1.43)$$

и

$$\mathbf{1}(t - t_0) = \int_{-\infty}^t \delta(t - t_0) dt. \quad (1.44)$$

Функция единичного скачка (1.42) представляет собой интеграл от функции единичного импульса (1.41), поэтому для определения его спектра можно воспользоваться свойством преобразования Фурье для спектра от интеграла исходного сигнала (с учетом (1.43),(1.44))

$$S(\omega) = \int_{-\infty}^{\infty} \mathbf{1}(t) e^{-j\omega t} dt = \pi \delta(\omega) + \frac{1}{j\omega}. \quad (1.45)$$

3.Постоянный сигнал

Выясним, каков спектр сигнала, имеющего постоянное значение на всем протяжении времени, то есть

$$x(t) = X_m = const \quad \text{при } \forall t.$$

Применим свойство дуальности преобразования Фурье к результатам, полученным нами при вычислении спектра дельта-функции. Поскольку мы ранее получили, что спектром единичного импульса является константа, то спектром константы должна быть дельта-функция с поправкой на коэффициент 2π , то есть

$$S(\omega) = \int_{-\infty}^{\infty} X_m e^{-j\omega t} dt = 2\pi X_m \delta(t). \quad (1.46)$$

4. Гармонический сигнал

Для гармонического сигнала, заданного выражением

$$x(t) = X_m \cos(\omega_0 t + \theta_0)$$

Спектральную функцию можно рассчитать либо непосредственно по уравнениям преобразования Фурье, либо рассмотрев ее как произведение постоянного сигнала на гармоническую функцию, и воспользоваться ранее полученными выражениями.

Естественно, в обоих случаях результат будет одинаков

$$S(\omega) = X_m \pi e^{j\theta_0} \delta(\omega - \omega_0) + X_m \pi e^{-j\theta_0} \delta(\omega + \omega_0). \quad (1.47)$$

График данной функции приведен на рис.1.27.

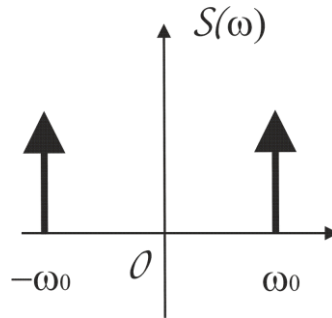


Рис.1.27. Спектр гармонического сигнала

5. Комплексная экспонента

Если исходный сигнал описывается функцией вида

$$x(t) = X_m e^{j\omega t},$$

то такой сигнал называется комплексной экспонентой. Используя формулу Эйлера его легко записать в ином виде

$$x(t) = X_m \cos \omega t + j X_m \sin \omega t.$$

Тогда с учетом ранее полученных уравнений легко записать

$$S(\omega) = \int_{-\infty}^{\infty} X_m e^{j\omega_0 t} e^{-j\omega t} dt = 2X_m \pi \delta(\omega - \omega_0). \quad (1.48)$$

Поскольку исходный сигнал не является вещественным, то полученная его спектральная функция не обладает свойством симметрии.

6. Произвольный периодический процесс

Ранее мы рассматривали связь преобразования Фурье и коэффициентов ряда Фурье для периодически продолженного сигнала. Теперь рассмотрим близкую по сути задачу об определении спектральной функции произвольного периодического сигнала.

Если рассматриваемый сигнал $x(t)$ является периодическим с периодом $T_0 = 2\pi/\omega_0$, то при выполнении условий Дирихле он может быть представлен в виде ряда

$$x(t) = \sum_{n=-\infty}^{\infty} c_n e^{jn\omega_0 t}.$$

С учетом свойств спектра комплексной экспоненты для суммы экспонент получаем

$$S(\omega) = \sum_{k=-\infty}^{\infty} 2\pi c_k \delta(\omega - \omega_0 k). \quad (1.49)$$

Следовательно, спектральная функция периодического сигнала представляет собой набор дельта-функций, расположенных на частотах, кратных основанной частоте сигнала, а множители при этих дельта-функциях равны коэффициентам ряда Фурье, умноженным на 2π .

Уравнения (1.41) – (1.49) являются чрезвычайно важными и полезными при анализе свойств более сложных сигналов, что будет рассмотрено далее в материалах лекций и практических занятий.

1.9. Корреляционный анализ детерминированных сигналов

Рассмотренная ранее спектральная плотность сигналов позволяет оценить их характеристики в частотной области. Для оценки некоторых характеристик сигнала во временной области часто используется другая характеристика – корреляционная функция.

Корреляционная функция детерминированного сигнала $x(t)$ в общем случае определяется выражением

$$B(\tau) = \int_{-\infty}^{\infty} x(t)x^*(t - \tau)dt \quad (1.50)$$

Для сигналов, длительность которых во времени ограничена интервалом $t_1 \leq t \leq t_2$, (так называемые финитные сигналы) пределы интегрирования (1.50) соответствующим образом сужаются.

Для вещественных сигналов приведенная формула упрощается из-за отсутствия необходимости проводить операцию комплексного сопряжения.

В качестве примера рассмотрим вычисление корреляционной функции прямоугольного импульса, определенного на интервале от 0 до T и имеющего амплитуду X_m . Для положительных значений сдвига τ легко получить

$$B(\tau) = \int_{\tau}^T X_m^2 dt = X_m^2 (T - \tau)$$

Для отрицательных значений τ аналогично находится

$$B(\tau) = \int_0^{T-\tau} X_m^2 dt = X_m^2 (T - |\tau|)$$

Объединяя оба выражения для общего интервала $-T \leq \tau \leq T$, можно записать окончательное выражение для искомой корреляционной функции

$$B(\tau) = X_m^2 (T - |\tau|)$$

Графическая иллюстрация процесса вычисления корреляционной функции для указанного случая представлена на рис.1.28.

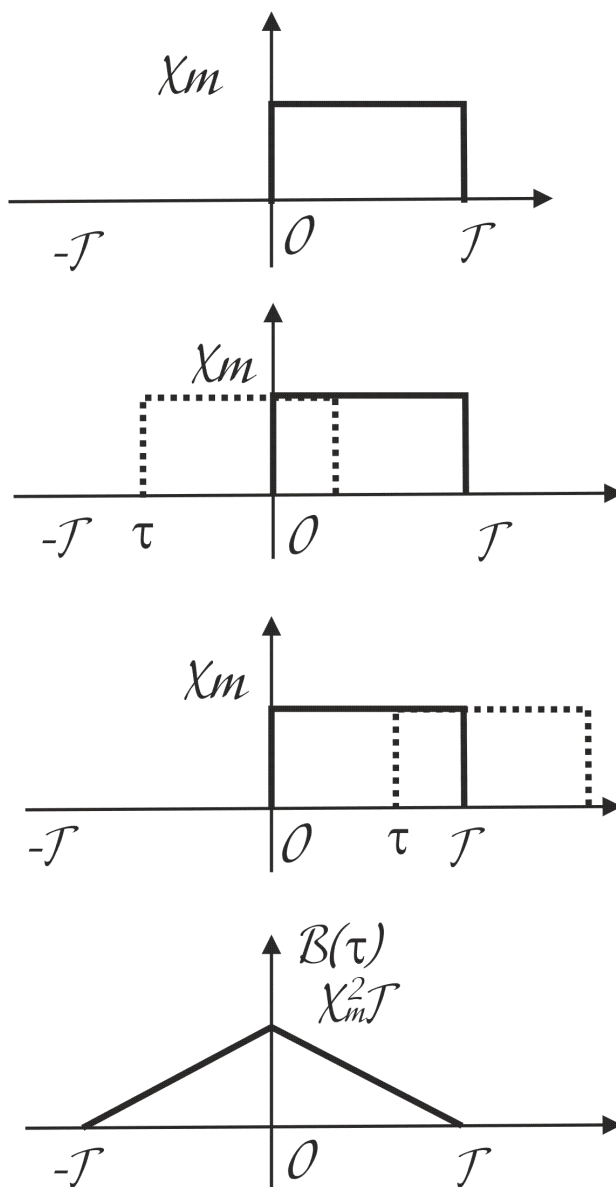


Рис.1.28. Вычисление корреляционной функции прямоугольного импульса.

Аргументом корреляционной функции является переменная τ , которая показывает временной сдвиг копии сигнала относительно своего оригинала. Поскольку сдвиг может осуществляться в обе стороны по оси времени, то область определения корреляционной функции для финитного сигнала вдвое больше области задания самого сигнала.

Построение графика корреляционной функции для рассмотренного примера с помощью программы MathCad показано на рис.1.29.

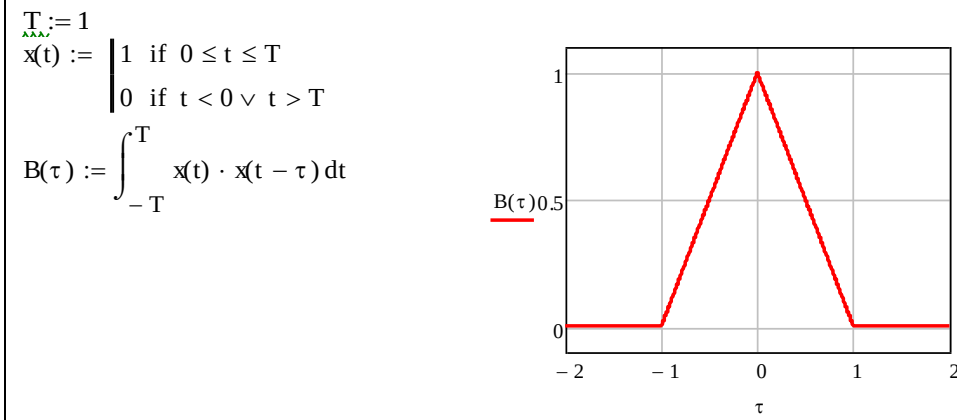


Рис.1.29. Листинг построения графика корреляционной функции

Из рассмотренного ясно, что корреляционная функция характеризует степень связи сигнала со своей копией, сдвинутой по оси времени, при этом максимальное значение корреляционной функции равно энергии сигнала (если сигнал имеет ограниченную энергию) и это максимальное значение достигается при нулевом сдвиге между оригиналом и его копией, то есть

$$B(0) = \int_{-\infty}^{\infty} x(t)x(t)dt = \mathcal{E} \quad (1.51)$$

Введенная корреляционная функция сигнала является четной, то есть

$$B(\tau) = B(-\tau) \quad (1.52)$$

и для ее определения могут использоваться две равнозначные формы записи

$$B(\tau) = \int_{-\infty}^{\infty} x(t)x(t - \tau)dt = \int_{-\infty}^{\infty} x(t)x(t + \tau)dt \quad (1.53)$$

Рассмотренная корреляционная функция характеризует связь сигнала со своей собственной копией, поэтому она также называется автокорреляционной функцией. Для характеристики степени связи двух различных сигналов используется вводимая подобным же образом взаимная корреляционная функция

$$B_{12}(\tau) = \int_{-\infty}^{\infty} x_1(t)x_2(t - \tau)dt \quad (1.54)$$

Рассмотренная ранее автокорреляционная функция $B(\tau)$ является частным случаем $B_{12}(\tau)$ при $x_1(t) = x_2(t)$.

Взаимная корреляционная функция также может быть представлена двумя равнозначными выражениями

$$B_{12}(\tau) = \int_{-\infty}^{\infty} x_1(t)x_2(t - \tau)dt = \int_{-\infty}^{\infty} x_1(t)x_2(t + \tau)dt, \quad (1.55)$$

но в отличие от автокорреляционной функции значение $B_{12}(0)$ из (1.54) и (1.55) не обязательно является максимальным значением, а сама взаимная корреляционная функция

не обязательно является четной относительно τ . Этот факт проиллюстрирован на рис.1.30, который визуализирует процесс вычисления взаимной корреляционной функции прямоугольного и треугольного импульсов.

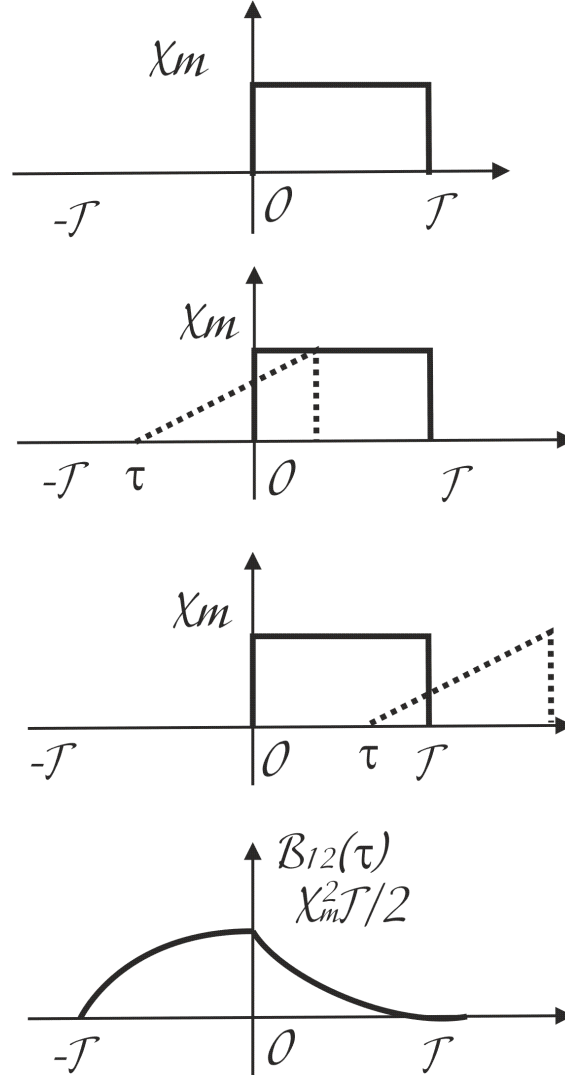


Рис.1.30. Вычисление взаимной корреляционной функции

Для отрицательных значений τ нетрудно записать

$$B_{12}(\tau) = \int_0^{T-\tau} \frac{X_m^2}{T} (t - \tau) dt = \frac{X_m^2}{2T} (T^2 - \tau^2),$$

а для положительных значений τ аналогично получаем

$$B_{12}(\tau) = \int_{\tau}^T \frac{X_m^2}{T} (t - \tau) dt = \frac{X_m^2}{2T} (\tau - T)^2.$$

Полученная в результате взаимная корреляционная функция двух сигналов не обладает свойством четности, поэтому в данном случае нам не удастся объединить выражения для положительных и отрицательных значений τ в одну общую формулу, как это было сделано при расчете автокорреляционной функции (1.51)-(1.53).

Если к уравнению, описывающему определение корреляционной функции (1.50) применить правило свертки для изображений Фурье, то легко получить, что

$$B(\tau) = \int_{-\infty}^{\infty} x(t)x(t-\tau)dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega)S^*(\omega)e^{-j\omega\tau}d\omega \quad (1.56a)$$

или

$$B(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S^2(\omega)e^{-j\omega\tau}d\omega \quad (1.56b)$$

что в соответствии со свойствами преобразования Фурье позволяет записать

$$S^2(\omega) = \int_{-\infty}^{\infty} B(\tau)e^{j\omega\tau}d\tau \quad (1.57)$$

Таким образом, приведенные уравнения (1.56) и (1.57) показывают, что прямое преобразование Фурье от корреляционной функции дает спектральную плотность энергии сигнала, а обратное преобразование от энергии сигнала дает его корреляционную функцию.

Степень близости двух сигналов можно оценивать и иным способом. Так например, физике возник термин "когерентность" в приложении к описанию суммы гармонических колебаний с совпадающими фазами. Постепенно термин получил более широкое распространение, подразумевая связь между фазами более сложных сигналов.

В качестве меры когерентности используют несколько различных критериев, один из которых, например, можно представить в следующем виде

$$K = \frac{\int_{-\infty}^{\infty} x_1(t)x_2(t)dt}{\int_{-\infty}^{\infty} |x_1(t)| |x_2(t)|dt}.$$

В радиотехнике и особенно радиолокации понятием когерентности характеризуют степень согласованности во времени двух колебательных процессов. При этом колебания полагают когерентными, если разность их фаз за время наблюдения остается постоянной или изменяется по известному закону.

1.10. Вопросы для самоконтроля и повторения

1. Дайте определение понятию информация.
2. Какие системы называются информационными?
3. Какова связь между понятиями информация, данные и сообщение?
4. Что такое сигнал?
5. Какие компоненты сигнала выделяют в кибернетике?
6. Какова количественная мера информации?
7. Как определяется информационная емкость сигнала?
8. Что такое энтропия дискретного источника информации?
9. Что такое энтропия конечного ансамбля?
10. Перечислите основные свойства энтропии.
11. Как определяется техническая скорость передачи информации?
12. Чем определяется информационная пропускная способность канала связи?
13. Дайте определение аналогового сигнала.

14. Дайте определение дискретного сигнала.
15. Дайте определение цифрового сигнала.
16. Дайте определение модулирующего, модулированного и манипулированного сигналов.
17. Какие виды функциональных пространств вам известны?
18. Каковы основные свойства метрического пространства?
19. Каковы основные свойства линейного пространства?
20. Каковы основные свойства унитарного пространства?
21. Какие функции называются ортогональными?
22. Какое пространство называется Гильбертовым?
23. Какая система функций называется полной?
24. В чем состоят преимущества использования ортогональной системы функций?
25. Какой сигнал называется гармоническим?
26. Перечислите основные параметры гармонического сигнала.
27. Каково среднее значение гармонического сигнала за период?
28. Что такое действующее значение гармонического сигнала?
29. Перечислите условия Дирихле.
30. Сформулируйте теорему Фурье.
31. Дайте определение амплитудного спектра сигнала.
32. Дайте определение фазового спектра сигнала.
33. Какой тип спектра характерен для периодических сигналов?
34. Каким образом изменяется спектр при временном сдвиге сигнала?
35. Каким образом изменяется спектр сигнала при добавлении к нему постоянной составляющей?
36. Как вводится понятие спектральной плотности сигнала?
37. Запишите уравнения для прямого и обратного преобразования Фурье.
38. Перечислите основные свойства преобразования Фурье.
39. Найдите преобразование Фурье для прямоугольного импульса.
40. Найдите преобразование Фурье для радиоимпульса.
41. Как выглядит преобразование Фурье для единичного импульса, единичного скачка, гармонического сигнала и постоянного сигнала?
42. Как определяется корреляционная функция детерминированного сигнала?
43. Что такое взаимная корреляционная функция двух сигналов?

ГЛАВА 2. Модулированные сигналы

2.1. Общие сведения о модуляции

Для передачи сигналов на большие расстояния с помощью радиоволн используются высокочастотные колебания с некоторой частотой ω_0 , которая называется несущей частотой. Информация, передаваемая при этом, должна быть некоторым образом заложена в параметры данного колебания. В наиболее общем виде информационный радиосигнал, можно представить в виде модулированного колебания

$$x(t) = A(t) \cos[\omega_0 t + \theta(t)] = A(t) \cos \psi(t), \quad (2.1)$$

где $A(t)$ и $\theta(t)$ являются информационными параметрами передаваемого сигнала и изменяются во времени соответствующим образом. Схема формирования модулированного сигнала приведена на рис.2.1.

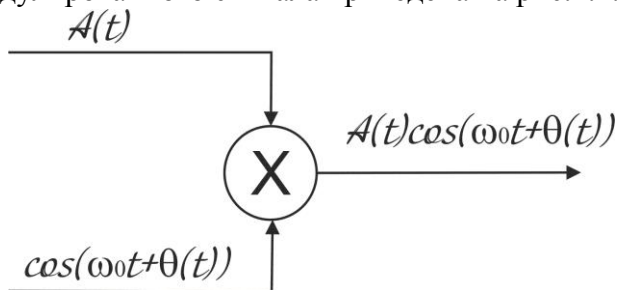


Рис.2.1. Схема формирования модулированного сигнала

Если $A(t) = \text{const}$ и $\theta(t) = \text{const}$, то приведенное уравнение (2.1) описывает обычное гармоническое колебание. Если же $A(t)$ и $\theta(t)$, представляют собой функции времени, задаваемые специально для передачи информации, то колебание становится **модулированным**.

В зависимости от того какой именно параметр изменяется, различают два основных вида модуляции: амплитудную (АМ) и угловую, последняя в свою очередь подразделяется на два типа: частотную (ЧМ) и фазовую (ФМ). Частотная и фазовая модуляции тесно связаны между собой, и различие проявляется только в характере изменения во времени функции $\psi(t)$ при одной и той же модулирующей функции.

Для качественной передачи информации частота несущего колебания ω_0 должна быть существенно больше, чем наивысшая частота в спектре передаваемого сообщения. Поэтому чем выше требуемая скорость передачи информации (чем шире спектр передаваемого информационного сигнала), тем выше должна быть и несущая частота.

Данное положение позволяет считать $A(t)$ и $\theta(t)$ функциями, медленно изменяющимися во времени. Термин «медленный» в данном контексте носит относительный характер, предполагая, что относительные изменения $A(t)$ и $\theta(t)$ за один период несущего колебания $T_0 = 2\pi/\omega_0$ много меньше единицы.

Аппроксимируя изменения амплитуды $A(t)$ за один период несущего колебания T_0 линейным приближением, приведенное выше условие «медленного» изменения можно записать в виде

$$\frac{|dA(t)/dt|T_0}{|A(t)|} \ll 1, \quad \text{для } \forall t.$$

Разделив обе части неравенства на T_0 и переходя к частоте ω_0 , получаем

$$\left| \frac{dA(t)/dt}{A(t)} \right| \ll \frac{\omega_0}{2\pi}.$$

Условие «медленного» изменения для фазовой функции $\theta(t)$, получается аналогичным образом с поправкой на то, линейная аппроксимация изменения частоты за один период несущего колебания будет включать не первую, а вторую производную от фазы

$$\left| \frac{d\omega(t)/dt}{\omega(t)} \right| T_0 = \left| \frac{d^2\theta(t)/dt^2}{\omega(t)} \right| T_0 \ll 1$$

или принимая во внимание близость $\omega(t)$ и ω_0 ,

$$\left| \frac{d^2\theta(t)}{dt^2} \right| \ll \frac{\omega_0^2}{2\pi} \text{ для } \forall t.$$

В большинстве реальных ситуаций полученные условия «медленного» изменения выполняются. Поэтому практически любой радиосигнал можно считать «узкополосным» относительно несущей частоты.

2.2. Амплитудная и угловая модуляция

2.2.1. Амплитудная модуляция

Амплитудная модуляция (АМ) 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 обусловлено низким КПД использования энергии модулированных сигналов.

При использовании амплитудной модуляции огибающая несущего колебания изменяется по закону, определяемому функцией $A(t)$, а значения начальной фазы и частоты несущего колебания поддерживаются постоянными

$$x(t) = A(t) \cos[\omega_0 t + \theta_0]. \quad (2.2)$$

Если реализовывать приведенную форму сигнала (2.2) без каких-либо изменений, то, к сожалению, при отрицательных значениях $A(t)$, то огибающая выделяемая амплитудным детектором из принятого сигнала не будет соответствовать исходному модулирующему сигналу. Поэтому в модулирующий множитель вводят постоянную составляющую. Результаты процесса модуляции без введения постоянной составляющей ($x_1(t)$) и при ее наличии ($x_2(t)$) приведены на рис.2.2.

Значение M характеризует *глубину* амплитудной модуляции.

В простейшем случае, называемом **тональной амплитудной модуляцией**, в качестве модулирующего сигнала выбирается колебание вида

$$A(t) = A_0 (1 + M \cos(\Omega t + \theta_{\Omega 0})), \quad (2.3)$$

где M – коэффициент амплитудной модуляции.

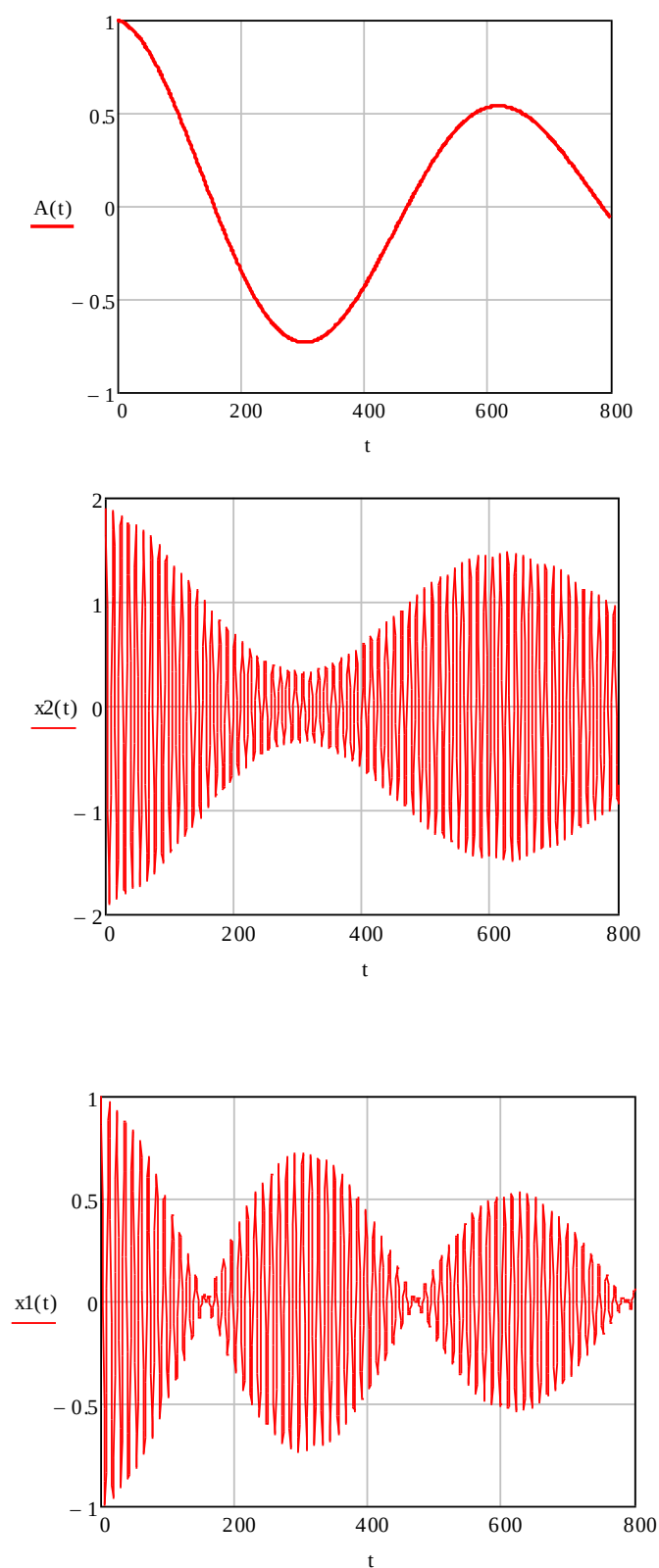


Рис.2.2. Формирование амплитудно-модулированного сигнала

Тогда итоговое колебание может быть описано следующей функцией

$$x(t) = (A_0 + A(t)) \cos[\omega_0 t + \theta_0] = A_0 (1 + M \cos(\Omega t + \theta_{\Omega 0})) \cos[\omega_0 t + \theta_0]. \quad (2.4)$$

График результирующего модулированного колебания представлен на рис.2.3.

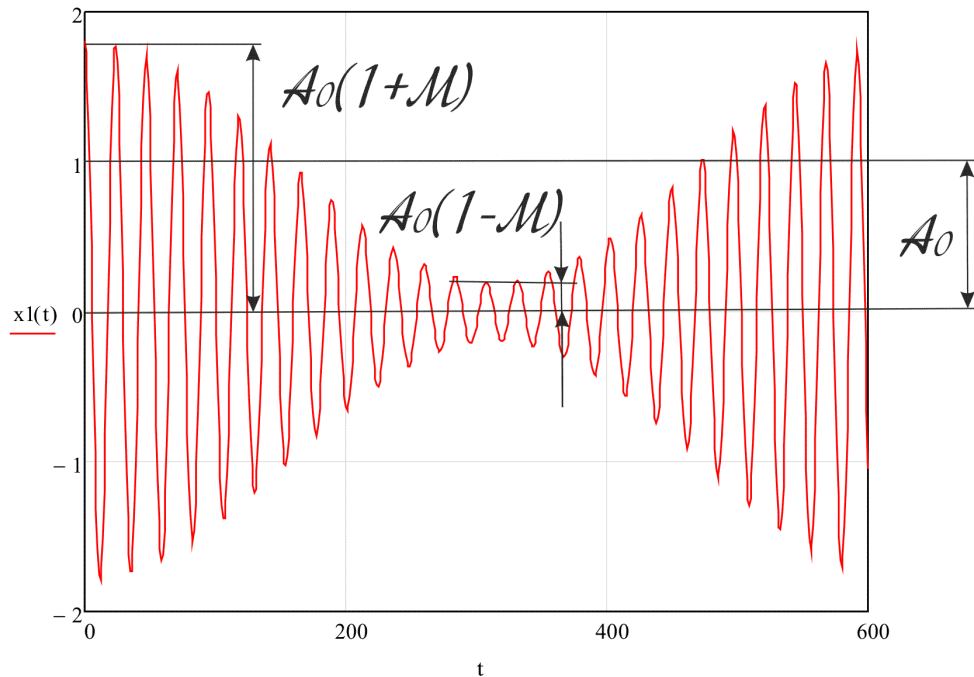


Рис.2.3. График сигнала с тональной модуляцией

Произведя почленное умножение и преобразуя полученное выражение с помощью тригонометрической формулы произведения косинусов двух углов $\cos(\alpha) \cdot \cos(\beta) = (1/2)[\cos(\alpha+\beta) + \cos(\alpha-\beta)]$, нетрудно получить

$$x(t) = A_0 \cos[\omega_0 t + \theta_0] + \frac{MA_0}{2} \cos\{(\omega_0 + \Omega)t + \theta_0 + \theta_{\Omega 0}\} + \frac{MA_0}{2} \cos\{(\omega_0 - \Omega)t + \theta_0 - \theta_{\Omega 0}\}. \quad (2.5)$$

Значение M в выражениях (2.4) и (2.5) 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 $A(t)$. Малую глубину модуляции для основных гармоник модулирующего сигнала ($M \ll 1$) применять нецелесообразно, т.к. при этом мощность передаваемого информационного сигнала будет много меньше мощности несущего колебания, и мощность передатчика используется неэкономично.

Стопроцентная модуляция ($M=1$) может приводить к искажениям сигналов при перегрузках передатчика, если последний имеет ограниченный динамический диапазон по амплитуде несущих частот или ограниченную мощность передатчика (увеличение амплитуды несущих колебаний в пиковых интервалах сигнала $A(t)$ в два раза требует увеличения мощности передатчика в четыре раза).

При $M > 1$ возникает так называемая *перемодуляция*, пример которой приведен на рис. 2.4. Форма огибающей при перемодуляции искажается относительно формы модулирующего сигнала и после демодуляции, если применяются ее простейшие методы, информация может исказиться.

Начальные фазы несущего и информационного сигнала в данном случае принципиального значения не имеют, поэтому для улучшения наглядности полученной формулы их можно опустить

$$x(t) = A_0 \cos[\omega_0 t] + \frac{MA_0}{2} \cos\{(\omega_0 + \Omega)t\} + \frac{MA_0}{2} \cos\{(\omega_0 - \Omega)t\}. \quad (2.6)$$

Формула (2.6) наглядно показывает, что в спектре амплитудно-модулированного сигнала по сравнению со спектрами несущего колебания и информационного сигнала происходят определенные изменения, а именно: 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. Изменение фазы происходит аналогичным образом: начальная фаза модулирующего колебания для верхней боковой частоты складывается с начальной фазой несущей, а для нижней боковой – вычитаются из фазы несущей. Графики амплитудного и фазового спектра представлены на рис.2.5а.

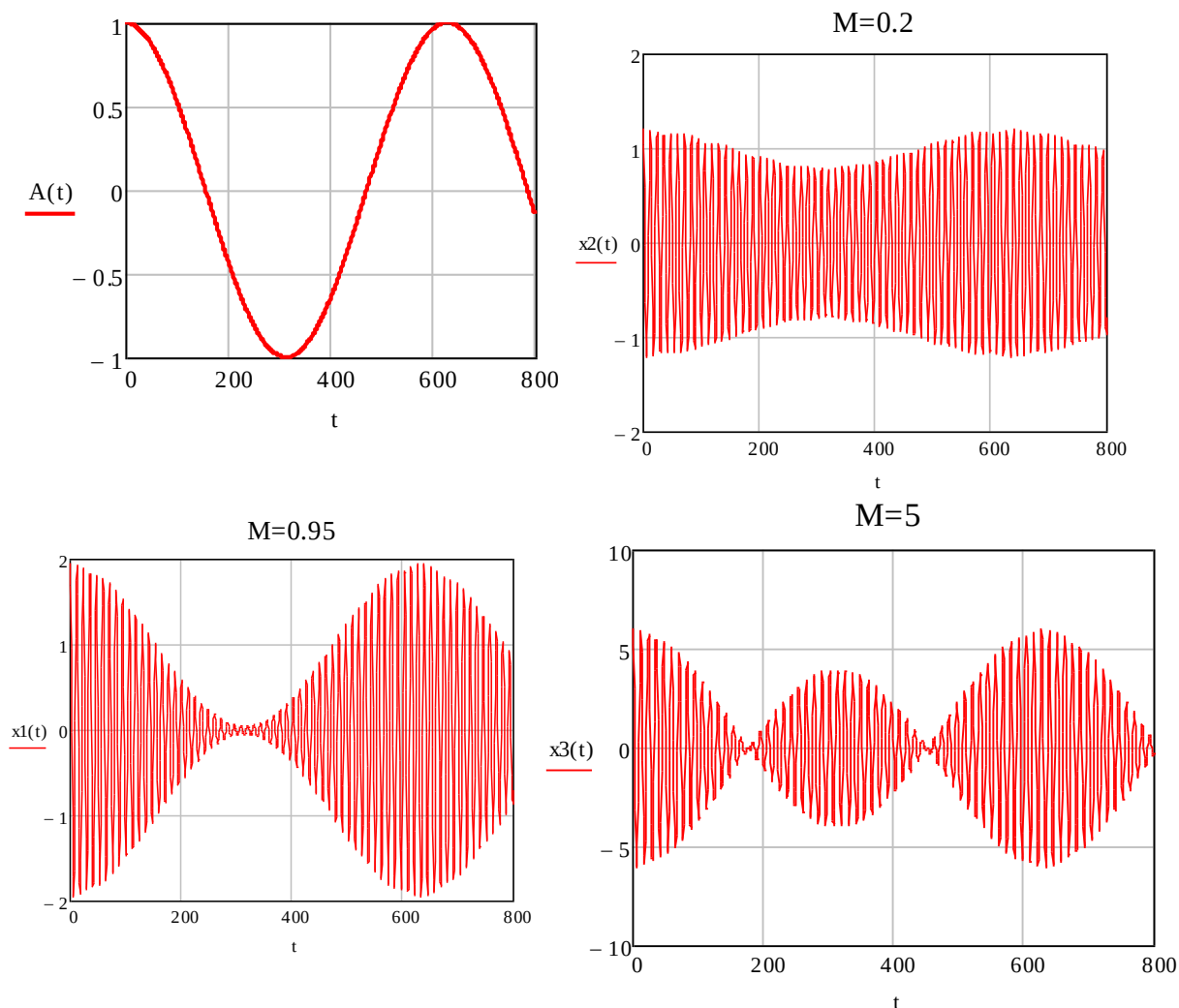


Рис.2.4. Амплитудная модуляция при различных значениях M .

Из приведенных графиков хорошо видно, что эффективная ширина спектра радиосигнала с тональной амплитудной модуляцией равна удвоенному значению частоты модулирующего колебания, то есть

$$\Delta\omega_3 = 2\Omega.$$

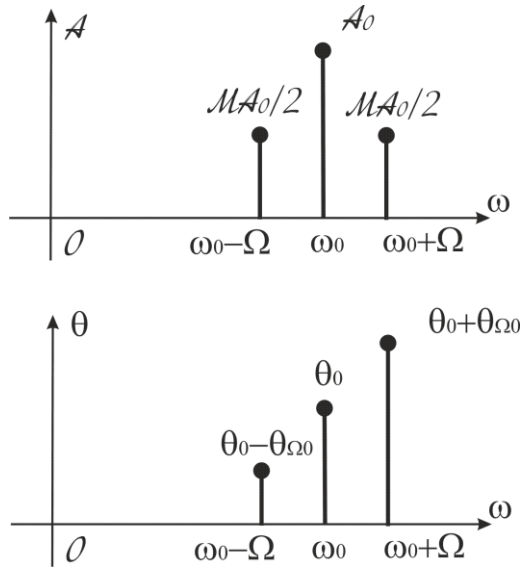


Рис.2.5. Спектры радиосигнала с тональной модуляцией

$$\begin{aligned} x(t) &= A_0 \cos(\omega_0 t + \theta_0) = \operatorname{Re}\{x^\circ(t)\} = \operatorname{Re}\{A_0 e^{j(\omega_0 t + \theta_0)}\} = \operatorname{Re}\{A_0 e^{j\theta_0} e^{j\omega_0 t}\} \\ &= \operatorname{Re}\{A_0^\circ e^{j\omega_0 t}\}, \end{aligned}$$

где

$x^\circ(t) = A_0^\circ e^{j\omega_0 t}$ комплексный сигнал с комплексной амплитудой A_0° , содержащей информацию как об амплитуде сигнала, так и о его начальной фазе, что позволяет удобно представлять периодические сигналы на комплексной плоскости.

Подобным образом можно представить и модулированный радиосигнал

$$\begin{aligned} x(t) &= \operatorname{Re}\left\{A_0 e^{j\theta_0} + \frac{MA_0}{2} e^{j(\theta_0 + \theta_{\Omega})} e^{j\Omega t} + \frac{MA_0}{2} e^{j(\theta_0 - \theta_{\Omega})} e^{-j\Omega t}\right\} e^{j\omega_0 t} \\ &= \operatorname{Re}\{A_{AM}^\circ(t) e^{j\omega_0 t}\} \end{aligned} \quad (2.7)$$

В данном случае комплексная амплитуда модулированного сигнала, называемая также комплексной огибающей АМ-сигнала, оказалась функцией времени

$$A_{AM}^\circ(t) = A_0 e^{j\theta_0} + \frac{MA_0}{2} e^{j(\theta_0 + \theta_{\Omega})} e^{j\Omega t} + \frac{MA_0}{2} e^{j(\theta_0 - \theta_{\Omega})} e^{-j\Omega t} \quad (2.8)$$

Соответствующая уравнениям (2.7) и (2.8) векторная диаграмма представлена на рис.2.6

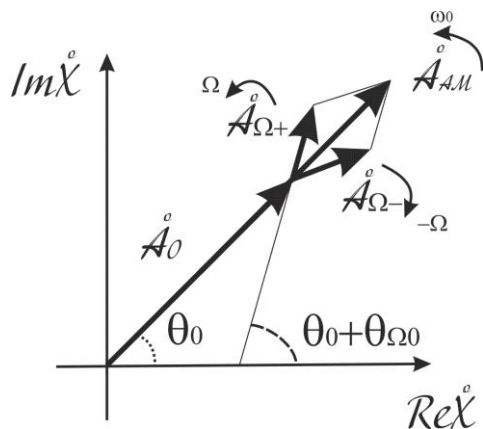


Рис.2.6. Векторная диаграмма тонального АМ-сигнала

Сигналы с тональной АМ-модуляцией часто используются в качестве тестовых сигналов для лабораторной практики и настройки аппаратуры.

Для наглядного представления процесса формирования АМ-сигналов удобно использовать их представление на комплексной плоскости на основе символического метода (метода комплексных амплитуд), широко применяемого в теоретической электротехнике. В соответствии с данным методом гармонический сигнал может быть представлен как вещественная часть комплексной функции:

Вектор $A_{AM}^\circ(t)$ в соответствии с вышеполученной формулой представляет собой сумму трех векторов, вращающихся относительно начала координат против часовой стрелки с угловой скоростью ω_0 . Вектора $\frac{MA_0}{2} e^{j(\theta_0 + \theta_{\Omega})}$ и $\frac{MA_0}{2} e^{j(\theta_0 - \theta_{\Omega})}$ помимо этого совершают вращение с угловыми скоростями Ω и $-\Omega$ соответственно. Поскольку вращение векторов $\frac{MA_0}{2} e^{j(\theta_0 + \theta_{\Omega})}$ и $\frac{MA_0}{2} e^{j(\theta_0 - \theta_{\Omega})}$ происходит синхронно и равными по величине угловыми скоростями, то они всегда располагаются симметрично относительно вектора A_0° . В результате

этого вектора $\mathbf{A}^{\circ}_{AM}(t)$ и $\mathbf{A}^{\circ}_0$ всегда коллинеарны.

Модуль вектора $\mathbf{A}^{\circ}_{AM}(t)$ изменяется по закону модуляции

$$A_{AM}(t) = A_0 (1 + M \cos(\Omega t + \theta_{\Omega 0})). \quad (2.9)$$

Подробно рассмотренный процесс монотональной амплитудной модуляции позволяет обобщить полученные результаты на случай многотональной модуляции, когда модулирующий сигнал представляет собой сумму отдельных гармоник

$$A(t) = A_0 \left(1 + \sum_{n=1}^N M_n \cos(\Omega_n t + \theta_{\Omega n}) \right) \quad (2.10)$$

при условии $\sum_{n=1}^N M_n \leq 1$.

Проделявая преобразования подобные тем, которые применялись при рассмотрении монотональной модуляции (2.9) для случая многотональной модуляции (2.10) нетрудно получить

$$\begin{aligned} x(t) = A_0 \cos[\omega_0 t + \theta_0] + \sum_{n=1}^N \frac{M_n A_0}{2} \cos\{(\omega_0 + \Omega_n)t + \theta_0 + \theta_{\Omega n}\} + \\ + \sum_{n=1}^N \frac{M_n A_0}{2} \cos\{(\omega_0 - \Omega_n)t + \theta_0 - \theta_{\Omega n}\}. \end{aligned} \quad (2.11)$$

В предельном переходе количество используемых для многотональной модуляции колебаний N может быть бесконечным, что позволяет перейти от уравнения (2.10) к еще более общему случаю формирования АМ-модулированного сигнала

$$A(t) = A_0 (1 + M_{AM} a(t)), \quad (2.12)$$

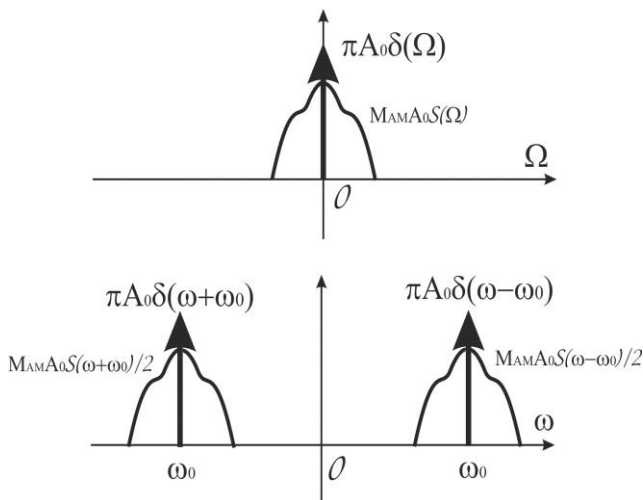
где

$a(t)$ – передаваемый информационный сигнал, имеющий спектральную плотность $S(\Omega)$.

В данном случае спектральная плотность сигнала $A(t)$ будет иметь две составляющие

$$S_A(\Omega) = 2\pi A_0 \delta(\Omega) + M_{AM} A_0 S(\Omega). \quad (2.13)$$

Спектральные плотности информационного сигнала и модулированного им высокочастотного колебания представлены на рис.2.7.



Автокорреляционная функция АМ-сигнала может быть рассчитана по формуле

$$B_{AM}(\tau) = \int_{-\infty}^{\infty} x_{AM}(t) x_{AM}(t - \tau) dt.$$

Поскольку постоянная фаза не влияет на форму получаемой автокорреляционной функции, то можно воспользоваться упрощенным выражением для сигнала $x_{AM}(t)$.

Рис.2.7. Спектры информационного и модулированного сигналов

Тогда

$$B_{AM}(\tau) = \int_{-\infty}^{\infty} A(t)A(t-\tau)\cos\omega_0 t \cos\omega_0(t-\tau)dt$$

Принимая во внимание, что

$$\cos\omega_0 t \cos\omega_0(t-\tau) = \frac{1}{2} (\cos\omega_0\tau + \cos\omega_0(2t-\tau)),$$

можно записать

$$B_{AM}(\tau) = \frac{\cos\omega_0\tau}{2} \int_{-\infty}^{\infty} A(t)A(t-\tau)dt + \frac{1}{2} \int_{-\infty}^{\infty} A(t)A(t-\tau) \cos\omega_0(2t-\tau)dt.$$

В полученной сумме первый интеграл представляет собой автокорреляционную функцию сигнала $A(t)$. Второй интеграл включает в себя осцилляции с удвоенной несущей частотой, поэтому им можно пренебречь. Тогда

$$B_{AM}(\tau) \approx \frac{\cos\omega_0\tau}{2} B_A(\tau). \quad (2.14)$$

Полная энергия модулированного сигнала с учетом усреднения по высокой частоте составит

$$\mathfrak{E} = B_{AM}(0) \approx \frac{1}{2} B_A(0).$$

Для случая бесконечной энергии сигнала при определении автокорреляционной функции необходимо выполнить предельный переход

$$B_{AM}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x_{AM}(t)x_{AM}(t-\tau)dt.$$

Обозначим соответствующими индексами составляющие АМ-сигнала

$$x(t) = x_{\omega}(t) + x_{\Omega}(t) + x_{-\Omega}(t).$$

Тогда средняя мощность сигнала также может быть представлена как сумма трех составляющих (взаимное влияние слагаемых не учитывается вследствие ортогональности)

$$P = P_{\omega} + P_{\Omega} + P_{-\Omega} = \frac{A_0^2}{2} + \frac{M^2 A_0^2}{8} + \frac{M^2 A_0^2}{8} = \frac{A_0^2}{2} \left(1 + \frac{M^2}{2} \right).$$

Под **полезной мощностью модулированных сигналов** понимают ту часть мощности, которая используется для передачи информации, т.е. в данном случае мощность боковых частот. Коэффициент полезного действия данного типа модуляции определяется отношением мощности боковых частот к общей средней мощности модулированного сигнала

$$\eta_{AM} = \frac{P_{\Omega} + P_{-\Omega}}{P} = \frac{M^2}{2 + M^2}. \quad (2.15)$$

График зависимости (2.15) коэффициента полезного действия АМ-модуляции от величины коэффициента модуляции представлен на рис.2.8.

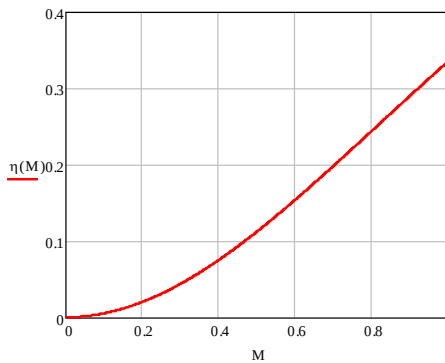


Рис.2.8. График зависимости $\eta_{AM}(M)$

Как можно видеть из приведенного графика даже при предельном значении $M=1$ коэффициент полезного действия достигает только 33%.

Для повышения коэффициента полезного действия предложено несколько способов модификации процесса амплитудной модуляции. Самым первым и очевидным из них является применение так называемой **балансовой амплитудной модуляции**.

Как следует из вышеприведенных данных, основная доля мощности АМ – сигнала приходится на несущую частоту. При балансной модуляции 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

$$A(t) = M \cos(\Omega t).$$

получаем

$$x(t) = A_0 M \cos(\Omega t) \cos \omega_0 t = \frac{A_0 M}{2} (\cos\{(\omega_0 + \Omega)t\} + \cos\{(\omega_0 - \Omega)t\}), \quad (2.16)$$

т.е. биения двух одинаковых по амплитуде гармонических сигналов с верхней и нижней боковыми частотами. По существу, однотоновый модулирующий сигнал в этом случае переносится на биения двух высоких частот. Физическая сущность подавления несущей частоты заключается в следующем. При переходе огибающей биений $A(t)$ через нуль фаза несущей частоты высокочастотного заполнения скачком изменяется на 180° , поскольку функция косинуса огибающей имеет разные знаки слева и справа от нуля. График, приведенный на рис.2.9, иллюстрирует данный процесс.

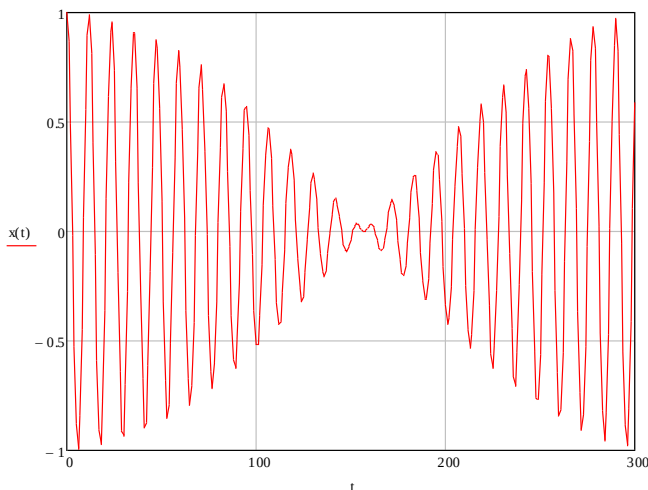


Рис.2.9. График сигнала с балансной гармонической модуляцией.

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. В принципе, синхронное детектирование 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 данном типе модуляции 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.

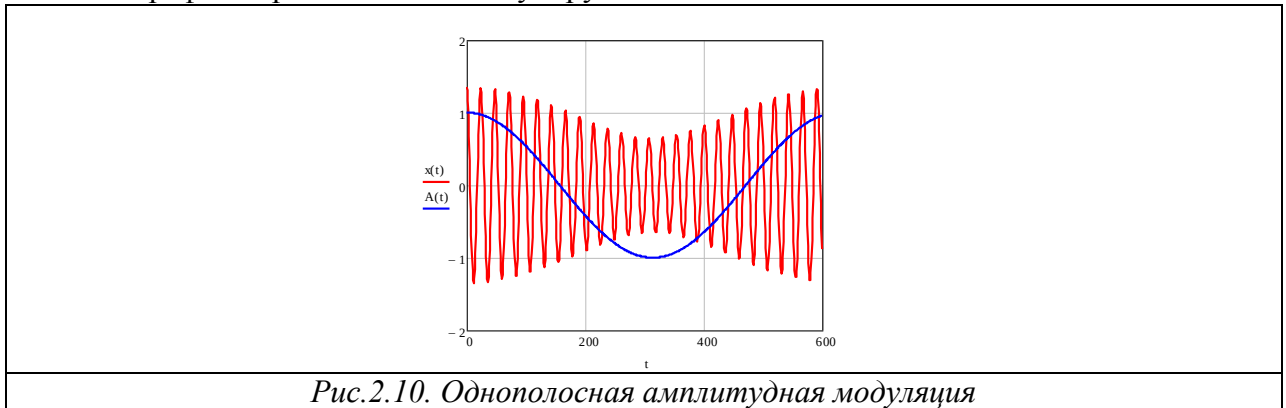
Другой способ повышения КПД базируется на идее о том, что при идентичности информации в группах верхних и нижних боковых частот нет никакой необходимости в их одновременной передаче. Одна из них перед подачей сигнала в канал связи может быть удалена, чем достигается двукратное сокращение полосы занимаемых сигналом частот.

Сигналы такого типа получили наименование ОБП-сигналы (сигналы с одной боковой полосой) или SSB (single side band).

Уравнение для данного вида модуляции легко получить из предыдущих путем сокращения соответствующего слагаемого. Так, например, для случая подавления нижней боковой полосы можно записать

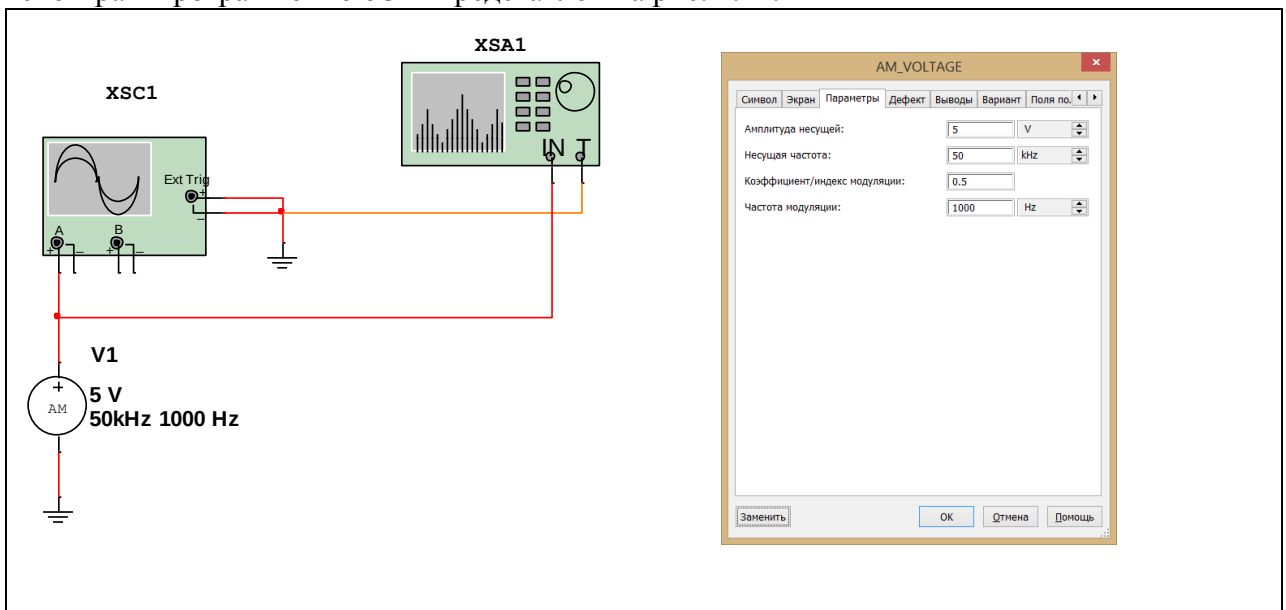
$$x(t) = A_0 \cos[\omega_0 t + \theta_0] + \sum_{n=1}^N \frac{M_n A_0}{2} \cos\{(\omega_0 + \Omega_n)t + \theta_0 + \theta_{\Omega n}\} \quad (2.17)$$

Пример сигнала с таким видом модуляции для случая $n=1$ приведен на рис.2.10. На этом же графике представлен и модулирующий сигнал.



Пример 2.1.

Пример генерирования амплитудно-модулированного сигнала и определения его спектра в программе Multisim представлен на рис. 2.11.



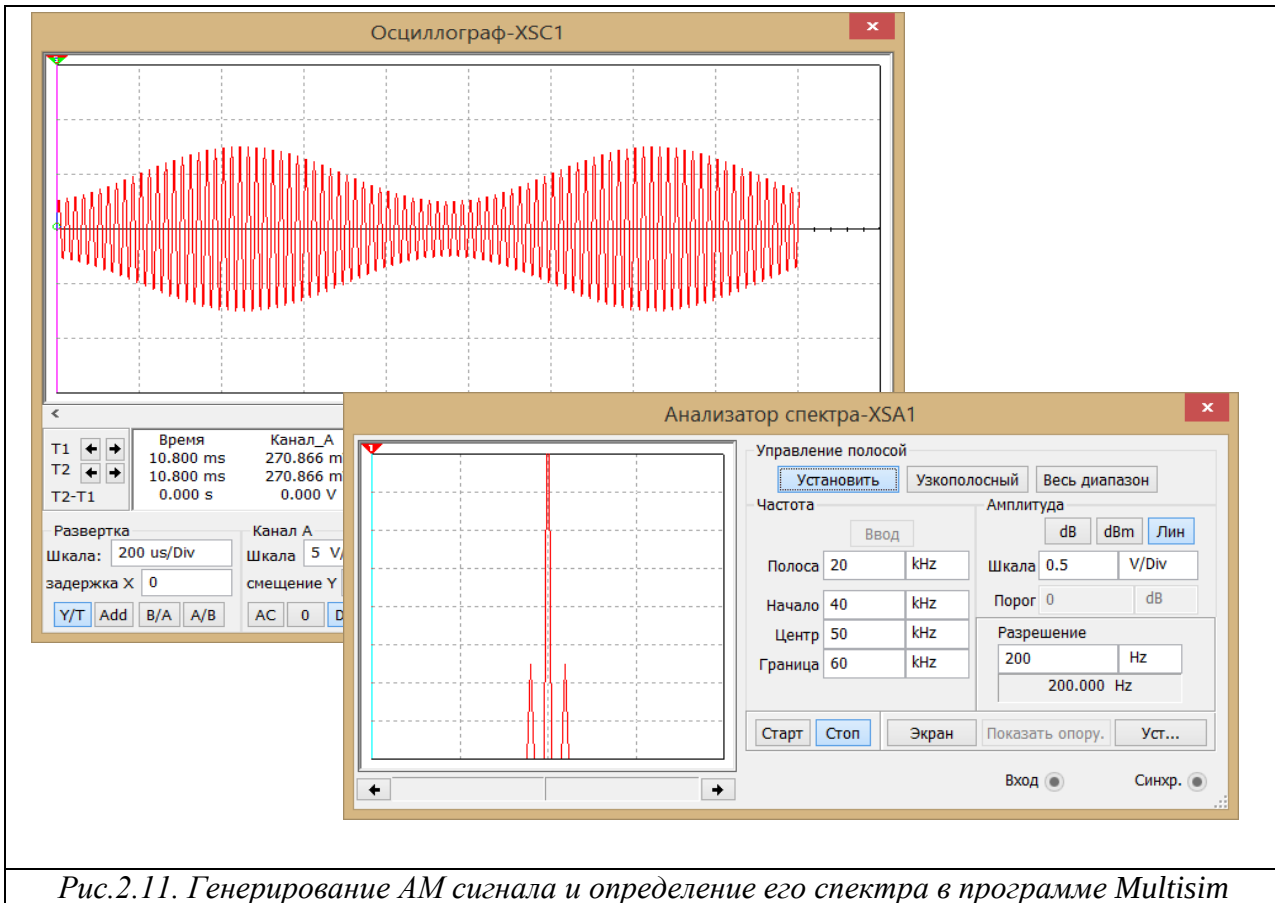


Рис.2.11. Генерирование АМ сигнала и определение его спектра в программе Multisim

2.2.2. Угловая модуляция

Как отмечалось выше в наиболее общем виде информационный радиосигнал, можно представить в виде модулированного колебания

$$x(t) = A(t) \cos[\omega_0 t + \theta(t)] = A(t) \cos \psi(t), \quad (2.18)$$

где $A(t)$ и $\theta(t)$ являются информационными параметрами передаваемого сигнала.

Если $A(t) = \text{const}$ и $\theta(t) \neq \text{const}$, то получается модель радиосигнала с угловой модуляцией.

Вводя определение мгновенной частоты колебаний как производной от полной фазы колебания

$$\omega(t) = \frac{d}{dt} \psi(t) = \omega_0 + \frac{d}{dt} \theta(t),$$

можно логичным образом расширить понятие угловой частоты колебаний, используемое для описания гармонических сигналов.

Тогда полную фазу произвольного радиосигнала можно определить как интеграл от предыдущего выражения

$$\psi(t) = \int_{-\infty}^t \frac{d}{dt} \omega(t) dt.$$

Данный интеграл представим в виде суммы трех слагаемых

$$\psi(t) = \omega_0 t + \theta(t) + \theta_0, \quad (2.19)$$

представляющих собой линейный набег фазы, фазовую функцию и начальную фазу.

Угловую модуляцию принято подразделять на **фазовую** и **частотную** модуляцию.

Рассмотрим разницу между ними. В общем случае выражение для сигнала с угловой модуляцией записывается в виде

$$x(t) = A_0 \cos(\omega_0 t + \theta(t) + \theta_0), \quad (2.20)$$

где $A_0 = \text{const}$.

Если информационный сигнал задается функцией времени $a(t)$, то для сигнала с фазовой модуляцией справедливо

$$x_{\text{фм}}(t) = A_0 \cos(\omega_0 t + ka(t)),$$

где $k = \text{const}$ – некоторый коэффициент.

При частотной модуляции

$$x_{\text{чм}}(t) = A_0 \cos\left(\omega_0 t + k \int_{-\infty}^t a(t) dt\right). \quad (2.21)$$

Сравнивая приведенные уравнения для сигналов с фазовой и частотной модуляцией нетрудно заметить, что они тесно связаны между собой. Особенно наглядно эта связь прослеживается при выборе в качестве модулирующего сигнала

$$a(t) = \cos \Omega t.$$

В этом случае сигнал с фазовой модуляцией можно представить в виде

$$x_{\text{фм}}(t) = A_0 \cos(\omega_0 t + \Delta\varphi \cos \Omega t), \quad (2.22)$$

где $\Delta\varphi$ – величина, называемая девиацией фазы.

Вводя аналогичным образом величину, называемую девиацией частоты $\Delta\omega$, для частотно-модулированного сигнала запишем

$$\omega(t) = \omega_0 + \Delta\omega \cos \Omega t,$$

что после интегрирования приводит к

$$x_{\text{чм}}(t) = A_0 \cos\left(\omega_0 t + \frac{\Delta\omega}{\Omega} \sin \Omega t + \theta_0\right). \quad (2.23)$$

Для унификации рассматриваемых процессов модуляции (2.22) и (2.23) вводится индекс гармонической угловой модуляции, который для частотной модуляции равен

$$\beta = \frac{\Delta\omega}{\Omega}, \quad (2.24)$$

а для фазовой модуляции анализируется сигнал

$$x_{\text{фм}}(t) = A_0 \cos(\omega_0 t + \beta \sin \Omega t + \theta_0). \quad (2.25)$$

Вид графика сигнала с фазовой модуляцией представлен на рис.2.12. (для целей наглядности величина β специально завышена).

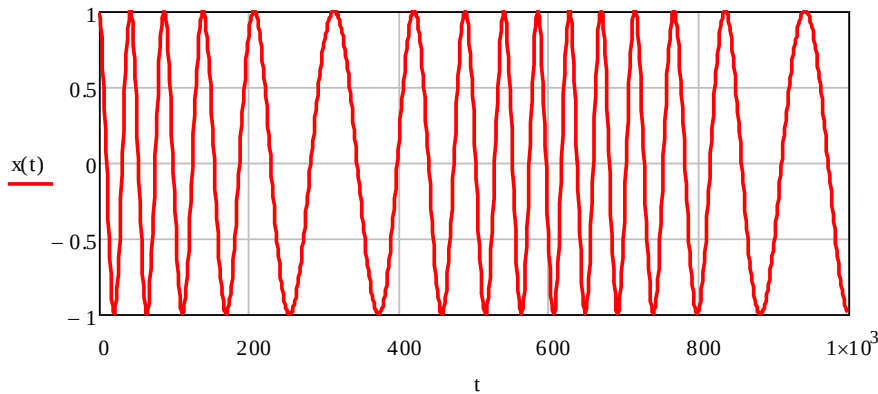


Рис.2.12. График сигнала с фазовой модуляцией

Для вычисления спектра радиосигнала при гармонической угловой модуляции удобно воспользоваться функциями Бесселя, с помощью которых можно выполнить следующее разложение комплексной экспоненты

$$e^{j\beta \sin t} = \sum_{k=1}^{\infty} J_k(\beta) e^{jkt},$$

где $J_k(\beta)$ – функция Бесселя 1-го рода порядка k от вещественного аргумента.

С применением функций Бесселя для сигнала с угловой модуляцией нетрудно записать

$$x_{\text{ум}}(t) = \text{Re}\{A_0 e^{j(\omega_0 t + \theta_0)} \sum_{k=1}^{\infty} J_k(\beta) e^{jk\Omega t}\} = A_0 \sum_{k=1}^{\infty} J_k(\beta) \cos((\omega_0 + k\Omega)t + \theta_0).$$

Из полученного видно, что амплитудный спектр сигнала с угловой модуляцией будет представлять собой бесконечную сумму гармоник, величина которых определяется как $A_0 |J_k(\beta)|$.

Схематический график такого спектра приведен на рис.2.13.

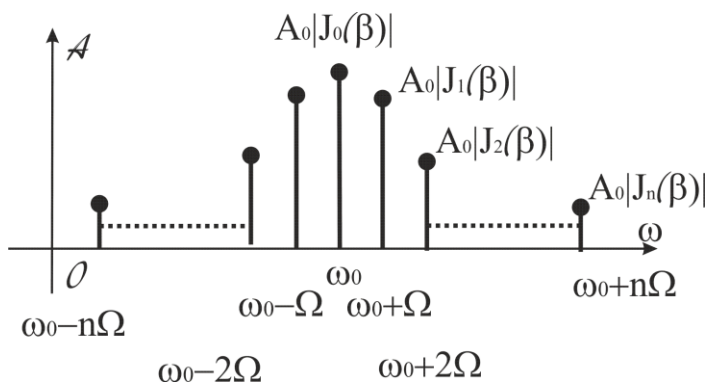


Рис.2.13. Амплитудный спектр сигнала с гармонической угловой модуляцией

Анализ структуры полученного для сигнала с угловой модуляцией уравнения показывает потенциально большую КПД, по отношению к способу амплитудной модуляции.

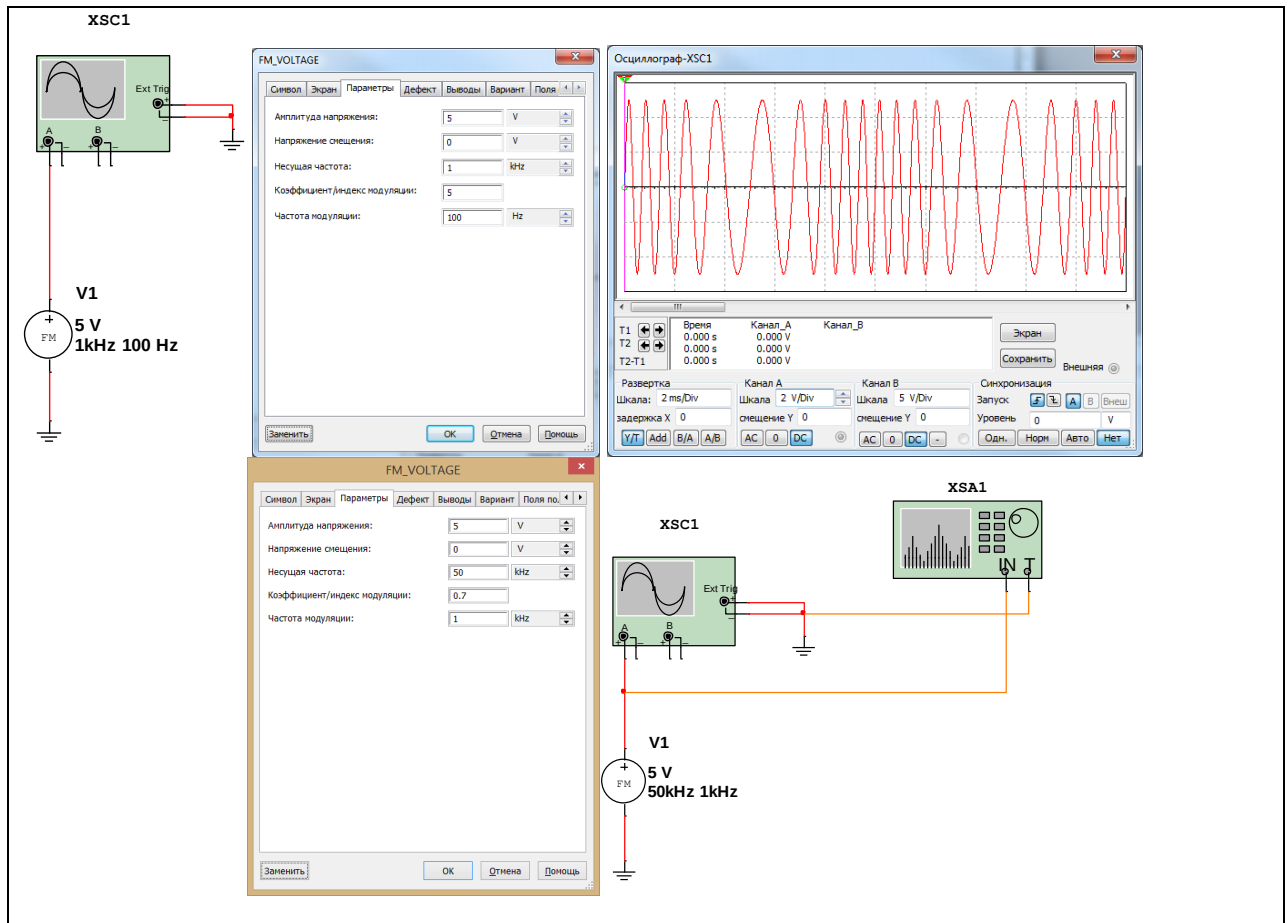
При малой величине индекса β значимые амплитудные значения имеют только первые гармоники. С ростом величины β количество значимых боковых составляющих увеличивается, а энергия сигнала перераспределяется на боковые составляющие. Функции Бесселя имеют колебательный характер, поэтому спектр при удалении от несущей частоты ω_0 спадает немонотонно.

В случае если гармоническая угловая модуляция проводится с малым значением индекса модуляции $\beta \ll 1$, можно произвести следующие упрощения ($\cos 0 = 1$, $\sin \alpha = \alpha$) уравнения для сигнала с угловой модуляцией.

$$\begin{aligned} x(t) &= A_0 \cos(\omega_0 t + \beta \sin \Omega t + \theta_0) = \\ &= A_0 \cos(\omega_0 t + \theta_0) \cos(\beta \sin \Omega t) - A_0 \sin(\omega_0 t + \theta_0) \sin(\beta \sin \Omega t) = \\ &= A_0 \cos(\omega_0 t + \theta_0) - \beta A_0 \sin \Omega t \sin(\omega_0 t + \theta_0) = \\ &= A_0 \cos(\omega_0 t + \theta_0) + \frac{\beta A_0}{2} \cos((\omega_0 + \Omega)t + \theta_0) - \frac{\beta A_0}{2} \cos((\omega_0 - \Omega)t + \theta_0). \end{aligned}$$

Пример 2.2.

Пример генерирования частотно-модулированного сигнала в программе Multisim с различным индексом модуляции представлен на рис. 2.14.



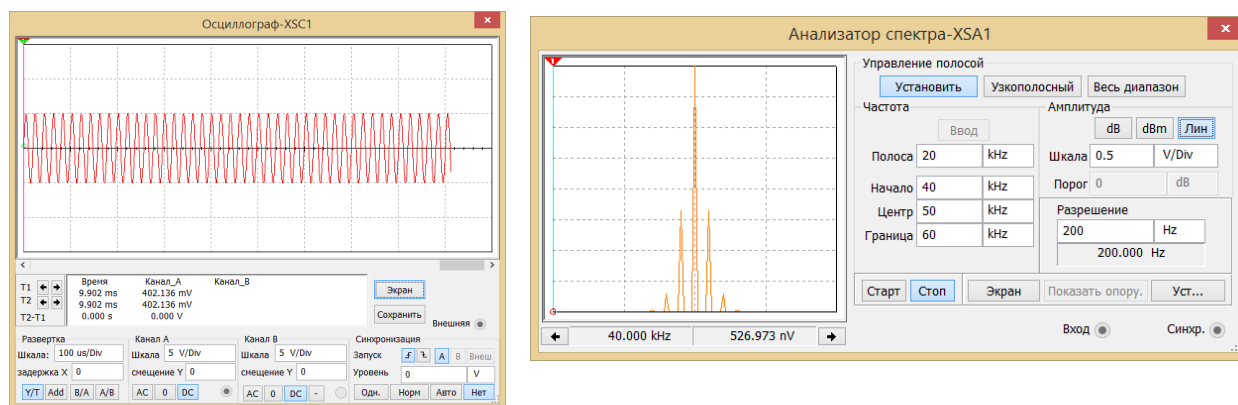


Рис.2.14. Генерирование частотно-модулированного сигнала и определение его спектра в программе Multisim

Пример 2.3.

Пример генерирования и визуализации частотно-модулированного сигнала в программе Simulink представлен на рис. 2.15.

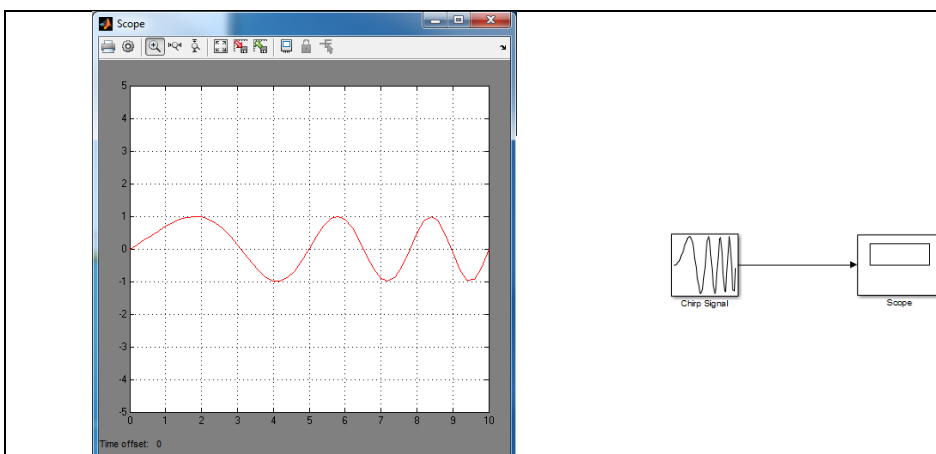


Рис.2.15. Генерирование частотно-модулированного сигнала в программе Simulink

2.3. Аналитические сигналы и преобразование Гильберта

Рассмотренные выше модулированные сигналы являются лишь простейшими видами радиотехнических сигналов. Реальные сигналы зачастую являются результатом одновременной модуляции и амплитуды и частоты.

В то же время заданный информационный сигнал $x(t)$ предполагается узкополосным. Это предположение означает, что все спектральные составляющие сигнала группируются в относительно небольшой ("узкой") по сравнению с центральной частотой полосе.

При представлении модулированного сигнала в виде

$$x(t) = A(t)\cos\psi(t)$$

вообще говоря, можно подобрать неограниченное число пар функций $A(t)$ и $\psi(t)$, удовлетворяющих приведенному уравнению.

Следовательно, необходимо принять какие-либо дополнительные правила по выбору $A(t)$ и $\psi(t)$ с целью обеспечения единственности представления. При этом естественно возникает желание, чтобы в случае простого гармонического сигнала с постоянной амплитудой и фазой

$$x(t) = A \cos[\omega_0 t + \theta_0]$$

искомые функции имели вид

$$A(t) = A, \quad \psi(t) = \omega_0 t + \theta_0.$$

Решение проблемы однозначного представления при сохранении желаемых свойств обеспечивается путем применения к исходному сигналу преобразования Гильберта

$$x_1(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x(\tau)}{t - \tau} d\tau, \quad (2.26)$$

где $x_1(t)$ - некая новая функция, называемая функцией сопряженной (по Гильберту) исходной функции $x(t)$ или ее квадратурным дополнением.

Обратное преобразование Гильберта имеет вид

$$x(t) = -\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x_1(\tau)}{t - \tau} d\tau. \quad (2.27)$$

Нетрудно заметить, что преобразование Гильберта (2.26), (2.27) по своей сути представляет свертку исходного сигнала $y(t)$ и функции $1/(\pi t)$, а прямое и обратное преобразование Гильберта отличаются только знаком

$$x_1(t) = x(t) * \left(\frac{1}{\pi t} \right),$$

$$x(t) = x_1(t) * \left(-\frac{1}{\pi t} \right).$$

Структурная схема преобразования Гильберта и соответствующая ему импульсная характеристика представлены на рис.2.16.

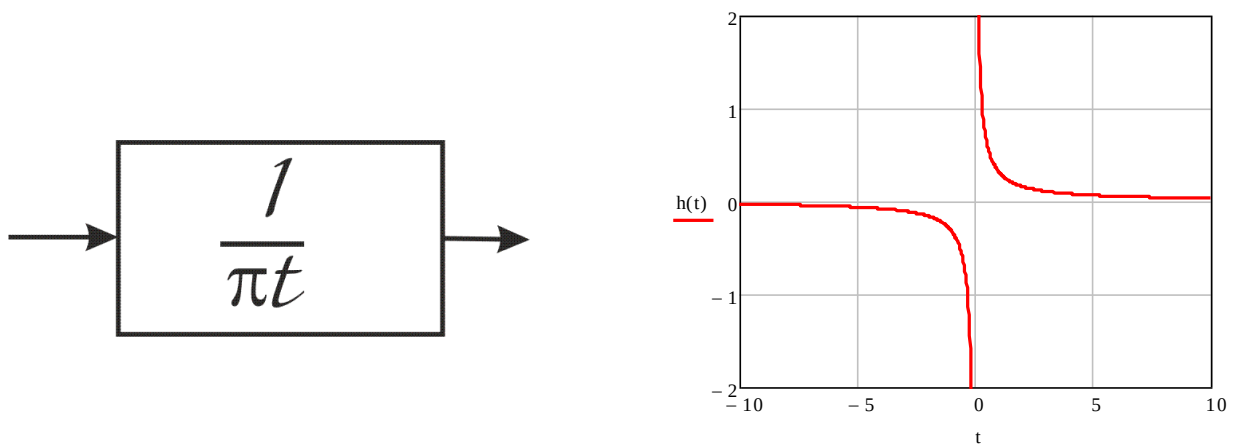


Рис.2.16. Преобразование Гильберта

Для любых произвольных функций $x(t)$ и $y(t)$, имеющих Фурье – образы $X(\omega), Y(\omega)$, и преобразования Гильберта $x_1(t) = H[x(t)], y_1(t) = H[y(t)]$, справедливы следующие свойства:

Линейность. $H[ax(t) + by(t)] = aH[x(t)] + bH[y(t)] = ax_1(t) + by_1(t)$

при любых постоянных значениях коэффициентов a и b для любых произвольных функций $x(t)$ и $y(t)$.

Сдвиг во временной области. $H[x(t - t_0)] = x_1(t - t_0)$

Преобразование константы. Преобразование постоянной составляющей сигнала, равно нулю. Это прямо следует из нечетности ядра преобразования Гильберта, то есть при преобразовании Гильберта из квадратурной составляющей исключается постоянная составляющая.

Свойство четности и нечетности определяется сдвигом всех гармоник сигнала на $\pi/2$. Соответственно четные сигналы $x(t)$ имеют нечетные сигналы-преобразования $x_1(t)$, и наоборот.

Последовательное двойное преобразование Гильберта возвращает исходную функцию с обратным знаком $H[H[x(t)]] = -x(t)$. Это определяется тем, что при двойном преобразовании фазы всех гармоники сигнала сдвигаются на π , то есть меняют свой знак.
Примечание: в силу исключения из сигнала при первом применении преобразования Гильберта постоянной составляющей, при двойном преобразовании сигнал $x(t)$ восстанавливается с исключенным средним значением по заданному временному интервалу.

Энергетическая эквивалентность:

$$\int_{-\infty}^{\infty} x^2(t) dt = \int_{-\infty}^{\infty} x_1^2(t) dt.$$

Свойство ортогональности:

$$\int_{-\infty}^{\infty} x(t)x_1(t) dt = 0$$

Свойство свертки:

$$H[x(t) * y(t)] = x_1(t) * y(t) = x(t) * y_1(t)$$

Для гармонических функций преобразование Гильберта дает очень простые соотношения

$$x(t) = \cos \omega_0 t \quad \rightarrow \quad x_1(t) = \sin \omega_0 t$$

и

$$x(t) = \sin \omega_0 t \quad \rightarrow \quad x_1(t) = -\cos \omega_0 t$$

Учитывая приведенные соотношения, для огибающей гармонического сигнала с единичной амплитудой, легко получить

$$A(t) = \sqrt{x^2(t) + x_1^2(t)} = \sqrt{(\cos \omega_0 t)^2 + (\sin \omega_0 t)^2} = 1.$$

Тогда для сигнала, представляющего собой сумму (в том числе бесконечную) синусных и косинусных слагаемых

$$x(t) = \sum_k^{\infty} (a_k \cos \omega_k t + b_k \sin \omega_k t)$$

получаем сопряженную функцию

$$x_1(t) = \sum_k^{\infty} (a_k \sin \omega_k t - b_k \cos \omega_k t).$$

Анализируя приведенные уравнения, нетрудно прийти к выводу, что спектральные функции исходного и сопряженного сигналов связаны соотношениям

$$S_1(\omega) = \begin{cases} -jS(\omega), & \text{при } \omega > 0 \\ jS(\omega), & \text{при } \omega < 0 \end{cases}$$

В силу того, что преобразование Гильберта исключает постоянную составляющую из сигнала, $S_1(0) = 0$.

На основе исходного сигнала и его квадратурного дополнения строится так называемый аналитический сигнал вида

$$x_a(t) = x(t) + jx_1(t). \quad (2.28)$$

В экспоненциальной форме аналитический сигнал можно представить в виде

$$x_a(t) = A(t)e^{j\psi(t)} = A(t)e^{j(\omega_0 t + \theta(t) + \theta_0)} = \{A(t)e^{j(\theta(t) + \theta_0)}\}e^{j\omega_0 t} = A^\circ(t)e^{j\omega_0 t}, \quad (2.29)$$

где $A^\circ(t) = A(t)e^{j(\theta(t) + \theta_0)}$ – комплексная огибающая узкополосного сигнала.

Тогда исходный сигнал может быть представлен как вещественная часть аналитического сигнала

$$x(t) = \operatorname{Re} x_a(t), \quad (2.30)$$

что позволяет однозначно определить модулирующие функции

$$\begin{aligned} A(t) &= \sqrt{x^2(t) + x_1^2(t)}, \\ \psi(t) &= \operatorname{arctg} \frac{x_a(t)}{x(t)}. \end{aligned} \quad (2.31)$$

Запишем исходный сигнал $x(t)$ в несколько ином виде

$$x(t) = A(t) \cos[\omega_0 t + \theta(t) + \theta_0] = \operatorname{Re}[A(t)e^{j(\omega_0 t + \theta(t) + \theta_0)}] = \operatorname{Re}[A^\circ(t)e^{j\omega_0 t}]. \quad (2.32)$$

Модуль комплексной огибающей содержит информацию только об амплитудной модуляции сигнала, а фазовый множитель $e^{j(\theta(t) + \theta_0)}$ – только об угловой модуляции. В целом комплексная огибающая $A^\circ(t)$ содержит всю информацию о сигнале за исключением значения частоты несущего колебания ω_0 , которое, впрочем, всегда

считается известным. Выбор центральной частоты ω_0 при определении комплексной огибающей, строго говоря, является произвольным. Для узкополосных сигналов существует интуитивно понятный вариант выбора центральной частоты равной значению несущей частоты.

Данное свойство комплексной огибающей позволяет при анализе исключить из рассмотрения частоту ω_0 , что существенно упрощает процесс анализа. (Такой метод анализа сигналов пришел из электротехники и называется методом низкочастотного эквивалента).

Кроме того комплексная огибающая наряду с аналитическим сигналом обладает рядом полезных практических свойств.

1.Произведение аналитического сигнала $x_a(t)$ на сопряженный ему сигнал равно квадрату огибающей исходного сигнала $x(t)$

$$\begin{aligned} x_a(t)x_a^*(t) &= (x(t) + jx_1(t))(x(t) - jx_1(t)) = \\ &= x^2(t) + x_1^2(t) = A^2(t) \end{aligned}$$

2.Спектральная функция аналитического сигнала имеет вид

$$S_a(\omega) = \begin{cases} 2S(\omega), & \text{при } \omega > 0 \\ 0, & \text{при } \omega < 0 \end{cases}$$

Этот интересный результат совсем нетрудно доказать. Если аналитический сигнал описывается уравнением

$$x_a(t) = x(t) + jx_1(t),$$

то его спектр имеет вид

$$S_a(\omega) = S(\omega) + jS_1(\omega).$$

Однако выше было показано, что при $\omega > 0$ $S_1(\omega) = -jS(\omega)$, а при $\omega < 0$ $S_1(\omega) = S(\omega)$, что после подстановки и приводит к искомому выражению.

3.Спектральная функция комплексной огибающей $\dot{A}(t)$ совпадает со спектром аналитического сигнала, смещенного влево по оси частот на величину несущей частоты ω_0 .

$$S_a(\omega) = \int_{-\infty}^{\infty} \{\dot{A}(t)e^{j\omega_0 t}\} e^{-j\omega t} dt = \int_{-\infty}^{\infty} \dot{A}(t) e^{-j(\omega - \omega_0)t} dt = S_{\dot{A}}(\omega - \omega_0).$$

В общем случае спектр комплексной огибающей не является симметричным относительно нулевой частоты.

На рис.2.17 изображены графики спектральных функций исходного и аналитического сигнала, а также спектральная функция соответствующей комплексной огибающей.

Введя обозначение

$$\Omega = \omega - \omega_0,$$

можно записать

$$S_a(\omega) = 2S(\omega_0 + \Omega).$$

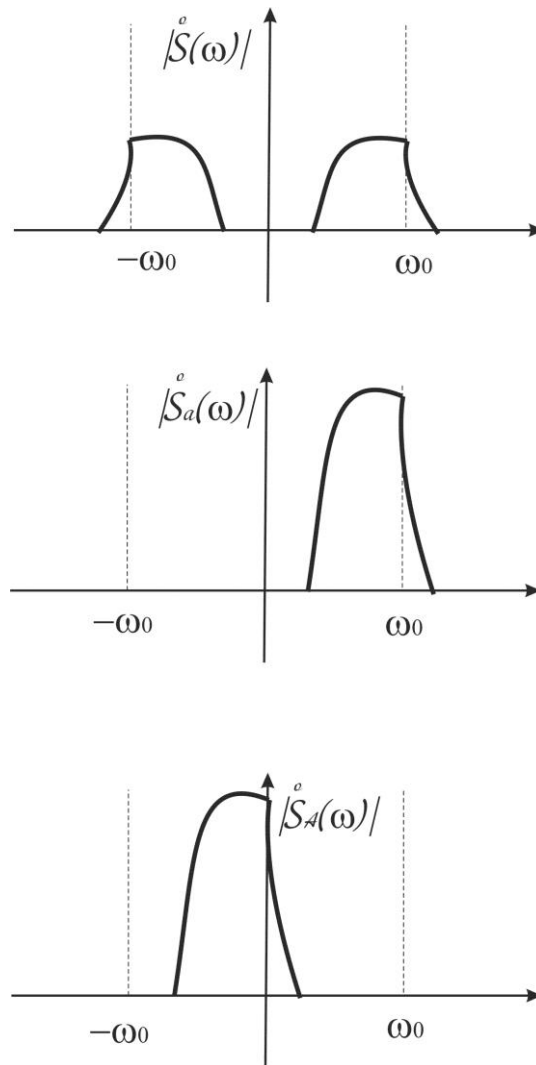


Рис.2.17. Спектр аналитического сигнала и комплексной огибающей

4. Корреляционная функция аналитического сигнала, определяемая выражением

$$B_a(\tau) = \int_{-\infty}^{\infty} x_a(t)x_a(t - \tau)dt,$$

связана с корреляционной функцией исходного сигнала $x(t)$ соотношением

$$B(\tau) = \frac{1}{2} \text{Re}\{B_a(\tau)\},$$

где $B(\tau) = \int_{-\infty}^{\infty} x(t)x(t - \tau)dt$.

Справедливость данного выражения рекомендуется в качестве упражнения проверить самостоятельно.

2.4. Внутриимпульсная частотная модуляция

Сигнал с внутриимпульсной частотной модуляцией представляет собой радиоимпульс, высокочастотное заполнение которого имеет переменную частоту.

Если закон изменения мгновенной частоты заполняющего сигнала имеет линейный характер, то такие сигналы называют ЛЧМ – сигналами (сигналами с линейной частотной модуляцией). Наиболее широкое применение они получили в радиолокации. Пример ЛЧМ – сигнала с огибающей прямоугольной формы приведен на рис. 2.18.

ЛЧМ – сигналы имеют одно важное свойство, состоящее в том, что 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

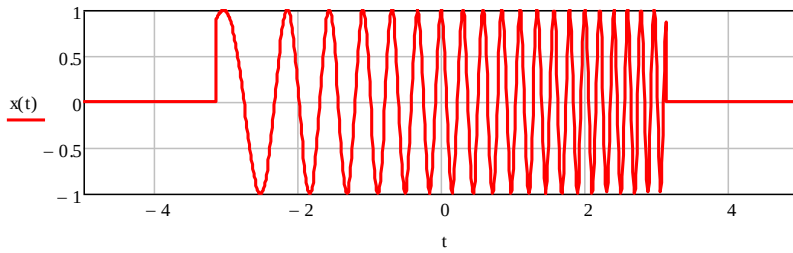


Рис.2.18. Сигнал с линейной частотной модуляцией

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.

Для модели радиоимпульса с прямоугольной огибающей прием его длительность равной $T_{и}$, и точку $t = 0$ поместим в центр радиоимпульса. Допустим также, что частота заполнения линейно нарастает от начала импульса к его концу со скоростью μ (c^{-2}), при этом:

$$\omega(t) = \omega_0 + \mu t \quad (2.33)$$

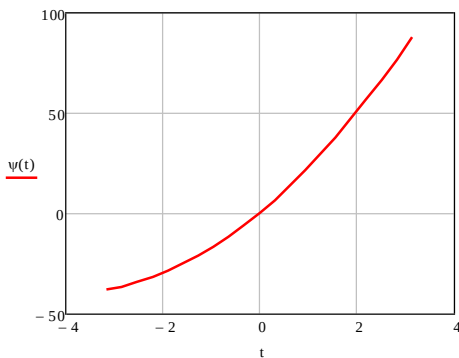


Рис.2.19. График изменения полной фазы

Тогда девиация частоты за время длительности импульса и полная фаза сигнала примут вид

$$\Delta\omega = \mu T_{и}$$

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

График изменения полной фазы представлен на рис.2.19

Итоговое уравнение ЛЧМ – сигнала с учетом полученных выражений можно представить в виде:

$$x(t) = \begin{cases} 0, & \text{при } t < -T_{и}/2 \\ A_0 \cos(\omega_0 t + \mu t^2 / 2), & \text{при } -T_{и}/2 \leq t \leq T_{и}/2 \\ 0, & \text{при } t > T_{и}/2 \end{cases}$$

Девиация частоты в подобных сигналах за время длительности импульса по сравнению с несущей частотой обычно мала и форма спектра зависит от так называемой базы импульса $\Delta\omega T_{и}$. Поскольку при выбранном типе модуляции

$$\Delta\omega T_{и} = \mu T_{и}^2,$$

то управление базой можно производить при сохранении длительности импульса $T_{и} = const$ и изменяя только девиацию частоты.

На практике значение базы сигналов обычно много больше 1. Увеличение базы сопровождается расширением полосы спектра, при этом в пределах этой полосы модуль спектральной плотности практически постоянен и равен $A_0 \sqrt{\frac{\pi}{2\mu}}$.

2.5. Общие сведения об импульсно модулированных сигналах

В импульсной модуляции в качестве носителя модулированных сигналов используются последовательности импульсов, которые, как правило, имеют прямоугольный вид. В радиотехнических системах (то есть при беспроводной передаче данных) эти последовательности заполняются высокочастотными колебаниями, что создает тем самым двойную модуляцию. В основном, такой вид модуляции применяется при передаче дискретизированных данных. Для прямоугольных импульсов наиболее широко используются амплитудно-импульсная (АИМ) и широтно-импульсная (ШИМ) модуляция.

Амплитудно-импульсная модуляция (АИМ) заключается в изменении приращения амплитуды импульсов пропорционально функции управляющего сигнала при постоянной длительности импульсов и периоде их следования.

$$x(t) = A_0 + ka(t) \quad (2.34)$$

Спектр АИМ рассмотрим на примере модулирования однотонального сигнала. Напишем уравнение модулированного сигнала в следующей форме:

$$x(t) = X_0(1 + M\cos\Omega t)f(t)\cos\omega_0 t$$

где $f(t)$ – периодическая последовательность прямоугольных импульсов с периодом следования $T_1 = \frac{2\pi}{\omega_1}$ и длительностью импульсов τ

$$f(t) = \sum_{k=-\infty}^{k=\infty} f(t - kT_1). \quad (2.35)$$

Для данной последовательности импульсов, как мы уже рассматривали ранее, существует представление в виде ряда Фурье. Воспользуемся стандартным представлением вещественной формы этого ряда

$$f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} A_n \cos(n\omega_1 t + \theta_n). \quad (2.36)$$

Подстановка данного выражения в исходное уравнение сигнала приводит к следующему

$$\begin{aligned} x(t) = X_0(1 + M\cos\Omega t) & \left\{ \frac{a_0}{2} \cos\omega_0 t + \right. \\ & + \frac{A_1}{2} \cos[(\omega_0 - \omega_1)t - \theta_1] + \frac{A_1}{2} \cos[(\omega_0 + \omega_1)t + \theta_1] + \\ & + \frac{A_2}{2} \cos[(\omega_0 - 2\omega_1)t - \theta_2] + \frac{A_2}{2} \cos[(\omega_0 + 2\omega_1)t + \theta_2] + \dots \left. \right\} \end{aligned}$$

Сигнал вида $f(t)\cos\omega_0 t$, входящий в полученное уравнение называют амплитудно-манипулированным сигналом.

При рассмотрении сигналов с тональной модуляцией нами было получено уравнение

$$x(t) = A_0(1 + M\cos(\Omega t + \theta_{\Omega 0})) \cos[\omega_0 t + \theta_0] = A_0 \cos[\omega_0 t + \theta_0] + \\ + \frac{MA_0}{2} \cos\{(\omega_0 + \Omega)t + \theta_0 + \theta_{\Omega 0}\} + \frac{MA_0}{2} \cos\{(\omega_0 - \Omega)t + \theta_0 - \theta_{\Omega 0}\},$$

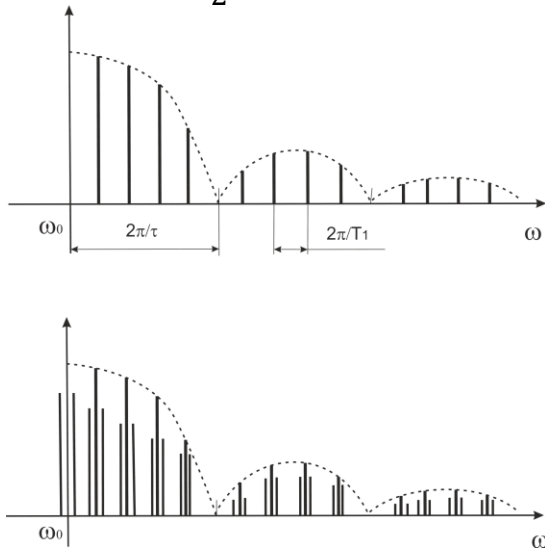


Рис.2.20 Спектры амлитудно-манипулированного и амплидудно-модулированного сигналов

которое теперь можно почленно применить к каждому слагаемому уравнения амплитудно-модулированного сигнала. Тогда становится очевидным, что в спектре амплитудно-модулированного сигнала по сравнению со спектром амплитудно-манипулированного сигнала дополнительно появится пара боковых частотных компонент, то есть на частоте ω_0 будет наблюдаться пара $\omega_0 \pm \Omega$, на частоте $\omega_0 + \omega_1$ будет наблюдаться пара $(\omega_0 + \omega_1) \pm \Omega$, на частоте $\omega_0 - \omega_1$ будет наблюдаться пара $(\omega_0 - \omega_1) \pm \Omega$ и так далее.

Графики спектров амплитудно-манипулированного и соответствующего ему амплитудно-модулированного сигналов приведены на рис. 2. 20 .

Широтно-импульсная модуляция (ШИМ, в английской терминологии pulse width modulation, PWM), которую иногда называют *модуляцией по длительности импульсов (ДИМ)*, заключается в управлении длительностью импульсов пропорционально функции управляющего сигнала при постоянной амплитуде импульсов и периоде следования по фронту импульсов:

$$\tau(t) = \tau_0 + \mu a(t), \quad T_n = \text{const}, A_0 = \text{const}. \quad (2.37)$$

Проиллюстрируем ШИМ на простейшем на примере, когда модулирующая функция изменяется по гармоническому закону. Соответствующие графики формирования сигнала приведены на рис.2.21 .

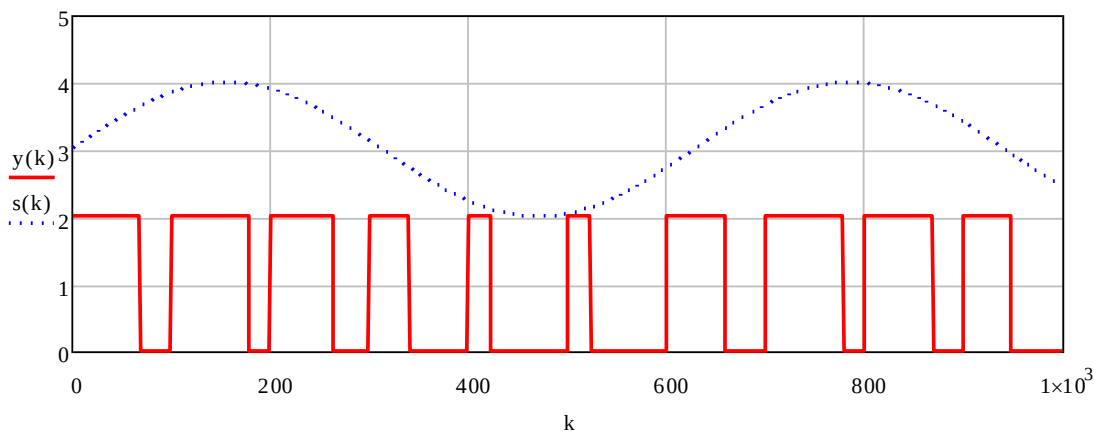


Рис.2.21 . Формирование ШИМ-сигнала

Передаваемый сигнал подвергается дискретизации, причем имеет значение, как интервал дискретизации, так и количество уровней квантования. При передаче данных прямоугольные импульсы начинаются в моменты дискретных отсчетов данных, а

длительность импульсов устанавливается пропорциональной значению отсчетов, при этом максимальная длительность импульсов не должна превышать интервала дискретизации данных. Естественно, что при малом числе уровней квантования погрешность восстановления исходного гармонического сигнала очень велика.

Заметим также, что широтно-импульсная модуляция с последующим выделением постоянной составляющей может эффективно использоваться для слежения за средним уровнем сигнала и автоматического регулирования его динамического диапазона, как, например, в системах установки громкости звука и яркости цветов и изображения в целом в современных телевизионных установках.

Временная импульсная модуляция (ВИМ) представляет собой девиацию импульсов по временной оси по закону модулирующего сигнала, и по существу аналогична угловой модуляции гармонической несущей и может быть реализована в виде фазовой (ФИМ) или частотной (ЧИМ) модуляции.

Суть кодово-импульсной модуляции заключается в том, что в точках дискретизации модулирующего сигнала производится квантование его значений и их кодирование (чаще всего в двоичной системе счисления). Кодированные значения затем передаются при помощи соответствующей кодовой последовательности стандартных символов.

2.6. Вопросы для самоконтроля и повторения

1. Запишите уравнение модулированного колебания.
2. Какие виды модуляции вам известны?
3. Запишите уравнение амплитудно-модулированного сигнала.
4. Что такое глубина амплитудной модуляции?
5. В чем состоит и когда возникает явление перемодуляции?
6. Что такое тональная амплитудная модуляция?
7. Изобразите спектр соответствующий случаю тональной амплитудной модуляции.
8. Как определяется полезная мощность модулированного сигнала.
9. Что такое балансовая амплитудная модуляция?
10. Что такое сигналы с одной боковой полосой?
11. Какие виды угловой модуляции вам известны?
12. Запишите уравнения соответствующие фазовой и частотной модуляции.
13. Что такое индекс гармонической угловой модуляции?
14. Изобразите амплитудный спектр сигнала с гармонической амплитудной модуляцией.
15. Поясните понятие «узкополосный сигнал».
16. Для каких целей используется преобразование Гильберта?
17. Перечислите основные свойства преобразования Гильберта.
18. Изобразите график сигнала с линейной частотной модуляцией.
19. Изобразите график сигнала с широтно—импульсной модуляцией.
20. Изобразите график сигнала с амплитудно-импульсной модуляцией.
21. Изобразите график сигнала с частотно-импульсной модуляцией.

ГЛАВА 3. Дискретные сигналы

3.1. Дискретизация непрерывных сигналов

Развитие компьютерной и микропроцессорной техники, сопровождающееся относительным удешевлением и миниатюризацией вычислительных устройств, привело к широкому распространению систем реального времени, использующих алгоритмы цифровой обработки сигналов.

Исходные физические сигналы являются непрерывными функциями, определенными на некотором континуальном интервале времени, а цифровые устройства оперируют дискретными последовательностями, представляющими собой, как правило, периодические отсчеты, получаемые из исходного физического сигнала.

Кроме того исходные физические сигналы, как правило, имеют непрерывную (или кусочно-непрерывную) область своих возможных значений. Цифровые устройства, исходя из принципа своего действия, используют квантованные по уровню значения сигналов. При этом допустимое количество уровней квантования определяется разрядной размерностью используемого системой слова (чаще всего 8 бит, 16 бит, 32 бита и т.д.).

Таким образом, процесс преобразования исходного физического сигнала в его цифровой аналог всегда сопровождается потерей информации. Это происходит, во-первых, из-за **дискретизации по времени** и, во-вторых, из-за **квантования по уровню** исходного сигнала.

Остановимся несколько подробнее на рассмотрении этих двух процессов с целью определить условия, при которых указанные искажения можно считать незначительными.

Вначале рассмотрим процесс квантования во времени для случая амплитудной модуляции импульсов, который является наиболее распространенным на практике. При данном способе модуляции амплитуда формируемых импульсов пропорциональна амплитуде исходного непрерывного сигнала, длительность импульсов постоянна, интервалы между импульсами одинаковы и равны такту квантования. Упрощенная схема, соответствующая такому способу временного квантования, представлена на рис.3.1.

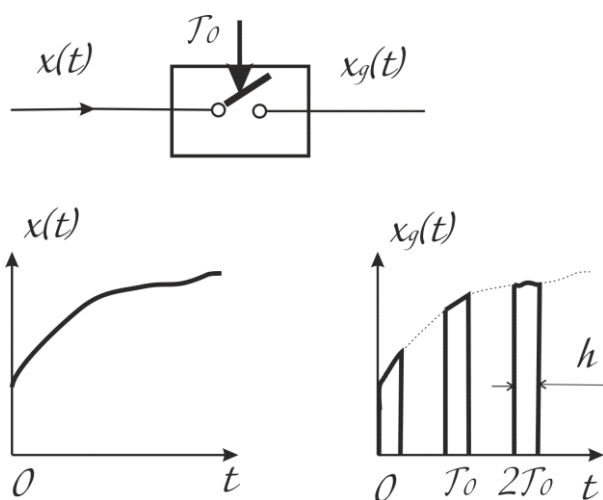


Рис.3.1. Схема дискретизации непрерывного сигнала

Рис.3.1. иллюстрирует принцип получения амплитудно-модулированных импульсов $x_g(t)$, основанный на пропускании исходного непрерывного сигнала $x(t)$ через ключ К, который с тактом T_0 периодически замыкается на время h .

Эффективный математический метод описания дискретных функций основывается на замене последовательности импульсов $x_g(t)$ **решетчатой функцией**, состоящей из идеальных импульсов.

Идеальный импульс (см.(1.40)), называемый также **единичным**

импульсом или функцией Дирака, определяется выражением

$$\delta(t) = \begin{cases} 0, & t \neq 0 \\ \infty, & t = 0, \end{cases}$$

а его площадь равна единице, то есть

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

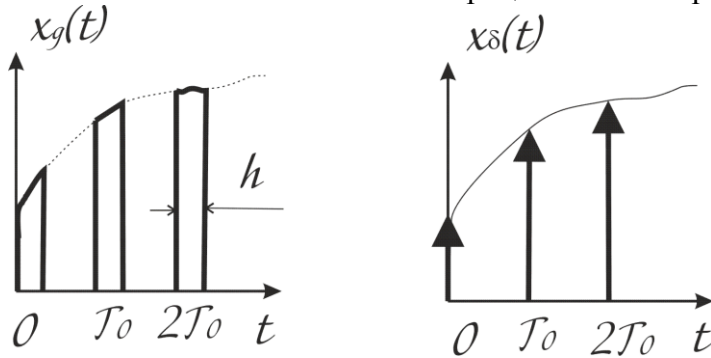
Если продолжительность замыкания ключа K значительно меньше такта квантования

$$h \ll T_0,$$

то импульсы последовательности $x_g(t)$, имеющие площадь $x_g(t)h$, можно приближенно заменить идеальными импульсами с весовым коэффициентом, равным той же площади

$$x_g(t) \approx x_\delta(t) = x(t)h \sum_{k=0}^{\infty} \delta(t - kT_0). \quad (3.1)$$

Схематично описанный процесс замены представлен на рис.3.2.



Решетчатая функция $x_\delta(t)$, изображенная на рис.9.2. физически нереализуема и служит лишь удобной формой представления последовательности реальных импульсов $x_g(t)$. Введение идеального квантователя, формирующего δ -импульсы, позволяет значительно упростить математический аппарат, используемый для описания

Рис.3.2. Переход к решетчатой функции

процессов прохождения дискретных сигналов через линейные динамические системы (при этом не явно предполагается, что за ключом K находится линейное звено с постоянной времени $\tau \gg h$).

Учитывая, что последовательность идеальных импульсов определена лишь в моменты времени, соответствующие замыканию ключа $t = kT_0$, где $k = 0, 1, 2, \dots$ уравнение для решетчатой функции можно переписать в виде

$$x_\delta(t) = h \sum_{k=0}^{\infty} x(kT_0) \delta(t - kT_0). \quad (3.2)$$

При анализе динамических систем с одинаковыми и синхронно работающими ключами на входе и выходе, продолжительность интервала замыкания h можно не учитывать. Если за ключом стоит экстраполятор (о нем речь будет ниже), то эта величина фактически не влияет на конечный результат. По этой причине величину h можно выбрать произвольно. В этом случае последовательность идеальных импульсов описывается выражением

$$x^*(t) = \sum_{k=0}^{\infty} x(kT_0)\delta(t - kT_0). \quad (3.3)$$

Вспомним теперь, как выглядит преобразование Фурье сигнала $x^*(t)$. Применение преобразования Фурье к единичному идеальному импульсу дает

$$F\{\delta(t)\} = \int_0^{\infty} \delta(t)e^{-j\omega t} dt = 1,$$

применение этого же преобразования к импульсу, сдвинутому во времени на kT_0 , очевидно приводит к выражению

$$F\{\delta(t - kT_0)\} = e^{-j\omega kT_0}.$$

Тогда с учетом ранее полученных уравнений легко записать

$$F\{x^*(t)\} = X^*(j\omega) = \sum_{k=0}^{\infty} x(kT_0)e^{-j\omega kT_0}$$

Из последнего уравнения видно, что преобразование Фурье дискретной функции является периодической функцией с частотой повторения

$$\omega_0 = 2\pi/T_0$$

так как

$$X^*(j\omega) = X^*(j\omega + jn\omega_0)$$

где $n = 0, \pm 1, \pm 2, \dots$,

j - мнимая единица.

Если непрерывный сигнал $x(t)$ имеет ограниченную полосу частот, то есть его преобразование Фурье удовлетворяет условиям

$$X(j\omega) \neq 0 \text{ при } \omega_{min} \leq \omega \leq \omega_{max}$$

$$X(j\omega) = 0 \text{ при } |\omega| > \omega_{max}$$

Предположим, что сигнал, удовлетворяющий приведенным условиям дискретизируется во времени с малым тактом T_0 , после чего аппроксимируется последовательностью $x^*(t)$. Если величина T_0 достаточно мала, то в соответствии с вышеизложенным преобразование Фурье для этого сигнала распадается на "основной спектр"

$$T_0 X^*(j\omega) \approx X(j\omega) \text{ для } |\omega| < \omega_{max}$$

и совокупность повторяющихся (по частоте) с периодом ω_0 "дополнительных спектров"

$$T_0 X^*(j\omega + jn\omega_0) \approx X(j\omega), \text{ для } n = 0, \pm 1, \pm 2, \dots, \quad (3.4)$$

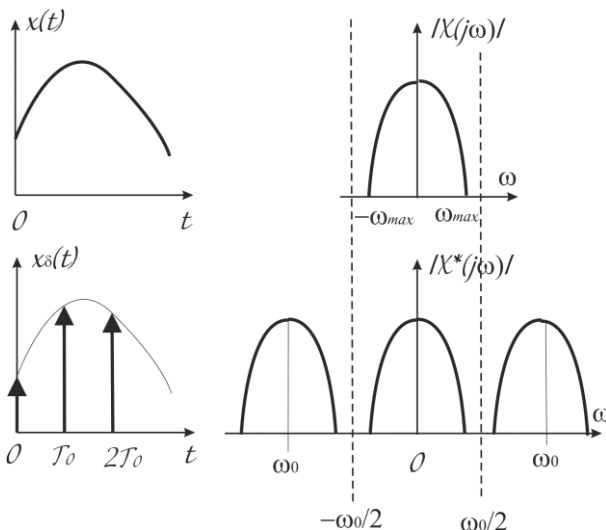


Рис.3.3. Спектры непрерывного и дискретного сигналов

Результат этого преобразования представлен на рис.3.3.

Таким образом, в спектре квантованного по времени сигнала по сравнению со спектром непрерывного сигнала появляются **дополнительные** высокочастотные составляющие.

Допустим, что исходный сигнал после квантования восстанавливается с помощью идеального полосового фильтра, имеющего характеристику, изображенную на рис.3.4.

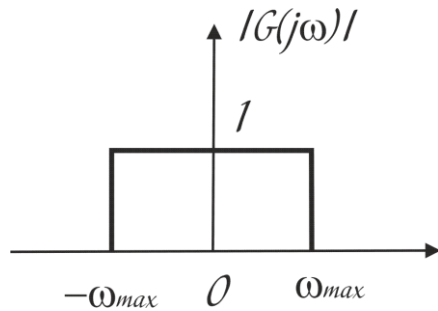


Рис.3.4. Характеристика идеального полосового фильтра

Понятно, что подобная операция восстановления может быть выполнена без ошибок только при условии

$$\omega_0/2 \geq \omega_{max}.$$

Если же частота квантования недостаточно велика, то на основной спектр накладываются дополнительные спектры и точное выделение исходного спектра

становится невозможным.

Из рассмотренного ясно, что для того чтобы непрерывный сигнал со спектром, ограниченным частотой ω_{max} , можно было однозначно представить последовательностью его дискретных значений, необходимо, чтобы частота квантования ω_0 удовлетворяла условию

$$\omega_0 \geq 2\omega_{max} \quad (3.5)$$

При этом непрерывный сигнал $x(t)$ может быть восстановлен из дискретного по интерполяционной формуле

$$x(t) = \sum_{k=-\infty}^{\infty} x(kT_0) \frac{\sin \omega_0(t - kT_0)/2}{\omega_0(t - kT_0)/2} \quad (3.6)$$

Последние два соотношения (3.5) и (3.6) составляют основное содержание **теоремы Котельникова-Шеннона**, которую часто называют **импульсной теоремой**.

В случае периодического квантования сигнала с ограниченным спектром восстановление сигнала осуществляется по формуле Котельникова-Шеннона и называется восстановлением Котельникова, однако это восстановление имеет два существенных недостатка:

- значение функции $x(t)$ в момент времени t выражается не только через предшествующие, но и через последующие значения дискретных отсчетов $x(kT_0)$;
- вычислительная сложность алгоритма восстановления.

Поэтому на практике используются более простые способы восстановления сигнала, основанные на применении экстраполяторов.

Наиболее часто в реальных системах используются фиксаторы (экстраполяторы нулевого порядка), которые на период, равный такту квантования, фиксируют мгновенное значение дискретного сигнала $x(kT_0)$, формируя на своем выходе соответствующий ступенчатый сигнал. Процесс работы квантователя с экстраполятором нулевого порядка представлен на рис.3.5.

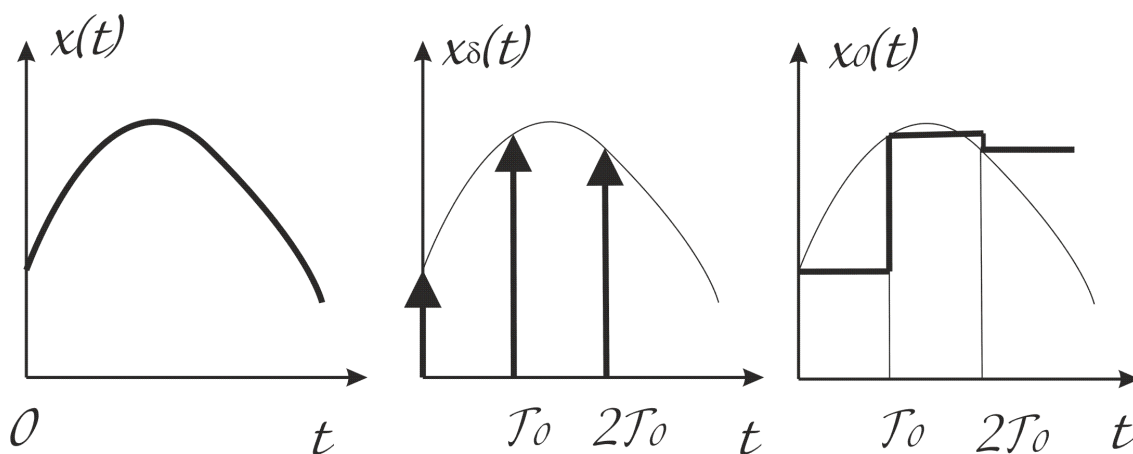


Рис.3.5. Восстановление сигнала экстраполятором нулевого порядка

Понятно, что в большинстве случаев восстановление с помощью экстраполятора дает ошибку. При этом наибольшее значение ошибки для случая квантования сигнала с гладкой первой производной может быть определено по формуле

$$e_{0_{max}} = \max_k |x(k+1)T_0 - x(kT_0)| \leq T_0 \max_t |x'(t)|$$

Для гладких функций $x(t)$ уменьшить ошибку восстановления во многих случаях позволяет применение экстраполяторов старших порядков. Так, например, экстраполятор первого порядка определяется выражением

$$x(t) = x(kT_0) + \frac{t - kT_0}{kT_0 - (k-1)T_0} [x(kT_0) - x((k-1)T_0)]$$

где $kT_0 \leq t \leq (k+1)kT_0$.

В этом случае максимальная ошибка может быть определена выражением

$$e_{1_{max}} \leq T_0^2 \max_t |x''(t)|$$

Для формирования представления о величине ошибки, возникающей при квантовании сигнала, рассмотрим случай синусоидального сигнала $x(t)$. Величины относительных ошибок получаемых при восстановлении сигнала приведены в табл.9.1.

Таблица 9.1.

Относительные ошибки, возникающие при восстановлении синусоидального сигнала

Число отсчетов за период	Максимальная относительная ошибка	
	Экстраполятор нулевого порядка	Экстраполятор первого порядка
2	1.5	2.5
10	0.3	0.19
100	0.03	0.002
500	0.006	0.00008

Пример.3.1. (Гармонический сигнал при дискретных наблюдениях)

Рассмотрим, как будет выглядеть модель обычного гармонического колебания

$$x(t) = A_0 \sin(\omega_0 t + \theta_0)$$

при использовании для его описания периодических отсчетов, производимых с периодом T_0 .

Для последовательности отсчетов $k = 0, 1, 2, \dots, n, \dots$ запишем

$$x(kT_0) = A_0 \sin(\omega_0 kT_0 + \theta_0).$$

Правую часть данного уравнения запишем иначе

$$\begin{aligned} A_0 \sin(\omega_0 kT_0 + \theta_0) &= A_0 \sin(\omega_0(k-1)T_0 + \omega_0 T_0 + \theta_0) = \\ &= A_0 \cos \omega_0 T_0 \sin(\omega_0(k-1)T_0 + \theta_0) + A_0 \sin \omega_0 T_0 \cos(\sin(\omega_0(k-1)T_0 + \omega_0 T_0)). \end{aligned}$$

Аналогичным образом получим

$$\begin{aligned} A_0 \sin(\omega_0(k-2)T_0 + \theta_0) &= \\ &= A_0 \cos \omega_0 T_0 \sin(\omega_0(k-1)T_0 + \theta_0) + A_0 \sin \omega_0 T_0 \cos(\sin(\omega_0(k-1)T_0 + \omega_0 T_0)). \end{aligned}$$

Объединяя два последних уравнения, получаем

$$A_0 \sin(\omega_0 kT_0 + \theta_0) = 2A_0 \cos \omega_0 T_0 \sin(\omega_0(k-1)T_0 + \theta_0) - A_0 \sin(\omega_0(k-2)T_0 + \theta_0).$$

В дискретном времени это уравнение принимает вид

$$x(k) = 2\cos \omega_0 T_0 x(k-1) - x(k-2) \quad (3.7a)$$

или

$$x(k) = a_1 x(k-1) - x(k-2), \quad (3.7b)$$

где $a_1 = 2\cos \omega_0 T_0 = \text{const}$.

Таким образом, нам удалось установить, что при переходе к дискретному времени гармонический сигнал может быть описан разностным уравнением второго порядка с одним коэффициентом, величина которого зависит от частоты колебаний сигнала и периода времени его дискретизации.

Обратим внимание на следующие особенности модели гармонического сигнала в дискретном времени:

1. модель одного и того же сигнала при различных значениях периода дискретизации имеет различное значение коэффициента a_1 ,
2. гармонические сигналы с различными частотами колебаний могут описываться одинаковыми дискретными моделями при соответствующем выборе периодов дискретизации.

Понятно, что для полигармонического сигнала порядок соответствующего разностного уравнения будет равняться числу гармоник, входящих в модель сигнала, умноженному на два.

3.2. Дискретное преобразование Фурье и его свойства

В предыдущем разделе мы рассмотрели процесс дискретизации во времени с периодом T_0 исходного сигнала $x(t)$ и переход к его описанию идеализированной последовательностью отсчетов

$$x(t) = \sum_{k=-\infty}^{\infty} x(kT_0)\delta(t - kT_0).$$

Ранее мы также выяснили, что спектр этого сигнала является периодическим с периодом $\frac{2\pi}{T_0}$.

Допустим, что исследуемая функция $x(t)$ является периодической. Тогда и соответствующая ей функция дискретного времени $x(k)$ также будет периодической с некоторым периодом дискретных отсчетов N

$$x(k) = x(k + N). \quad (3.8)$$

Предполагается, что N – наименьшее из чисел, для которых выполняется указанное условие.

В соответствии с принятым допущением о периодичности исследуемого сигнала его спектр должен быть дискретным с расстоянием между гармониками $\frac{2\pi}{NT_0}$, что позволяет воспользоваться общей формулой вычисления коэффициентов ряда Фурье.

$$\begin{aligned} X^\circ(n) &= \frac{1}{NT_0} \int_0^{NT_0} x(t) e^{-j\omega_0 t} dt = \frac{1}{NT_0} \int_0^{NT_0} \sum_{k=0}^{N-1} x(kT_0)\delta(t - kT_0) e^{-j\omega_0 t} dt = \\ &= \frac{1}{NT_0} \sum_{k=0}^{N-1} x(k) \int_0^{NT_0} \delta(t - kT_0) e^{-j\omega_0 t} dt = \frac{1}{NT_0} \sum_{k=0}^{N-1} x(k) e^{-j\omega_0 kT_0} = \\ &= \frac{1}{NT_0} \sum_{k=0}^{N-1} x(k) \exp\left(-j \frac{2\pi nk}{N}\right). \end{aligned}$$

В результате полученная формула показывает, что вычисление искомым комплексных амплитуд разложения исходного сигнала в ряд Фурье можно произвести путем линейной комбинации последовательности его дискретных отсчетов.

Поскольку при рассмотрении дискретных последовательностей величина T_0 не является существенной и используется только для перехода от последовательности отсчетов к реальному масштабу времени и частоты (см. пример предыдущего раздела), то его часто из окончательной формулы удаляют. Таким же образом поступают с множителем N , перенося его в формулу обратного преобразования.

В итоге приходим к следующим формулам прямого дискретного преобразования Фурье

$$X^\circ(n) = \sum_{k=0}^{N-1} x(k) \exp\left(-j \frac{2\pi nk}{N}\right) \quad (3.9)$$

и формулу обратного дискретного преобразования Фурье

$$x(k) = \frac{1}{N} \sum_{n=0}^{N-1} X^{\circ}(n) \exp\left(-j \frac{2\pi nk}{N}\right). \quad (3.10)$$

Исходя из аналогии непрерывного и дискретного вариантов преобразования Фурье, логично ожидать, что они должны обладать определенным подобием свойств. Рассмотрим некоторые свойства дискретного преобразования Фурье.

1. Линейность

$$\mathbb{F}(\alpha x(k) + \beta y(k)) = \alpha \mathbb{F}(x(k)) + \beta \mathbb{F}(y(k)),$$

где $\mathbb{F}(\cdot)$ – обозначение дискретного преобразования Фурье.

2. Задержка исходной функции во времени

Если исходную последовательность $x(k)$ задержать во времени на интервал, равный одному периоду дискретизации то для преобразования Фурье от новой последовательности будет справедливо следующее соотношение

$$\mathbb{F}(x(k-1)) = \mathbb{F}(x(k))e^{-\frac{j2\pi n}{N}} = X^{\circ}(n)e^{-\frac{j2\pi n}{N}},$$

где $X^{\circ}(n) = \mathbb{F}(x(k))$ – дискретное преобразование Фурье исходной последовательности.

3. Симметрия преобразования для вещественного сигнала.

Данное свойство полностью аналогично такому же свойству непрерывного преобразования Фурье

$$X^{\circ}(N-n) = X^{\circ}(-n) = X^{\circ*}(n).$$

4. Преобразование Фурье для произведения двух последовательностей.

Для двух последовательностей одинаковой длины N $x(k)$ и $y(k)$ можно произвести операцию поэлементного умножения $x(k)y(k)$.

Применение операции прямого дискретного преобразования Фурье к последовательности, полученной таким умножением, дает следующее

$$\mathbb{F}(x(k)y(k)) = \frac{1}{N} \sum_{m=0}^{N-1} X^{\circ}(m)Y^{\circ}(n-m),$$

где значения $Y^{\circ}(m)$, выходящие за диапазон $0 \leq m \leq N-1$, определяются исходя из свойства периодичности спектра $Y^{\circ}(m) = Y^{\circ}(m \pm N)$.

Если в последней формуле положить $m = 0$ и $x(k) = y(k)$, то легко получить дискретный аналог равенства Парсеваля

$$\sum_{k=0}^{N-1} x^2(k) = \frac{1}{N} \sum_{m=0}^{N-1} |X^{\circ}(m)|^2$$

5. Значения дискретного преобразования Фурье при $n = 0$.

Случаю $n = 0$ соответствует гармоника нулевого порядка, которая по своей сути является постоянной составляющей.

В соответствии с введенными уравнениями для прямого дискретного преобразования Фурье в данном случае можно записать

$$X^{\circ}(0) = \sum_{k=0}^{N-1} x(k).$$

В пояснениях, данных при введении уравнений прямого и обратного дискретных преобразований Фурье, отмечалось, что иногда множитель $1/N$ используется именно в формуле прямого преобразования, в том числе и для того, чтобы

сохранить обычный смысл среднего значения, как суммы слагаемых, деленной на их количество.

Дискретное преобразование Фурье может быть также использовано для восстановления непрерывного периодического сигнала с ограниченной частотной полосой. Для этого в формуле обратного дискретного преобразования Фурье заменяют номер отсчета непрерывной переменной, представляющей собой нормированное на период дискретизации реальное время, то есть t/T_0

$$x(t) = \frac{1}{N} \sum_{n=-N/2}^{\frac{N}{2}-1} X^{\circ}(n) e^{j \frac{2\pi t}{NT_0}} \quad (3.11)$$

при четном значении N и

$$x(t) = \frac{1}{N} \sum_{n=-\frac{N-1}{2}}^{\frac{N-1}{2}} X^{\circ}(n) e^{j \frac{2\pi t}{NT_0}} \quad (3.12)$$

при нечетных значениях N .

3.3. Растекание спектра и эффект Гиббса

При исследовании реальных сигналов с помощью преобразований Фурье мы вынужденно сталкиваемся с двумя видами ограничений. Во-первых, применение выше дискретного преобразования Фурье предполагает конечность ряда выборочных наблюдений из реального сигнала, а во-вторых, в реальности нам всегда приходится использовать ограниченное число членов ряда Фурье даже для непрерывных сигналов.

Явлением Гиббса называется особенность поведения частичных сумм ряда Фурье в окрестностях точки разрыва исследуемой функции. Эта особенность была впервые обнаружена Уилбрейамом в 1848 году, а позже переоткрыта Гиббсом в 1898 году.

Суть проблемы состоит в том, что большинство методов анализа и обработки сигналов представляют собой или имеют в своем составе операцию свертки сигналов и различных функций. При этом в аналитических уравнениях, как сам сигнал, так и оператор свертки, выполняющий определенную задачу обработки данных и реализующий определенную частотную функцию системы обработки, могут быть бесконечно большими. Реальная обработка сигналов или численное моделирование на ЭВМ может иметь дело только с ограниченными выборками данных и коэффициентов разложения.

При стремлении представить эту функцию ограниченным спектром (или частичной суммой ряда Фурье) возникают два неприятных эффекта:

- крутизна перепада в аппроксимирующей функции "размывается", так как она не может быть больше чем крутизна (в нулевой точке), которую имеет наибольшая используемая гармоника;

- по обе стороны от "размытых" перепадов появляются выбросы и затухающие осцилляции с частотой, близкой к частоте первого отброшенного члена ряда.

Указанные выбросы могут быть весьма значительными и достигают 9% на основной частоте и 18% на удвоенной частоте основной гармоники.

В качестве примера рассмотрим функцию единичного скачка в окрестности нулевой точки

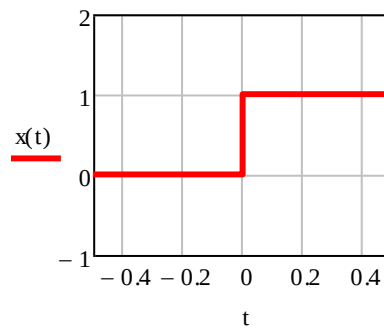


Рис.3.6. Функция единичного скачка

Ранее мы уже рассматривали этот сигнал и получали для него аналитическое выражение ряда Фурье в интервале $-T/2 < t < T/2$, которое с учетом сдвига по оси ординат и единичного значения амплитуды в данном случае примет вид

$$x(t) = \frac{1}{2} + \frac{2}{\pi} \left(\sin(\omega_0 t) + \frac{1}{3} \sin(3\omega_0 t) + \frac{1}{5} \sin(5\omega_0 t) + \dots \right)$$

Как видно из приведенного уравнения затухание коэффициентов ряда происходит довольно медленно и соответственно при восстановлении исходной функции усеченным рядом в точке скачка будут наблюдаться значительные осцилляции.

Поскольку сокращение членов ряда Фурье фактически означает ограничение частотного диапазона (полосы) в представлении сигнала, то для оценки влияния этой процедуры можно воспользоваться следующим приближением единичного скачка

$$x(t) \approx \frac{1}{2} + \frac{1}{\pi} Si(\beta, t),$$

где $Si(t) = \int_0^t \text{sinc}(z) dz$ – функция интегрального синуса

γ – граница ограничения полосы частот в преобразовании Фурье ($|\omega| < \beta$).

Пример, поясняющий возникающий эффект Гиббса приведен на рис.3.7.

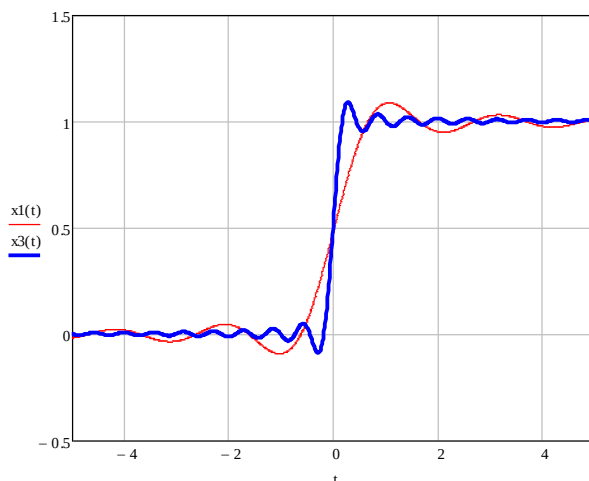


Рис.3.7. Эффект Гиббса

На графике представлены аппроксимации при значениях $\beta=3$ и $\beta=11$. При этом хорошо видно, эффект сохраняется даже при относительно большом увеличении спектральной полосы используемого разложения исходного сигнала.

Эффект Гиббса наблюдается во всех точках скачкообразного изменения сигналов.

Второй важнейший вопрос, который является основным для всех классических методов спектрального оценивания, связан с применением функции окна. Обработка сигналов

с помощью окон используется для управления эффектами, обусловленными наличием боковых лепестков в спектральных оценках. Причем, окна могут применяться как во временной области, то есть к наблюдаемым выборочным значениям самого исследуемого

сигнала, так и в частотной области, то есть к получаемым значениям спектральных составляющих.

Имеющуюся конечную последовательность отсчетов исходного сигнала удобно рассматривать как некоторую часть соответствующей бесконечной последовательности, так, например, последовательность из N наблюдаемых данных $x_0(k), k = 0, 1, 2, \dots, N - 1$ можно записать как произведение прямоугольной функции единичной амплитуды

$$\text{rect}(k) = \begin{cases} 1 & \text{при } 0 \leq k \leq N - 1 \\ 0 & \text{в остальных случаях} \end{cases}$$

и некой бесконечной последовательности $x(k)$ совпадающей с наблюдаемой последовательностью на участке наблюдения

$$x_{0(k)} = x(k)\text{rect}(k).$$

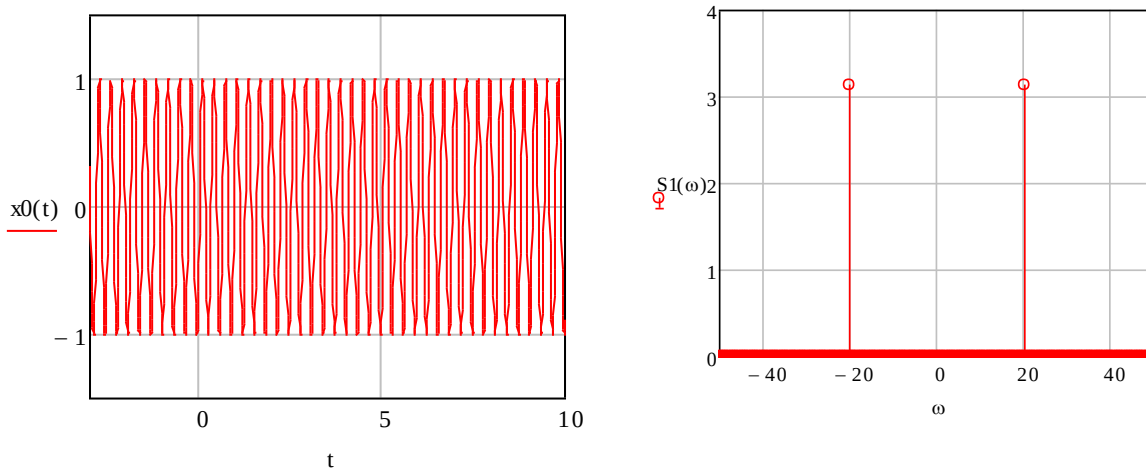
В общем случае значения последовательности $x(k)$ вне интервала наблюдения могут быть произвольными, но принимается допущение, что они равны нулю независимо от того так ли это на самом деле или нет.

Ранее мы уже рассматривали влияние окна на спектральную характеристику сигнала, хотя и не применяли такого наименования.

Вернемся еще раз вопросу вычисления спектра прямоугольного радиоимпульса. Но теперь рассмотрим его несколько иначе, а именно как наложение прямоугольного окна на бесконечный гармонический сигнал. Графики временных зависимостей исходного сигнала, окна и результата ограничения бесконечного сигнала конечным прямоугольным окном приведены на рис.3.8.

Рассмотренный пример наглядно демонстрирует каким образом спектр исходного гармонического сигнала расширился ("размылся") из-за образования многочисленных боковых лепестков, порожденных применением прямоугольного окна.

Для уменьшения данного эффекта, а также сокращения эффекта Гиббса разработано большое количество различных видов окон и способов коррекции. Эти



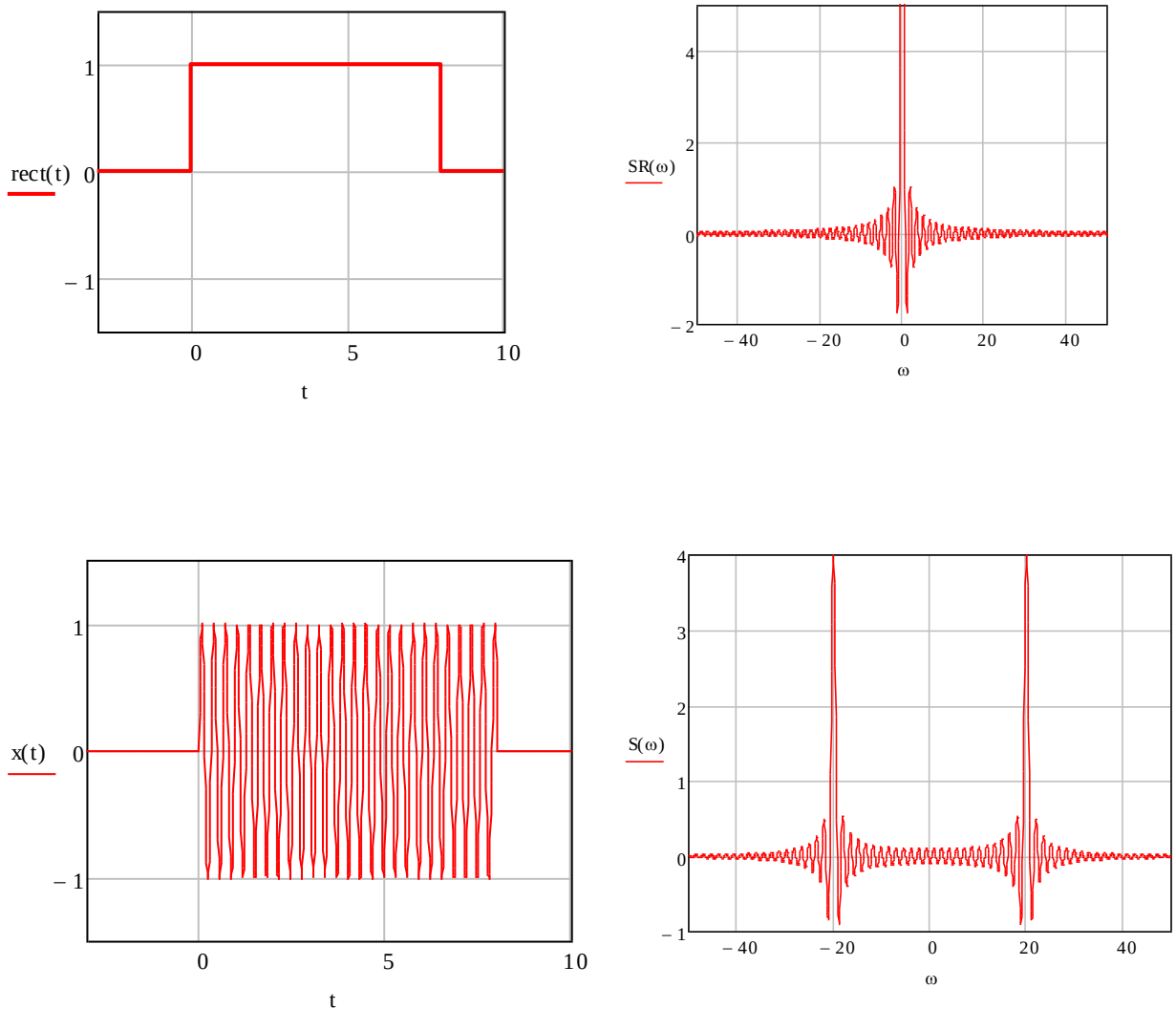


Рис.3.8. Размытие спектра гармонического сигнала при конечном интервале наблюдения

вопросы подробно рассматриваются в курсах по цифровой обработке сигналов.

3.4. Быстрое преобразование Фурье, дополнение нулями и частотное разрешение

Быстрое преобразование Фурье (БПФ) — это не отдельная разновидность преобразования Фурье, а название целого ряда эффективных вычислительных алгоритмов, предназначенных для быстрого вычисления дискретного преобразования Фурье.

Основная идея БПФ состоит в делении исходного N -точечного дискретного преобразования Фурье на два и более меньших (по длине) преобразований, каждое из которых можно вычислить отдельно, а затем линейно просуммировать с остальными, что в итоге дает преобразование исходной последовательности.

В задачи данного курса не входит подробное изучение алгоритмов формирования БПФ, поэтому мы остановимся лишь на описании общей идеи этого преобразования. В соответствии с рассмотренным выше дискретное преобразование Фурье было определено соотношением

$$X^o(n) = \sum_{k=0}^{N-1} x(k) \exp\left(-j \frac{2\pi nk}{N}\right). \quad (3.13)$$

Для вычисления одного спектрального отсчета требуется N операций комплексного умножения и сложения, поэтому общая вычислительная сложность данного преобразования оценивается в N^2 комплексных умножений и сложений. Если заменить одно преобразование на N точек двумя преобразованиями на $N/2$ точек каждое, то это приведет к сокращению числа необходимых операций в 2 раза. Схематично данная процедура разбиения представлена на рис.3.9.

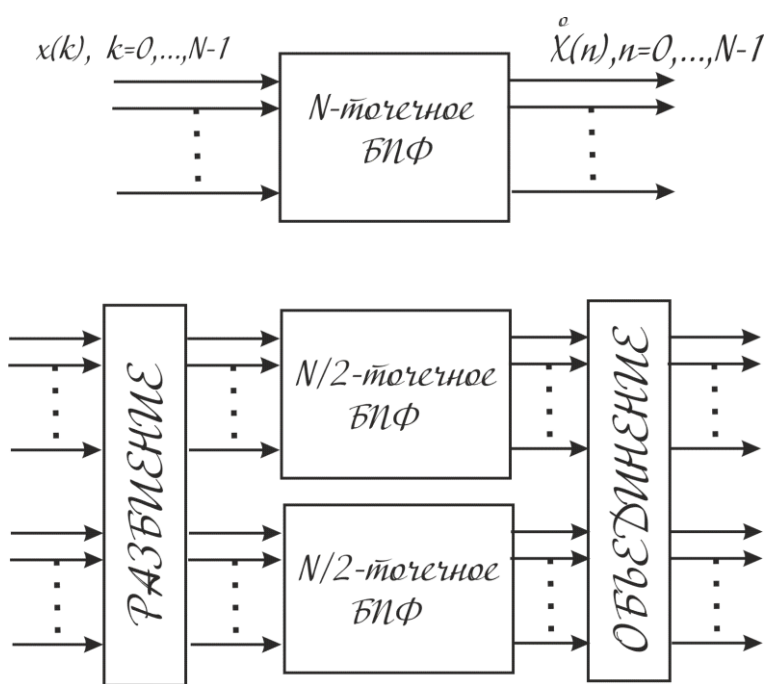


Рис.3.9. Замена N -точечного БПФ двумя $N/2$ -точечными БПФ

Процедура разбиения может быть продолжена далее и каждое из $N/2$ -точечных преобразований может быть разделено на пару $N/4$ -точечных БПФ и т.д. Если $N=2^n$, то такое разбиение может быть проведено n раз.

Алгоритмы БПФ, которые используют выборки длиной $N=2^n$, называются «алгоритмами БПФ по основанию 2». Данные алгоритмы получили наибольшее распространение, из-за удобства применения в машинной арифметике. Очевидно что делить последовательности на две можно по-разному, однако от этого зависит сможем ли мы при объединении получить неискаженный спектр сигнала

и чего с точки зрения вычислительных затрат это будет нам стоить. Эффективность алгоритма БПФ полностью зависит от способа разбиения и объединения последовательности, поскольку если не учитывать операции на разбиение-объединение, то для расчета спектра требуется $N/2$ раз посчитать ДПФ на 2 точки, в результате общее количество вычислительных операций составит $2^2 N/2 = 2N$, то есть количество операций линейно зависит от величины выборки.

В общем случае вычисление N -точечного БПФ требует выполнения примерно $M \log_2 N$ шагов с операциями сложения и $M \log_2 N/2$ операциями умножения на каждом шаге, что значительно меньше требуемых N^2 операций необходимых для вычисления обычного дискретного преобразования Фурье для последовательности той же длины.

Существует два основных способа разбиения — объединения: прореживание по времени и прореживание по частоте. По указанным выше причинам в данном курсе мы не будем подробно их рассматривать, но желающие могут ознакомиться с ними самостоятельно, так как подобные описания широко представлены в литературе.

Применение дискретного преобразования Фурье позволяет вычислить спектральное представление исходного дискретного сигнала в N точках частотной области. Для получения более гладкого представления получаемого спектра путем интерполяции его значений часто применяется способ получивший название "дополнения нулями".

Суть данного метода сводится к тому, исходная последовательность из N выборочных наблюдений из сигнала $x(k), k = 0, 1, \dots, N - 1$ (для БПФ $N = 2^m$) дополняется нулевыми значениями, например до формирования последовательности удвоенной длины, то есть

$$x(k) = 0 \text{ для } k = N, N + 1, \dots, 2N - 1.$$

Тогда для полученной удвоенной последовательности дискретное преобразование Фурье может быть записано как

$$X^\circ(n) = \sum_{k=0}^{2N-1} x(k) \exp\left(-j \frac{2\pi nk}{2N}\right),$$

что с учетом наличия нулевых значений приводит к

$$X^\circ(n) = \sum_{k=0}^{N-1} x(k) \exp\left(-j \frac{2\pi nk}{2N}\right).$$

Пусть k является четным числом ($k=2m$), тогда для четных значений $X^\circ(n)$ можно записать

$$X^\circ(m) = \sum_{k=0}^{N-1} x(k) \exp\left(-j \frac{2\pi nm}{N}\right),$$

то есть полностью совпадает со значениями N -точечного преобразования. Нечетные значения индекса k соответствуют интерполированным значениям дискретного преобразования Фурье, расположенным между значениями исходного N -точечного дискретного преобразования Фурье.

По мере увеличения количества дополняемых нулей можно добиться большей гладкости графика спектра за счет добавления дополнительных интерполированных значений.

Визуальная гладкость графика спектра, получаемая при дополнении исходной последовательности нулями, создает иллюзию более точного спектрального оценивания. Действительно, чем меньше шаг между двумя соседними линиями спектра, тем больше тонких деталей на нем можно различить. Однако в данном случае следует помнить, что дополнительные данные получены путем интерполяции и никакой дополнительной информации о спектре сигнала не содержат. То есть информации в интерполированном спектре содержится столько же, сколько ее было в исходном спектре, рассчитанном для последовательности длиной N отсчетов сигнала.

В действительности потенциально достижимая величина частотного разрешения связана с длиной наблюдаемой последовательности $x(k), k = 0, 1, \dots, N - 1$.

Для получения количественной оценки этой зависимости введем несколько полезных определений.

Сигнал $x(k)$ не может быть строго ограниченным по длительности и по ширине полосы своего спектра, но тем не менее может быть охарактеризован некоторым временным интервалом (эквивалентной длительностью), в котором сосредоточена большая часть его энергии во временной области, а также некоторым интервалом полосы частот (эквивалентной шириной полосы), в которой сосредоточена большая часть его энергии в частотной области. Для количественной характеристики меры казанных временной и частотной "концентрации" сигналов существует несколько подходов, но в рамках данного курса мы не будем подробно на них останавливаться. Отметим только, что длительность сигнала и эффективная ширина спектра связаны между собой обратно

пропорциональной зависимостью, то есть чем короче длительность сигнала, тем шире его спектр.

Произведение длительности сигнала на эффективную ширину его спектра называют базой сигнала. Например, для видеоимпульса эта величина составляет

$$\tau_3 \Delta \omega_3 \approx 2\pi \text{ или } \tau_3 \Delta f_3 \approx 1, \quad (2.14)$$

где τ_3 и $\Delta \omega_3$ эквивалентные длительность и ширина полосы сигнала.

Приведенное соотношение (2.14) иногда называемое соотношением неопределенности, не дает непосредственной возможности количественно определять степень разрешения отдельных спектральных линий. На сегодняшний день при оценке возможностей спектрального разрешения чаще всего рассматривается вопрос о возможности разрешать спектральные отклики двух близкорасположенных (по частоте и амплитуде) гармонических сигналов. Поскольку большинство применяемых критериев касается сигналов наблюдаемых на некотором конечном интервале времени, то естественным является использование в качестве параметра произведения интервала наблюдения на ширину полосы частот. Поэтому часто принято говорить, что разрешение в Герцах приближенно равно величине обратной времени наблюдения.

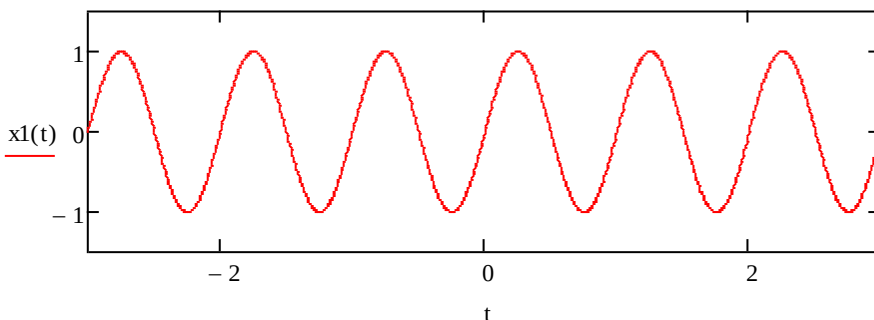
Следует отметить, что некоторые современные методы спектрального оценивания используют алгоритмы экстраполяции исходного сигнала за пределы окна наблюдения, что приводит к расширению эффективной длительности сигнала и соответствующему повышению частотного разрешения получаемых оценок.

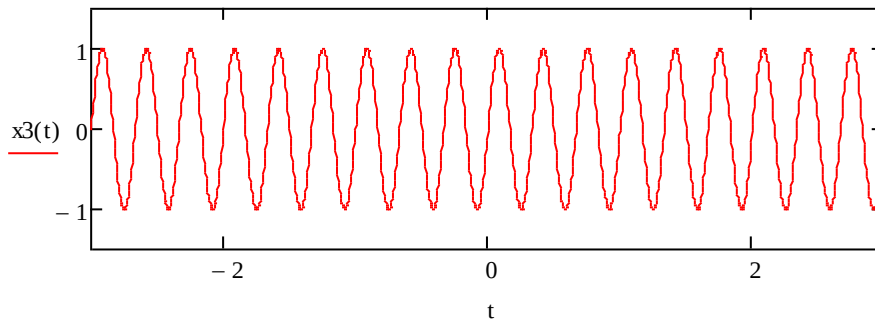
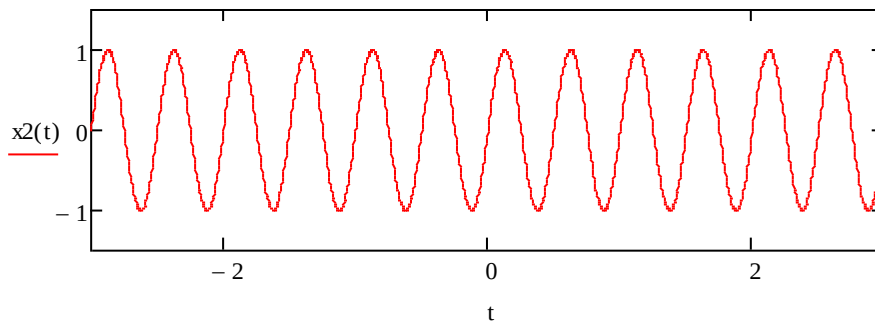
Пример 3.2

Рассмотрим применение БПФ для спектрального оценивания сигналов с использованием программы MathCad. Листинг рабочей сессии с комментариями приведен на рис.3.10.

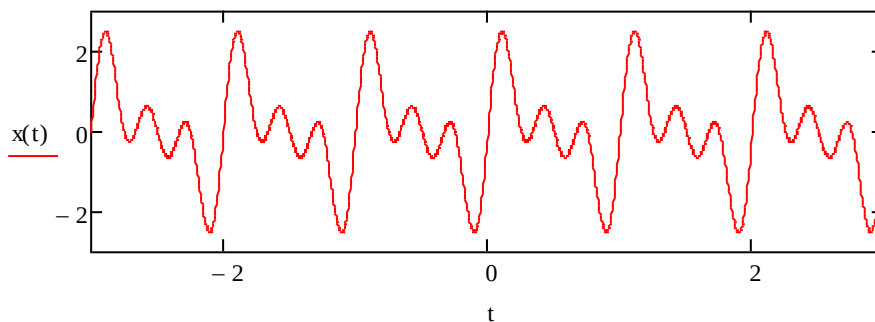
В качестве тестового сигнала сформируем колебание, являющееся суммой трех синусоид с единичной амплитудой, нулевой начальной фазой и кратными частотами.

```
f1 := 1
A1 := 1
f2 := 2
A2 := 1
f3 := 3
A3 := 1
x1(t) := sin(2 · π · f1 · t)
x2(t) := sin(2 · π · f2 · t)
x3(t) := sin(2 · π · f3 · t)
x(t) := A1 · sin(2 · π · f1 · t) + A2 · sin(2 · π · f2 · t) + A3 · sin(2 · π · f3 · t)
```





Таким образом, получаем итоговый тестовый сигнал



Перед тем, как применять к данному сигналу БПФ, сигнал требуется представить в дискретной форме.

Выберем частоту дискретизации превышающей частоту максимальной гармоники тестового сигнала в M раз, а количество данных используемых для БПФ-анализа 2^N .

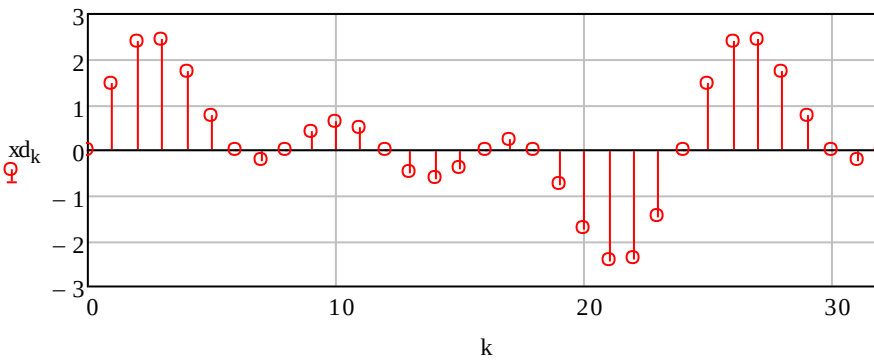
Выборку из значений исходного сигнала помещаем в отдельный массив.

```

M := 8
f0 := f3 · M
T0 := 1 / f0
m := 9
N := 2m
k := 0..N - 1
N = 512
nk := T0 · k
xdk := x(nk)

```

Фрагмент графика полученной дискретной функции имеет вид



Перед тем как проводить БПФ оценивание спектра исходного сигнала, проведем аналитический расчет теоретического спектра для случая бесконечного ряда наблюдений. (Выражение получается достаточно громоздким, поэтому здесь оно приводится частично)

$$x(t) \text{ fourier } t \rightarrow \pi \cdot (\Delta(\omega - 2 \cdot \pi) - \Delta(2 \cdot \pi + \omega) + \Delta(\omega - 4 \cdot \pi) - \Delta(4 \cdot \pi + \omega) + \Delta(\omega - 6 \cdot \pi) - \Delta(6 \cdot \pi + \omega))$$

Из полученного выражения видно, что как и ожидалось спектр исследуемого сигнала должен содержать три одинаковые составляющие, расположенные на частотах $\pm 2\pi f_1$, $\pm 2\pi f_2$, $\pm 2\pi f_3$.

В программе MathCad существует несколько встроенных функций для выполнения БПФ. Различие состоит в том, что они ориентированы на получение вещественного или комплексного результата, а также используют различные способы нормировки. (Это вопрос подробнее будет рассмотрен на практических занятиях).

Мы воспользуемся двумя из них.

Для вещественного и комплексного результата соответственно

$S2 := \text{FFT}(xd)$

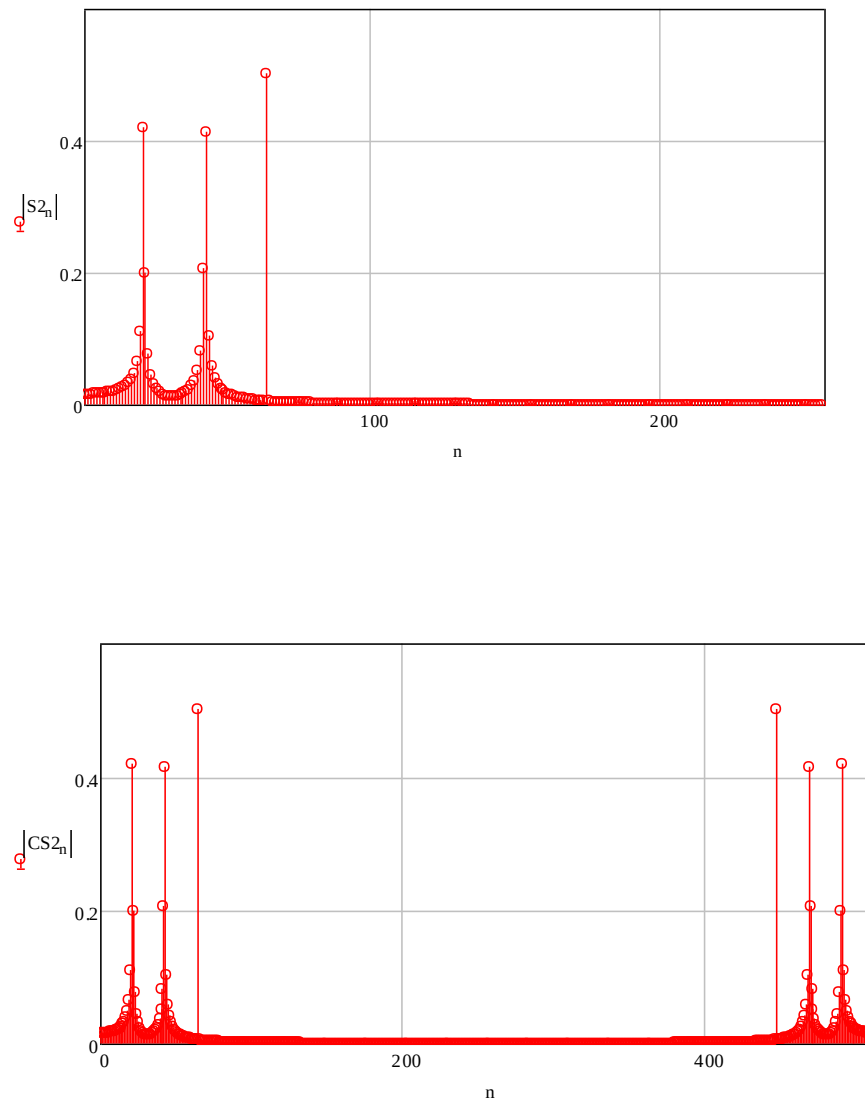
$CS2 := \text{CFFT}(xd)$

Проверив длину получаемых векторов, обнаруживаем, что вещественный вариант вдвое короче комплексного (из-за использования свойства симметрии спектра).

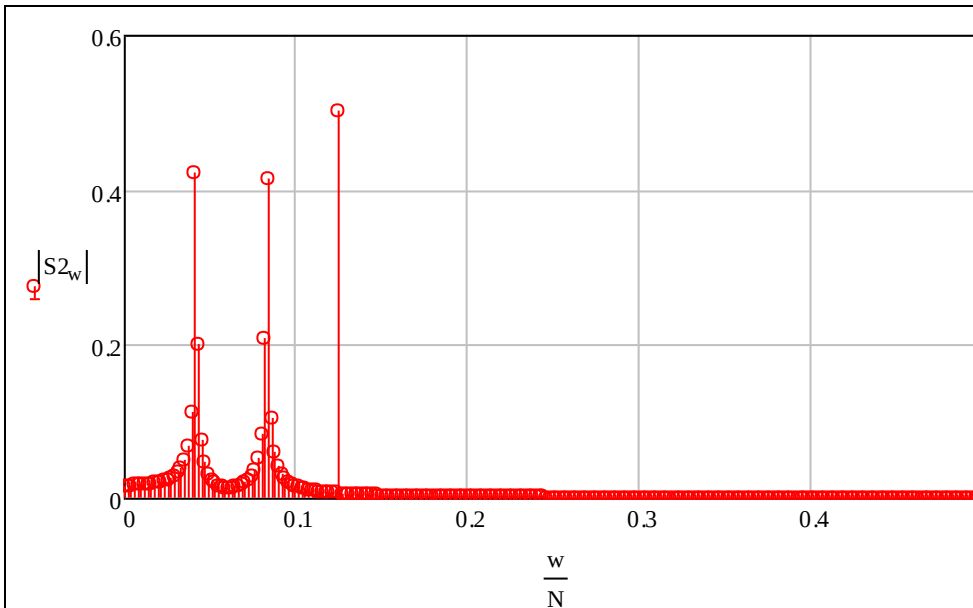
$\text{length}(S2) = 257$

$\text{length}(CS2) = 512$

Визуализация результатов имеет следующий вид



Пока на осях частот отложены только номера отсчетов в частотной области. Для практического использования этот результат лучше нормировать либо к частоте дискретизации (диапазон частот составит в вещественном случае от 0 до 0.5), либо представить его в реальных единицах. Вид графика при использовании нормированного к частоте дискретизации масштаба приведен ниже.



На полученном графике хорошо различимы все три составляющие исследуемого сигнала. При этом видно, что размытие спектра в областях первых двух гармоник существеннее по сравнению с областью, соответствующей третьей гармонике.

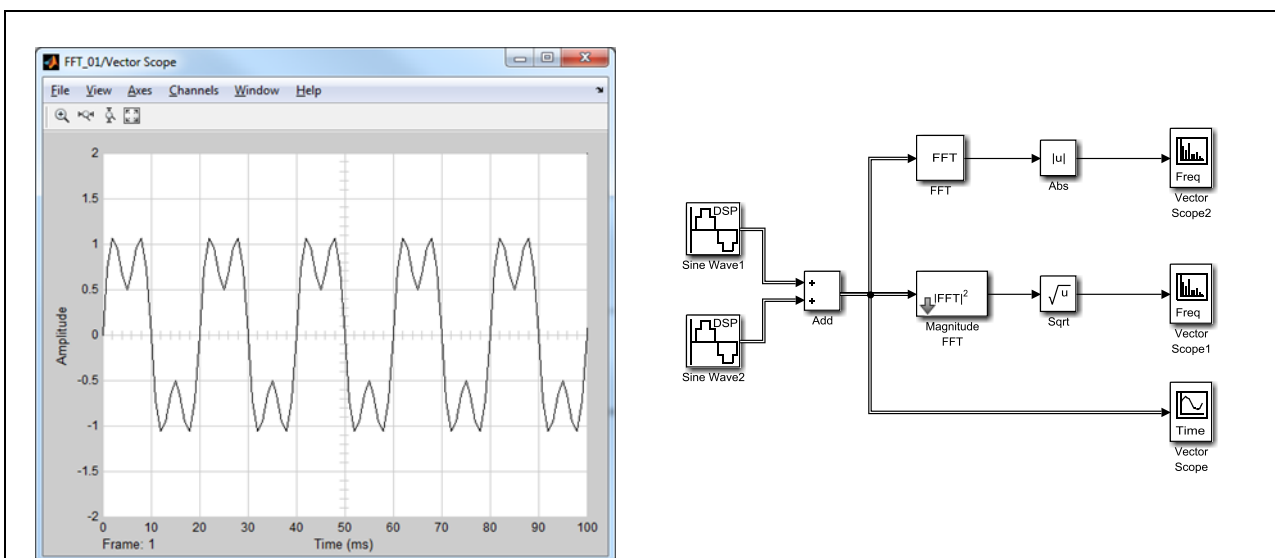
Причину этого явления рекомендуем объяснить самостоятельно.

Используя данный пример, рекомендуем самостоятельно провести моделирование с другими параметрами для ознакомления с проблемой частотного разрешения близких по частоте сигналов, а также провести моделирование процесса дополнения нулями.

Рис.3.10. Пример применения БПФ

Пример 3.3.

Пример определения спектра сигнала с применением быстрого преобразования Фурье в программе Simulink (Matlab) представлен на рис. 3.11. В качестве анализируемого сигнала выбран сигнал, представляющий собой сумму двух синусоидальных колебаний. Вычисление спектра анализируемого сигнала проведено двумя методами реализации алгоритма, которые дают одинаковый результат.



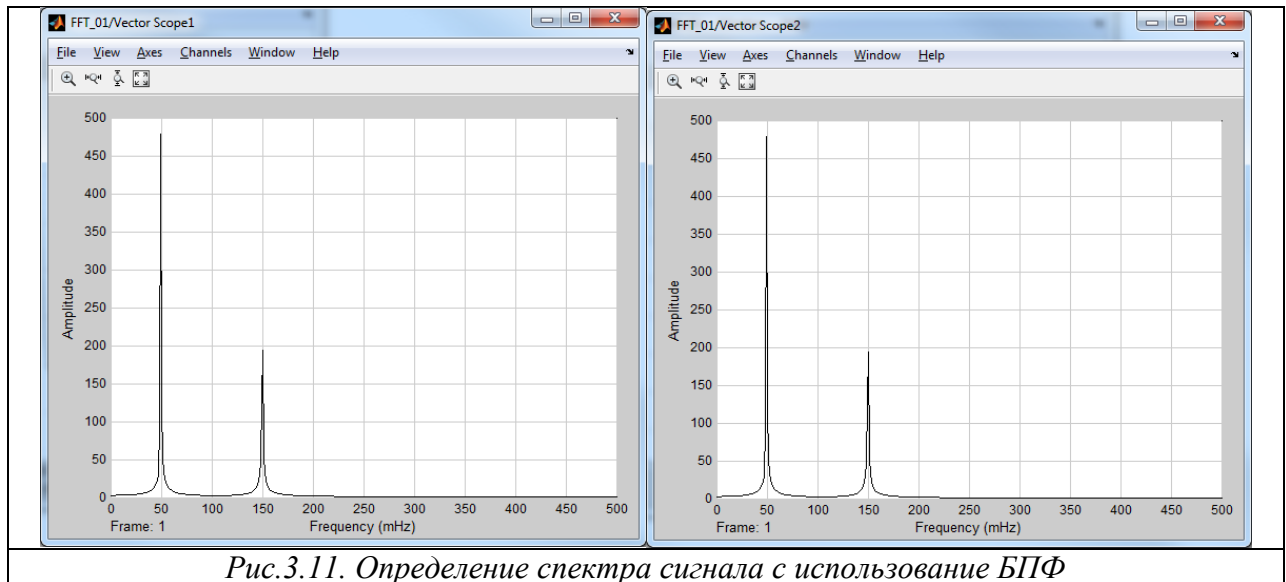


Рис.3.11. Определение спектра сигнала с использованием БПФ

3.5. Вопросы для самоконтроля и повторения

1. Каким образом из исходного аналогового сигнала формируется решетчатая функция?
2. Сформулируйте теорему Котельникова-Шеннона.
3. В чем состоит трудность восстановления исходного сигнала из последовательности дискретных наблюдений?
4. Каким образом работает фиксатор нулевого порядка?
5. Каким образом определяется дискретное преобразование Фурье?
6. Каковы основные свойства дискретного преобразования Фурье?
7. В чем состоит эффект Гиббса?
8. В чем состоит эффект «размывания» спектра?
9. Каким образом выполняется быстрое преобразование Фурье?
10. Для каких целей используется метод дополнения нулями?

ГЛАВА 4. Общие сведения о случайных процессах

4.1. Случайные процессы непрерывного времени и их основные характеристики

В предыдущих разделах мы рассматривали сигналы, которые являются определенными известными функциями времени. В этом случае значение сигнала может быть точно предсказано для любого момента времени. Такие сигналы называют детерминированными. Однако на практике часто встречаются сигналы, закон изменения которых имеет случайный характер, и, следовательно, значение сигнала в предстоящие моменты времени не может быть точно определено.

Функцию, значение которой при каждом значении независимой переменной является случайной величиной, называют случайной функцией. Если независимой величиной является время t , такие функции называют случайными (или **стохастическими**) процессами.

Случайный процесс $x(t)$ – представляет собой не определенную функцию, а **ансамбль** функций $x_1(t), x_2(t), \dots, x_n(t)$, каждая из которых может появиться во время проведения отдельного опыта. Каждую из функций ансамбля называют **реализацией случайного процесса**.

Для любого фиксированного значения времени $t = t_1$ конкретная реализация случайного процесса $x_n(t_1)$ представляет собой конкретную величину, а значение случайного процесса в этот же момент времени является величиной случайной $x(t_1)$ и называется **сечением** случайного процесса в момент времени t_1 .

Пример трех различных реализаций одного и того же случайного процесса приведен на рис.4.1.

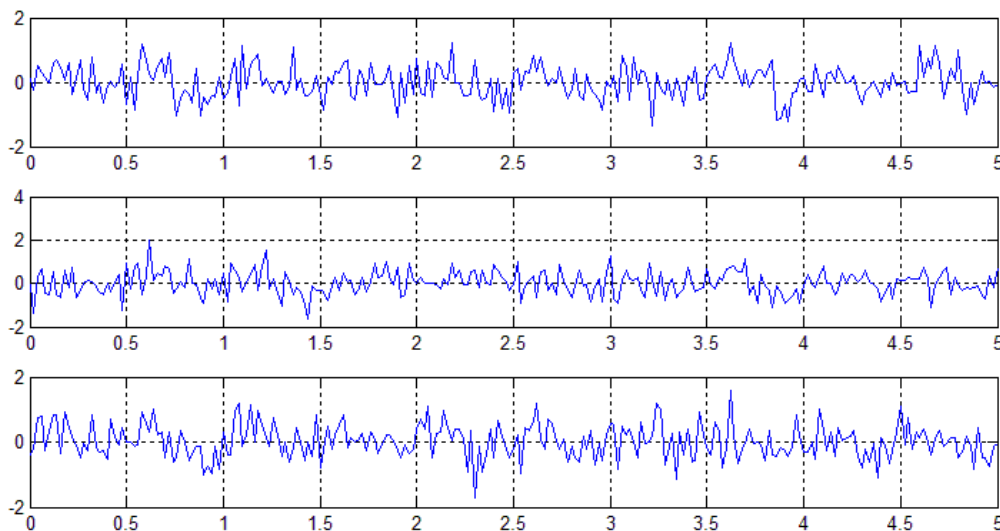


Рис.4.1. Пример трех реализаций одного случайного процесса

Статистические методы изучают не каждую из потенциально возможных реализаций, а обобщенные свойства всего множества реализаций случайного процесса

$x(t)$. Поэтому рассматривая стохастические свойства цепей, говорят не о свойствах относительно одного какого либо сигнала, а о свойствах цепи по отношению к целому классу однотипных (в стохастическом смысле) реализаций сигналов.

Таким образом, случайные процессы совмещают в себе свойства и случайных величин, и функций. Если зафиксировать время, то случайный процесс превратится в случайную величину, а если произвести отдельную реализацию процесса, то получается некоторая неслучайная функция.

Как известно, свойства случайной величины x определяются функциями распределения вероятности $F(x)$ и функцией плотности вероятности $p(x)$. Естественное обобщение этих понятий для случайных процессов позволяет ввести следующие определения.

Функция распределения вероятности случайного процесса $F(x, t_1)$ равна вероятности того, что в момент времени t_1 значение случайного процесса не превосходит x :

$$F(x, t_1) = P(x(t_1) \leq x). \quad (4.1)$$

Функция распределения вероятности $F(x, t)$ является неубывающей функцией, область значений которой лежит между нулем и единицей и при предельных значениях удовлетворяет условиям

$$\begin{aligned} F(-\infty, t_1) &= 0, \\ F(+\infty, t_1) &= 1. \end{aligned}$$

Вторая вероятностная характеристика случайного процесса называется **одномерной плотностью вероятности** $p(x, t_1)$ и представляет собой производную по x от функции распределения, то есть

$$p(x, t_1) = \frac{dF(x, t_1)}{dx}. \quad (4.2)$$

Плотность распределения вероятности представляет собой вероятность попадания значения случайного процесса $x(t)$ в момент времени t_1 в бесконечно малую окрестность значений около значения x . При расширении границ рассматриваемого интервала необходимо вычислить соответствующий интеграл

$$P(a \leq x(t_1) \leq b) = \int_a^b p(x, t_1) dx \quad (4.3)$$

Из (4.1) – (4.3) ясно, что

$$\int_{-\infty}^x p(x, t_1) dx = F(x, t_1) \quad (4.4)$$

и

$$\int_{-\infty}^{\infty} p(x, t_1) dx = F(\infty, t_1) = 1. \quad (4.5)$$

Знание одномерной плотности вероятности случайного процесса позволяет произвести различные усреднения, как самого процесса, так и некоторых функций от него. На практике в качестве усредненных параметров случайного процесса наиболее часто употребляются **математическое ожидание** и **дисперсия**.

Математическим ожиданием случайного процесса называют функцию времени, определяемую в соответствии со следующим уравнением

$$m_x(t) = M(x(t)) = \int_{-\infty}^{\infty} xp(x, t)dx. \quad (4.6)$$

Дисперсией случайного процесса называют функцию времени, определяемую в соответствии со следующим уравнением

$$D_x(t) = \int_{-\infty}^{\infty} (x - m_x(t))^2 p(x, t) dx = M\{(x(t) - m_x(t))^2\} = M\{x^2(t)\} - m_x^2(t). \quad (4.7)$$

Помимо дисперсии на практике часто используют величину, называемую средним квадратичным отклонением и связанную с дисперсией простым соотношением

$$\sigma(t) = \sqrt{D_x(t)}. \quad (4.8)$$

Наиболее часто в аналитических расчетах используется так называемый нормальный (или гауссовский) закон распределения случайной величины. Плотность вероятности нормального распределения описывается уравнением

$$p(x) = \frac{1}{\sigma_x \sqrt{2\pi}} e^{-\frac{(x-m_x)^2}{2\sigma_x^2}}. \quad (4.9)$$

График нормальной плотности распределения приведен на рис.4.2.

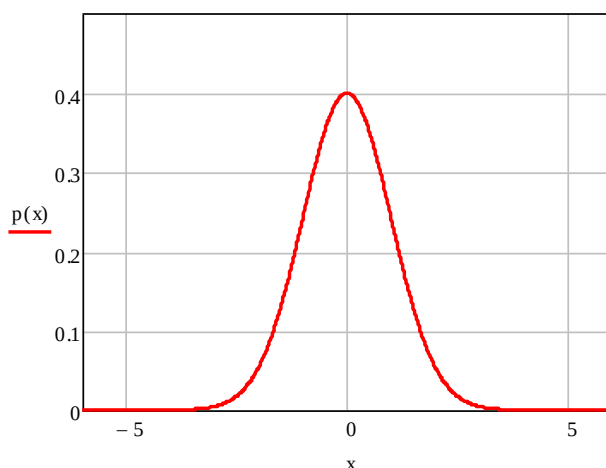


Рис.4.2. График плотности вероятности для нормального закона распределения

Несколько реже, но тоже достаточно часто используются такие характеристики случайного процесса как **мода** и **медиана**.

Мода представляет собой наиболее вероятное в момент времени t значение случайного процесса, то есть значение случайного процесса соответствующее максимуму плотности вероятности.

Медиана представляет собой такое значение случайного процесса в момент времени t , что отклонения от него в любую сторону равновероятны, то есть

$$F(\text{med}(x(t)), t) = \frac{1}{2}. \quad (4.10)$$

4.2. Корреляционная функция

Одномерная плотность вероятности случайного процесса и вводимые на ее основе количественные вероятностные характеристики (4.6) – (4.10) позволяют получить важную информацию о свойствах процесса, однако в ряде случаев ее оказывается недостаточно. Одномерные характеристики позволяют судить о предполагаемом поведении процесса в отдельные моменты времени, но не дают никакой информации о связи между разнесенными во времени значениями процесса между собой.

Существенно больше информации о свойствах случайного процесса можно получить, располагая большим количеством сечений изучаемого процесса. Таких сечений может быть несколько и в общем случае можно определить многомерную (n-мерную) функцию распределения для случайного процесса $x(t)$ в виде

$$F(x_1, x_2, \dots, x_n, t_1, t_2, \dots, t_n) = P(x(t_1) \leq x_1, x(t_2) \leq x_2, \dots, x(t_n) \leq x_n). \quad (4.11)$$

Тогда соответствующая многомерная плотность вероятности может быть связана с многомерной функцией распределения (4.11) уравнением

$$\begin{aligned} F(x_1, x_2, \dots, x_n, t_1, t_2, \dots, t_n) &= \\ &= \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_n} p(\xi_1, \xi_2, \dots, \xi_n, t_1, t_2, \dots, t_n) d\xi_1 d\xi_2 \dots d\xi_n. \end{aligned} \quad (4.12)$$

Аналогично одномерной плотности вероятности многомерная плотность (4.12) является неотрицательной величиной и при интегрировании в бесконечных пределах дает единицу.

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} p(x_1, x_2, \dots, x_n, t_1, t_2, \dots, t_n) dx_1 dx_2 \dots dx_n = 1. \quad (4.13)$$

При этом если известна n-мерная плотность вероятности, то для $m < n$ можно рассчитать m-мерную плотность с помощью соответствующего интегрирования по $m + 1, m + 2, \dots, n$ координатам:

$$\begin{aligned} p(x_1, x_2, \dots, x_m, t_1, t_2, \dots, t_m) &= \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} p(x_1, x_2, \dots, x_n, t_1, t_2, \dots, t_n) dx_{m+1} dx_{m+2} \dots dx_n. \end{aligned} \quad (4.14)$$

К счастью, в большинстве практических ситуаций необходимые результаты удается получить уже при использовании двумерных плотностей вероятности и функций распределения, поэтому вычисление и преобразование многомерных функций (4.12), (4.14) не требуется.

Использование двумерной плотности вероятности позволяет определить характеристику случайного процесса, называемую ковариационной функцией и определяемой в соответствии с уравнением

$$K_x(t_1, t_2) = M\{x(t_1)x(t_2)\}. \quad (4.15)$$

Ковариационная функция (4.15) представляет собой статистическое усреднение произведения значений случайного процесса в моменты времени t_1 и t_2 .

В соответствии с определением операции вычисления математического ожидания (4.6) для ковариационной функции (4.15) можно записать следующее уравнение

$$K_x(t_1, t_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 p(x_1, x_2, t_1, t_2) dx_1 dx_2. \quad (4.16)$$

На практике довольно часто применяют рассмотрение не самого случайного процесса, а процесса получаемого из исходного путем вычитания его же математического ожидания. Такие процессы называют **центрированными** и они предоставляют информацию о флуктуационной составляющей исходного процесса.

Для центрированных процессов вместо ковариационной функции (4.16) используют корреляционную функцию, которая может быть определена в соответствии со следующим уравнением

$$\begin{aligned} R_x(t_1, t_2) &= M\{[x(t_1) - m_x(t_1)][x(t_2) - m_x(t_2)]\} = \\ &= K_x(t_1, t_2) - m_x(t_1)m_x(t_2). \end{aligned} \quad (4.17)$$

Корреляционная функция (4.16), (4.17) характеризует степень стохастической связи значений случайного процесса $x(t)$, наблюдаемых в моменты времени t_1 и t_2 .

Рассмотрим основные свойства корреляционной функции.

1. При предельном сближении этих моментов времени $t = t_1 = t_2$, введенные определения для корреляционной функции совпадут с уравнениями, определяющими дисперсию случайного процесса, то есть

$$R_x(t, t) = K_x(t, t) - m_x^2(t) = D_x(t).$$

2. Корреляционная функция симметрична относительно своих аргументов

$$R_x(t_1, t_2) = R_x(t_2, t_1).$$

3. $|R_x(t_1, t_2)| \leq \sigma_x(t_1)\sigma_x(t_2)$ или что то же самое

$$|R_x(t_1, t_2)| \leq \sqrt{R_x(t_1, t_1)R_x(t_2, t_2)}.$$

По аналогии с корреляционной функцией, представляющей оценку статистической связи значений одно и того же случайного процесса для двух моментов времени, для двух различных случайных процессов можно ввести **взаимную корреляционную функцию** в виде

$$R_{xy}(t_1, t_2) = M\{[x(t_1) - m_x(t_1)][y(t_2) - m_y(t_2)]\}. \quad (4.18)$$

Взаимная корреляционная функция (4.18) обладает следующими основными свойствами

1. $R_{xy}(t_1, t_2) \neq R_{yx}(t_1, t_2)$, но $R_{xy}(t_1, t_2) = R_{yx}(t_2, t_1)$.
2. $R_{xx}(t_1, t_2) = R_x(t_1, t_2)$.
3. $|R_{xy}(t_1, t_2)| \leq \sigma_x(t_1)\sigma_y(t_2)$.

Свойства №3, приведенные для корреляционной и взаимной корреляционных функций, связаны с понятиями коррелированности и независимости случайных величин. Условием статистической независимости двух случайных величин x_1, x_2 является взаимная независимость их совместной плотности вероятности, которая в этом случае представляется произведением плотностей вероятности каждой из величин, то есть $p(x_1, x_2) = p_1(x_1)p_2(x_2)$.

При наличии статистической связи между двумя случайными величинами, их статистические свойства зависят от значений, принимаемых другой величиной. Эта связь может быть слабой и сильной, иметь линейный и нелинейный характер.

Для оценки линейной статистической связи между двумя случайными величинами используется коэффициент корреляции, определяемый как

$$r_{12} = \frac{M\{x_1 x_2\} - M\{x_1\}M\{x_2\}}{\sqrt{D\{x_1\}D\{x_2\}}}. \quad (4.19)$$

Модуль коэффициента корреляции (4.19) может принимать значения от нуля до единицы. При этом значение единицы достигается, если между сравниваемыми случайными величинами существует связывающее их линейное соотношение.

Равенство нулю коэффициента корреляции свидетельствует об отсутствии линейной корреляционной связи между сравниваемыми случайными величинами, которые в данном случае называются некоррелированными.

Следует иметь в виду, что из статистической независимости случайных величин следует их некоррелированность, но обратное утверждение в общем случае является неверным, так как некоррелированные величины могут быть зависимыми.

По аналогии с коэффициентом корреляции (4.19) для случайных процессов иногда используют нормированную корреляционную функцию

$$r_x = \frac{R_x(t_1, t_2)}{\sigma_x(t_1)\sigma_x(t_2)}. \quad (4.20)$$

Как мы отмечали ранее, каждая отдельная реализация случайного процесса представляет собой некую детерминированную функцию, поэтому к ней может быть применено преобразование Фурье. Поскольку все отдельные реализации процесса будут отличаться друг от друга, то, естественно, что и их спектры будут отличны друг от друга.

4.3. Понятия стационарности и эргодичности

Различные случайные процессы в зависимости от того, каким образом изменяются их статистические характеристики с течением времени, принято подразделять на **стационарные** и **нестационарные**.

Стационарные случайные процессы в свою очередь подразделяются на **стационарные в узком смысле** и **стационарные в широком смысле** процессы.

Стационарным в узком смысле называют случайный процесс $x(t)$, если его n -мерные функция распределения и плотность вероятности не зависят от положения начала координат во времени, то есть

$$\begin{aligned} F(x_1, x_2, \dots, x_n, t_1, t_2, \dots, t_n) &= F(x_1, x_2, \dots, x_n, t_1 + \tau, t_2 + \tau, \dots, t_n + \tau), \\ p(x_1, x_2, \dots, x_n, t_1, t_2, \dots, t_n) &= p(x_1, x_2, \dots, x_n, t_1 + \tau, t_2 + \tau, \dots, t_n + \tau). \end{aligned}$$

Это свойство означает, что два процесса $x(t)$ и $x(t + \tau)$ имеют одинаковые статистические свойства для любого произвольного значения τ , то есть статистические характеристики процесса неизменны во времени.

Если ограничить указанные требования стационарности только к одномерным и двумерным функциям распределения и плотностям вероятности, то случайный процесс будет **стационарным в широком смысле**.

Для стационарного процесса математическое ожидание и дисперсия не зависят от времени, а корреляционная функция зависит не от самих моментов времени t_1, t_2 , а от разницы между ними

$$R_x(t_1, t_2) = R_x(t_2 - t_1) = R_x(\tau). \quad (4.21)$$

Распространение данного условия на два случайных процесса приводит к понятию стационарно связанных процессов, для которых справедливо

$$R_{xy}(t_1, t_2) = R_{xy}(t_2 - t_1). \quad (4.22)$$

Понятие стационарного в широком смысле случайного процесса вводится тогда, когда в качестве статистических характеристик используются только математическое ожидание и корреляционные функции. Часть теории случайных процессов, которая описывает их свойства через математическое ожидание и корреляционные функции носит название корреляционной теории. Для случайных процессов с нормальным законом распределения понятия стационарности в узком и широком смыслах совпадают.

В теории случайных процессов пользуются двумя понятиями средних значений: *среднее по множеству* значений и *среднее по времени* наблюдения.

В общем случае среднее по множеству и среднее по времени для одного и того же процесса будут различными. Однако для некоторого класса случайных процессов они совпадают. Класс таких процессов получил название *эргодических*.

Основные свойства характеристик стационарных случайных процессов (4.21), (4.22) следующие.

1. Корреляционная функция стационарного процесса является четной, то есть

$$R_x(\tau) = R_x(-\tau).$$

2. При $\tau = 0$

$$R_x(0) = D_x = \text{const.}$$

3. Корреляционная функция постоянной величины равна квадрату этой величины

$$x(t) = A = \text{const}$$

$$R_x(\tau) = A^2.$$

4. Корреляционная функция суммы двух процессов $x(t)$ и $y(t)$ определяется следующим образом

$$R_{x+y}(\tau) = R_x(\tau) + R_y(\tau) + R_{xy}(\tau) + R_{yx}(\tau)$$

5. Ни при каких значениях своего аргумента корреляционная функция стационарного процесса не может быть по абсолютной величине больше своего значения в начале координат

$$|R_x(\tau)| \leq R_x(0) = D_x.$$

4.4. Спектральная плотность мощности и ее связь с корреляционной функцией

При введении характеристик случайного процесса с целью получения усредненных параметров вводилось усреднение по всему ансамблю реализаций случайного процесса.

Такую же операцию можно произвести и со спектрами.

$$\overline{S(\omega)} = \overline{\int_{-\infty}^{\infty} x(t)e^{-j\omega t} dt} = \int_{-\infty}^{\infty} \overline{x(t)}e^{-j\omega t} dt = \int_{-\infty}^{\infty} m_x(t)e^{-j\omega t} dt, \quad (4.23)$$

где верхней чертой обозначена операция усреднения по ансамблю значений. Однако для центрированных процессов математическое ожидание равно нулю, что приведет и к нулевому значению усредненного по ансамблю спектра. Это явление связано с тем, что фазы спектральных составляющих в различных реализациях являются случайными и независимыми величинами.

Для того чтобы устранить данный эффект, можно от амплитудных значений перейти к мощности случайного сигнала. Поскольку мощность сигнала не зависит от фазы, то при таком подходе эффекта нулевого усреднения наблюдаться не должно. Пусть задан некоторый случайный процесс $x(t)$. Для его реализации, рассматриваемой на интервале $[-\frac{T}{2}; \frac{T}{2}]$, в соответствии с равенством Парсеваля его энергию на данном интервале можно вычислить следующим образом

$$\mathfrak{E}_T = \int_{-\frac{T}{2}}^{\frac{T}{2}} x^2(t) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |X_T(\omega)|^2 d\omega, \quad (4.24)$$

где $X_T(\omega)$ – спектральная плотность для наблюдаемого сегмента реализации случайного сигнала.

Разделив энергию (4.24) на величину интервала наблюдения можно перейти к средней мощности сигнала на рассматриваемом интервале

$$P_T = \frac{1}{T} \mathfrak{E}_T = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{|X_T(\omega)|^2}{T} d\omega. \quad (4.25)$$

При увеличении интервала наблюдения сигнала T в (4.25) энергия неограниченно возрастает. Если при этом средняя мощность стремится к некоторому пределу, то путем усреднения по ансамблю наблюдений с учетом предельного перехода можно получить следующее

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \lim_{T \rightarrow \infty} \frac{|X_T(\omega)|^2}{T} d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_P(\omega) d\omega, \quad (4.26)$$

где $S_P(\omega) = \lim_{T \rightarrow \infty} \frac{|X_T(\omega)|^2}{T}$ – спектральная плотность мощности средней мощности рассматриваемой реализации случайного процесса.

Поскольку для стационарного и эргодического процесса усреднение по времени и усреднение по ансамблю совпадают, то для таких процессов найденная плотность мощности (4.25), (4.26) характеризует весь процесс в целом. При этом понятно, что

$$D_x = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_P(\omega) d\omega, \quad (4.27)$$

где D_x – дисперсия процесса $x(t)$.

$S_P(\omega)$ представляет собой не отрицательную вещественную функцию, обладающую свойствами четности.

Вид спектральной плотности мощности и корреляционной функции случайного процесса характеризуют скорость ее изменчивости. Чем медленнее меняется во времени случайный процесс, тем уже его энергетический спектр и тем шире его корреляционная функция.

Точную связь между корреляционной функцией и спектральной плотностью мощности случайного процесса определяет теорема Винера-Хинчина, в соответствии с которой

$$S_{Px}(\omega) = \int_{-\infty}^{\infty} R_x(\tau) e^{-j\omega\tau} d\tau, \quad (4.28)$$

$$R_x(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{Px}(\omega) e^{j\omega\tau} d\omega. \quad (4.29)$$

Другими словами, корреляционная функция случайного процесса и его спектральная плотность мощности связаны между собой преобразованием Фурье.

Поскольку корреляционная функция и спектральная плотность мощности в (4.28), (4.29) являются четными вещественными функциями, то можно перейти от комплексной формы их записи к вещественному эквиваленту с соответствующей заменой пределов интегрирования и удвоением нормирующих коэффициентов

$$S_{Px}(\omega) = 2 \int_0^{\infty} R_x(\tau) \cos(\omega\tau) d\tau, \quad (4.30)$$

$$R_x(\tau) = \frac{1}{\pi} \int_0^{\infty} S_{Px}(\omega) \cos(\omega\tau) d\omega. \quad (4.31)$$

Чем шире энергетический спектр случайного сигнала, тем меньше интервал его корреляции и, с другой стороны, чем больше интервал корреляции случайного сигнала, тем уже его спектр.

Числовыми характеристиками, служащими для оценки интервала корреляции и ширины энергетического спектра случайного сигнала служат величины, получившие название **интервала корреляции** и **эффективной ширины спектра** соответственно. Однако единого общепринятого способа определения этих характеристик не существует. Различные авторы используют для этих целей отличающиеся определения. Один из способов задания указанных величин, например, выглядит следующим образом

$$T_R = \frac{1}{R_x(0)} \int_0^{\infty} |R_x(\tau)| d\tau, \quad (4.32)$$

$$\Delta\omega_{эф} = \frac{1}{S_{Pmax}} \int_0^{\infty} S_P(\omega) d\omega, \quad (4.33)$$

где S_{Pmax} – максимальное значение спектральной плотности мощности.

Соотношение между эффективной шириной спектра и интервалом корреляции (4.32), (4.33) задается так называемым соотношением неопределенности, в соответствии с которым

$$T_R \Delta\omega_{эф} \sim 2\pi \quad (4.34a)$$

или

$$T_R \Delta f_{эф} \sim 1. \quad (4.34b)$$

Данные соотношения определяют только общий характер зависимостей, а их конкретные числовые соотношения в (4.34a), (4.34b) зависят от способа определения интервала корреляции и эффективной ширины спектра.

Особое место среди всех случайных процессов занимает стационарный случайный процесс, спектральная плотность мощности которого постоянна на всех частотах, то есть

$$S_{Px}(\omega) = S_0 = \text{const.} \quad (4.35)$$

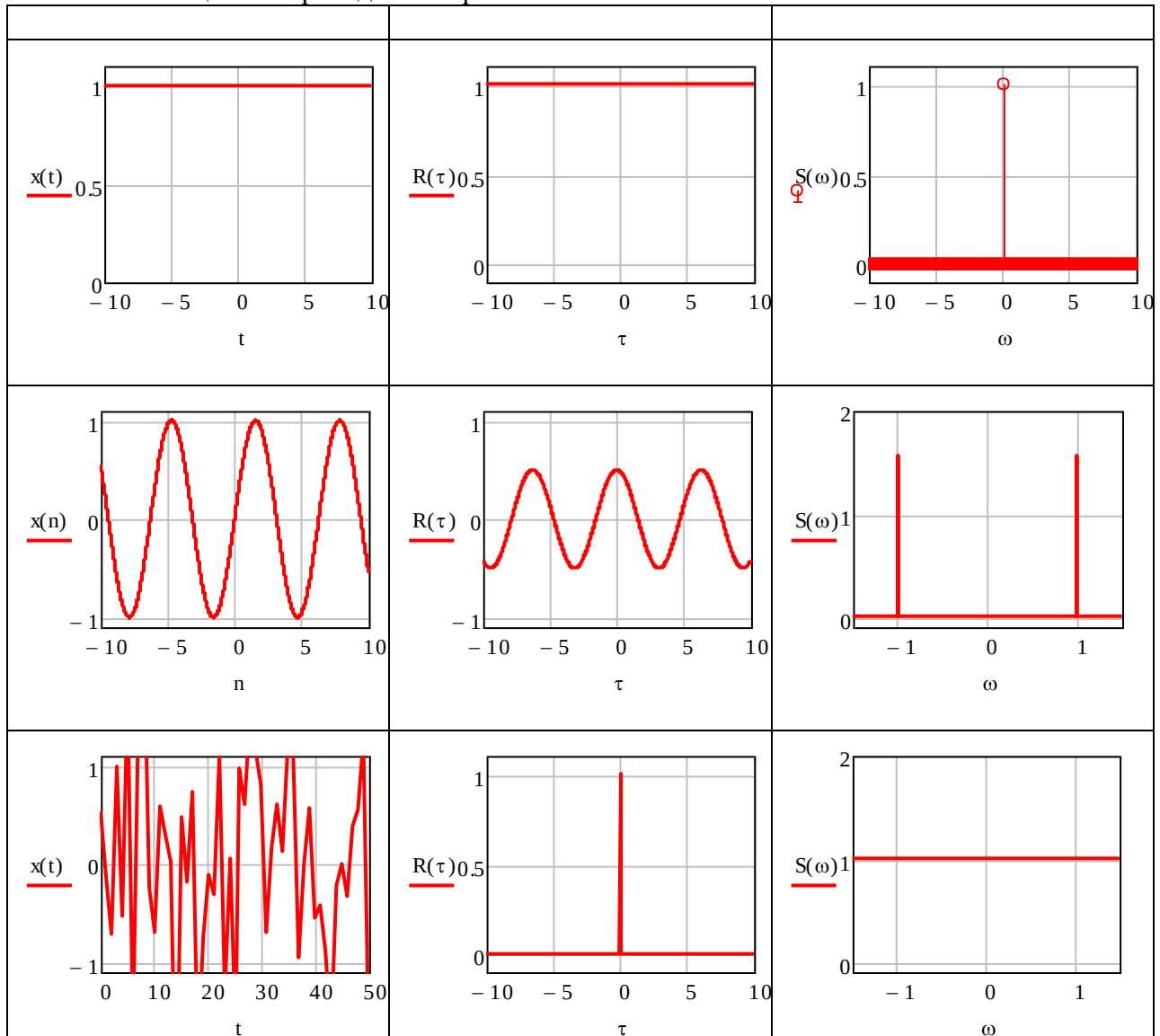
Такой сигнал получил название белого шума. Его корреляционная функция может быть определена по теореме Винера-Хинчина в следующем виде

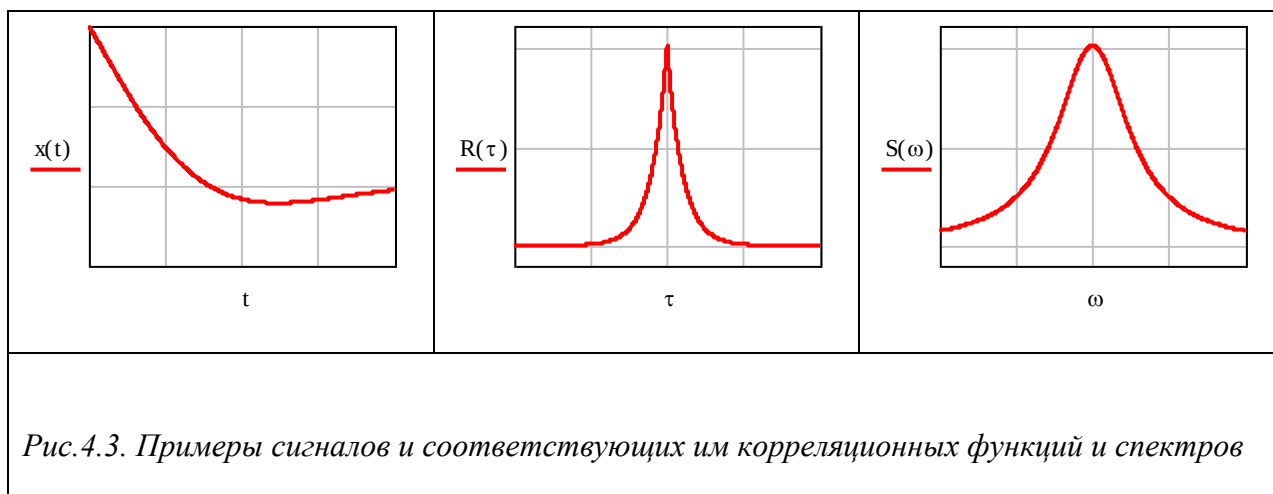
$$R(\tau) = \frac{S_0}{2\pi} \int_{-\infty}^{\infty} e^{j\omega\tau} d\omega = S_0 \delta(\tau). \quad (4.36)$$

Таким образом, корреляционная функция (4.36) белого шума представляет собой дельта функцию и равна нулю всюду, кроме начала координат.

Белый шум (4.35) является предельной идеализацией случайного процесса, значения которого некоррелированы в несовпадающие моменты времени. Белый шум – удобная абстрактная математическая модель, широко используемая в аналитических расчетах. Впрочем, на практике, когда реальная полоса пропускания цепи существенно уже эффективной ширины сигнала, в качестве приближенной модели также используется модель белого шума.

Примеры соотношения вида сигнала, его корреляционной функции и спектральной плотности мощности приведены на рис.4.3.





4.5. Случайный узкополосный сигнал

При рассмотрении модулированных сигналов мы отмечали важную роль, которую в радиотехнике играют процессы, спектр которых расположен в относительно узкой области вблизи некоторой заданной частоты.

Для рассмотрения статистических свойств узкополосных сигналов введем понятие квазигармонического колебания

$$x(t) = A(t) \cos[\omega_0 t + \theta(t)], \quad (4.37)$$

где $A(t), \theta(t)$ – случайные функции, скорость изменения которых во времени существенно меньше, чем у функции $\cos \omega_0 t$.

Для определения статистических параметров огибающей и фазовой функции в (4.37) воспользуемся концепцией аналитического сигнала, подробно рассмотренной ранее применительно к детерминированным сигналам. Аналогичным образом для случайного процесса $x(t)$ построим комплексный процесс

$$x_a(t) = x(t) + jx_1(t), \quad (4.38)$$

где $x_1(t)$ – сопряженный случайный процесс, реализации которого связаны с реализациями исходного процесса $x(t)$ преобразованием Гильберта

$$x_1(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x(\tau)}{t - \tau} d\tau.$$

Тогда исходный сигнал (4.37) может быть представлен как вещественная часть аналитического сигнала (4.38)

$$x(t) = \operatorname{Re} x_a(t), \quad (4.39)$$

что позволяет однозначно определить модулирующие функции

$$A(t) = \sqrt{x^2(t) + x_1^2(t)}, \quad (4.40)$$

$$\psi(t) = \operatorname{arctg} \frac{x_a(t)}{x(t)}, \quad (4.41)$$

где $\psi(t) = \omega_0 t + \theta(t)$.

Применение процедуры определения математического ожидания к сопряженной по Гильберту функции (4.39) дает следующее

$$M\{x_1(t)\} = M\left\{\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x(\tau)}{t - \tau} d\tau\right\} = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{M\{x(\tau)\}}{t - \tau} d\tau = 0. \quad (4.40)$$

Поскольку исходный процесс $x(t)$ является нормальным, то нормальным будет и сопряженный процесс $x_1(t)$, как результат применения линейного интегрального преобразования.

Кроме того при изучении преобразования Гильберта для детерминированных сигналов мы отмечали, что спектральные функции исходного и сопряженного сигналов связаны соотношением

$$S_1(\omega) = \begin{cases} -jS(\omega), & \text{при } \omega > 0 \\ S(\omega), & \text{при } \omega < 0, \end{cases}$$

а силу того, что преобразование Гильберта исключает постоянную составляющую из сигнала, $S_1(0) = 0$. Соответственно при переходе к спектральным плотностям мощности становится понятно, что $S_{Px1}(\omega) = S_{Px}(\omega)$.

Поскольку корреляционная функция связана со спектральной плотностью мощности преобразованием Фурье, то из равенства энергетических спектральных плотностей с очевидностью следует равенство соответствующих корреляционных функций $R_{x1}(\tau) = R_x(\tau)$.

Проверка статистической связи между процессами $x(t)$ и $x_1(t)$ требует довольно громоздких преобразований, поэтому мы ограничимся лишь демонстрацией факта ее отсутствия путем расчета взаимной корреляционной функции для нулевого значения аргумента $R_{xx1}(\tau)|_{\tau=0}$.

Для удобства записи получаемых уравнений при использовании формулы преобразования Гильберта применим другое обозначение интегрируемой переменной.

$$R_{xx1}(0) = M\left\{x(t) \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x(z)}{t - z} dz\right\} = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{M\{x(t)x(z)\}}{t - z} dz = -\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{R_x(\tau)}{\tau} d\tau = 0$$

Полученный результат показывает некоррелированность процессов $x(t)$ и $x_1(t)$ в совпадающие значения времени, что в совокупности со свойством нормальности процессов обеспечивает их статистическую независимость в эти моменты времени.

В соответствии с правилом задания комплексного случайного сигнала

$$x_a(t) = x(t) + jx_1(t),$$

его можно представить вектором на комплексной плоскости, проекции которого на вещественную и мнимую оси дают

$$\begin{aligned} x(t) &= \operatorname{Re} x_a(t), \\ x_1(t) &= \operatorname{Im} x_a(t). \end{aligned}$$

Как было выяснено ранее, эти значения статистически независимы, имеют нормальное распределение с нулевым средним и равными корреляционными функциями (то есть, в том числе и равными дисперсиями). Исходя из этого совместная плотность вероятности процессов $x(t)$ и $x_1(t)$ может быть представлена как произведение их одномерных плотностей:

$$p(x, x_1) = p(x)p(x_1) = \frac{1}{2\pi\sigma_x^2} e^{-\frac{x^2 + x_1^2}{2\sigma_x^2}} \quad (4.41)$$

Выражение (4.41) получено для декартовых координат представления комплексной плоскости. Для получения статистических свойств огибающей и фазы необходимо перейти к такой системе координат, в которой их представление является более удобным для анализа. Такой координатной системой в данном случае является полярная система координат. В этой системе связь между огибающей и фазой задается парой уравнений

$$\begin{aligned}x &= A \cos \psi, \\x_1 &= A \sin \psi.\end{aligned}$$

Переход от одной координатной системы к другой не должен приводить к изменению статистических свойств, поэтому вероятность попадания в бесконечно малую область в окрестности каждой из точек комплексной плоскости должна оставаться неизменной.

В декартовых координатах площадь малой области задается произведением $dx dx_1$, а в полярных координатах соответственно $A dA d\psi$.

Следовательно,

$$p_{xx_1}(x, x_1) dx dx_1 = p_{xx_1}(A \cos \psi, A \sin \psi) A dA d\psi = p_{A\psi}(A, \psi) dA d\psi.$$

Графически процесс перехода от прямоугольной системы координат к полярной представлен на рис.4.4.

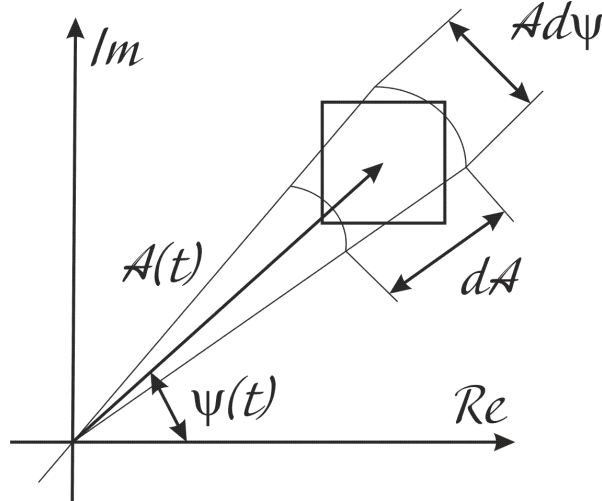


Рис.4.4. Переход от прямоугольной системы координат к полярной

Тогда для искомой плотности вероятности нетрудно записать

$$p_{A\psi}(A, \psi) = A p_{xx_1}(A \cos \psi, A \sin \psi) = \frac{A}{2\pi\sigma_x^2} e^{-\frac{A^2}{2\sigma_x^2}}. \quad (4.42)$$

Общее правило перехода к плотностям вероятности пониженного порядка мы рассматривали ранее. Оно состоит в необходимости проведения интегрирования плотности вероятности более высокого порядка по координатам, порядок которых превышает требуемый. В данном случае, пользуясь этим методом, мы можем перейти в (4.42) от совместной плотности вероятности двух случайных величин к одномерным плотностям вероятности для огибающей и фазы, то есть

$$p_A(A) = \int_0^{2\pi} p_{A\psi}(A, \psi) d\psi,$$

$$p_\psi(\psi) = \int_0^{2\pi} p_{A\psi}(A, \psi) dA.$$

Проводя указанное интегрирование для плотности вероятности амплитуды запишем

$$\begin{aligned}p_A(A) &= \int_0^{2\pi} p_{A\psi}(A, \psi) d\psi = p_A(A) = \int_0^{2\pi} \frac{A}{2\pi\sigma_x^2} e^{-\frac{A^2}{2\sigma_x^2}} d\psi = \\&= \frac{A}{2\pi\sigma_x^2} e^{-\frac{A^2}{2\sigma_x^2}} (2\pi - 0) = \frac{A}{\sigma_x^2} e^{-\frac{A^2}{2\sigma_x^2}}.\end{aligned} \quad (4.43)$$

Плотность вероятности, описываемая полученным выражением (4.43), носит название закона Релея и в относительных безразмерных единицах принимает вид

$$p(\alpha) = \alpha e^{-\frac{\alpha^2}{2}}. \quad (4.44)$$

График, соответствующий закону Релея (4.44), приведен на рис.4.5.

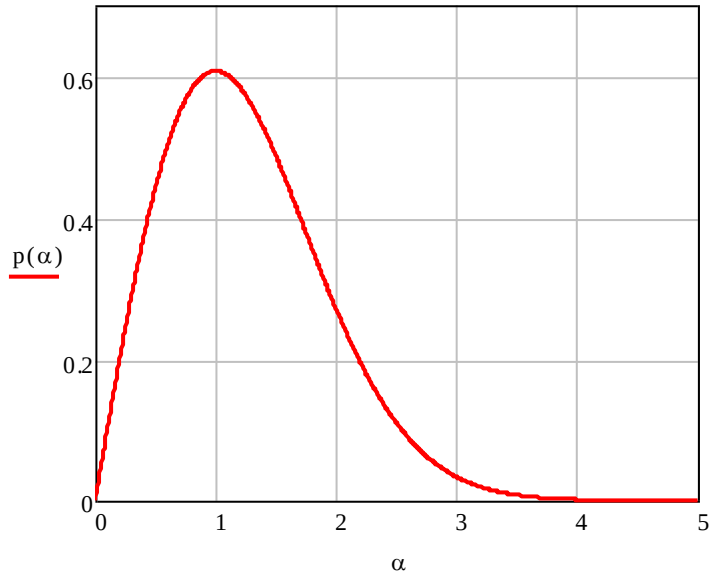


Рис.4.5. Плотность вероятности соответствующая закону Релея

Знание плотности вероятности позволяет рассчитать математическое ожидание и дисперсию огибающей:

$$M\{A\} = \int_0^{\infty} A p_A(A) dA = \int_0^{\infty} \frac{A^2}{\sigma_x^2} e^{-\frac{A^2}{2\sigma_x^2}} dA = \sqrt{\frac{\pi}{2}} \sigma_x,$$

$$D_A = M\{A^2\} - M^2\{A\} = \left(2 - \frac{\pi}{2}\right) \sigma_x^2.$$

Вычисление плотности вероятности для фазы проводится аналогично и дает

$$p_{\psi}(\psi) = \int_0^{2\pi} p_{A\psi}(A, \psi) dA = \int_0^{2\pi} \frac{A}{2\pi\sigma_x^2} e^{-\frac{A^2}{2\sigma_x^2}} dA = \frac{1}{2\pi},$$

что означает равномерное распределение фазы на всем интервале от 0 до 2π .

В радиотехнике большой интерес представляет анализ статистических свойств суммы детерминированного (например, гармонического) сигнала и узкополосного нормального (гауссовского) случайного сигнала. Именно этот случай имеет место на выходе некоторого частотно-избирательного устройства, например, резонансного усилителя.

Если детерминированную составляющую сигнала представить в виде

$$u(t) = U_m \cos \omega_0 t, \quad (4.45)$$

то суммарный сигнал может быть представлен ранее использованной формулой случайного узкополосного случайного сигнала, однако плотность вероятности для суммарного сигнала в этом случае принимает вид распределения, называемого законом распределения Релея-Райса.

$$p(A) = \frac{A}{\sigma_x^2} e^{-\frac{(U_m^2 + A^2)}{2\sigma_x^2}} I_0\left(\frac{AU_m}{\sigma_x^2}\right), \quad (4.46)$$

где $I_0(z) = \frac{1}{2\pi} \int_0^{2\pi} e^{z \cos \varphi} d\varphi$ - модифицированная функция Бесселя нулевого порядка.

На рис.4.6. представлены графики плотности вероятности случайной величины, распределенной по закону Райса (4.46) при различных соотношениях $\alpha = U_m/\sigma_x$

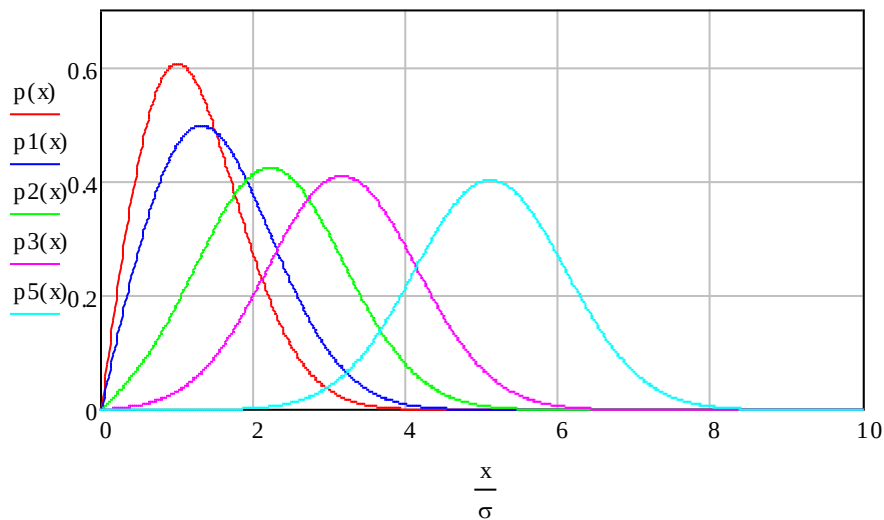


Рис.4.6. Плотность вероятности закона Релея-Райса

4.6. Модели случайных сигналов в дискретном времени

Ранее мы уже рассматривали вопросы представления непрерывных детерминированных сигналов в дискретном времени, то есть последовательностью выборочных значений. Такая же процедура дискретизации может быть применена и к случайным сигналам.

Рассмотрим дискретную последовательность $x(k)$, $k = 0, 1, 2, \dots, n, \dots$, получаемую из непрерывного случайного процесса $x(t)$ путем выборки значений через равные интервалы времени T_0 . Поскольку отдельные реализации случайного процесса являются детерминированными функциями, то все ограничения, вызванные процессом дискретизации и рассмотренные ранее для детерминированных сигналов, в данном случае также будут иметь место.

Для дискретных случайных сигналов справедливы все общие свойства, касающиеся законов распределения и плотностей распределения случайных величин. Сохраняются понятия стационарности и эргодичности случайных процессов, корреляционных функций и спектральных плотностей мощности.

Главное отличие от непрерывных процессов состоит в том, что время в дискретном случае не является непрерывным континуумом, а представляется целочисленной последовательностью.

Поэтому несколько изменяются и формулы для определения основных вероятностных характеристик этих процессов, в которых в отличие от случая

непрерывного времени вместо переменных t_1, t_2 дискретные отсчеты. Так, например, для плотности вероятности стационарного в широком смысле процесса форма записи будет выглядеть следующим образом

$$p_x(x_1, x_2, k_1, k_2) = p_x(x_1, x_2, \Delta k). \quad (4.47)$$

Для стационарного эргодического дискретного процесса (в дальнейшем мы будем рассматривать именно такие процессы) математическое ожидание может быть определено как

$$\bar{x} = M\{x(k)\} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N x(k). \quad (4.48)$$

Аналогично случаю непрерывного времени используется понятие центрированного случайного процесса, как процесса, из которого вычтено его математическое ожидание

$$x_0(k) = x(k) - \bar{x}. \quad (4.49)$$

Корреляционная функция для центрированного стационарного процесса может быть получена из выражения

$$R_x(m) = M\{x_0(k)x_0(k+m)\} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N x_0(k)x_0(k+m), \quad (4.50)$$

где m – принимает только целочисленные значения.

При $m = 0$ аналогично случаю непрерывного времени корреляционная функция принимает значение дисперсии случайного процесса

$$D_x = \sigma_x^2 = R_x(0) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N (x(k) - \bar{x})^2. \quad (4.51)$$

Введенные определения (4.47) – (4.51) аналогичны соответствующим определениям, ранее введенным для сигналов непрерывного времени.

В случае, когда значения сигнала в конкретный момент времени описывается нормальным законом распределения, то он полностью описывается своим математическим ожиданием и корреляционной функцией.

Спектральная плотность мощности для дискретного случайного сигнала вводится по аналогии со случаем непрерывного времени путем применения дискретного преобразования Фурье, которое мы рассматривали при анализе дискретных детерминированных сигналов.

В общем случае спектральная плотность мощности дискретного сигнала $x(k)$ может быть определена по формуле

$$S_{Px}(\omega) = T_0 \sum_{m=-\infty}^{\infty} R_x(m) e^{-j\omega m T_0}, \quad (4.52)$$

где T_0 – интервал временного квантования непрерывного сигнала.

Поскольку дискретные сигналы часто рассматриваются в относительном безразмерном времени, то нормирующий множитель T_0 в формуле для спектральной плотности мощности в (4.52) заменяют единицей аналогично спектрам дискретных детерминированных сигналов.

На практике количество доступных для наблюдения и обработки значений процесса $x(k)$ всегда ограничено. Для конечного набора из N наблюдений используют несколько вариантов оценки корреляционной функции, простейший из которых для вещественных сигналов выглядит следующим образом

$$\widehat{R}_x(m) = \frac{1}{N} \sum_{n=0}^{N-m-1} x(n+m)x(n). \quad (4.53)$$

Из приведенного уравнения видно, что при таком способе вычисления корреляционной функции (4.53) ее значения, соответствующие разным временным сдвигам, оцениваются с использованием различного количества отсчетов исходного сигнала. (Для устранения этого эффекта и предлагаются различные методы расчета).

Выборочная оценка спектральной плотности мощности в этом случае определяется как

$$\widehat{S}_{Px}(\omega) = \sum_{m=-(N-1)}^{N-1} \widehat{R}_x(m) e^{-j\omega m T_0} \sum_{m=-(N-1)}^{N-1} \left(\frac{N-|m|}{N}\right) R_x(m) e^{-j\omega m T_0}. \quad (4.54)$$

Переход от переменной ω к дискретным наборам частотной сетки (например, при применении БПФ) можно легко осуществить посредством следующей замены переменных в (4.54)

$$\omega_k = 2\pi f_k = \frac{2\pi k}{NT_0}. \quad (4.55)$$

В большинстве практических задач приходится рассматривать сигналы, представленные конечным набором N своих значений. В этом случае часто пользуются таким представлением характеристик сигнала, как корреляционная матрица, определяемая следующим выражением

$$\mathbf{R}_x = \begin{bmatrix} R_x(0) & R_x(1) & \dots & R_x(N-1) \\ R_x(-1) & R_x(0) & \dots & R_x(N-2) \\ \vdots & \vdots & \ddots & \vdots \\ R_x(-N+1) & R_x(-N+2) & \dots & R_x(0) \end{bmatrix}. \quad (4.56)$$

В случае вещественного сигнала ввиду симметрии четности корреляционной функции вид матрицы (4.56) существенно упрощается

$$\mathbf{R}_x = \begin{bmatrix} R_x(0) & R_x(1) & \dots & R_x(N-1) \\ R_x(-1) & R_x(0) & \dots & R_x(N-2) \\ \vdots & \vdots & \ddots & \vdots \\ R_x(N-1) & R_x(N-2) & \dots & R_x(0) \end{bmatrix}. \quad (4.57)$$

В дискретном случае аналогично непрерывному времени вводится идеализированный случайный сигнал, называемый дискретным белым шумом. Для описания этого сигнала используется функция Кронекера, являющаяся в определенном смысле аналогом импульсной функции Дирака, применяемой при описании сигналов непрерывного времени

$$\delta(k) = \begin{cases} 1, & \text{при } k = 0 \\ 0, & \text{при } k \neq 0 \end{cases} \quad (4.58)$$

Поскольку взаимосвязь между значениями белого шума в различающиеся моменты дискретного времени отсутствует, то его корреляционная функция записывается в виде

$$R_x(\tau) = \sigma_x^2 \delta(k).$$

Понятно, что корреляционная матрица белого шума будет иметь диагональную форму

$$\mathbf{R}_x = \mathbf{I} \sigma_x^2 \quad (4.59)$$

Важным типом случайных процессов являются так называемые марковские случайные процессы, обладающие особыми свойствами плотности условного распределения. Случайный процесс называют марковским процессом первого порядка, если его условная плотность вероятности подчиняется следующему соотношению

$$p(x(k)|x(k-1), x(k-2), \dots, x(0)) = p(x(k)|x(k-1)), \quad (4.60)$$

то есть очередное значение сигнала зависит только от предшествующего значения.

Данному определению марковского процесса (4.60) соответствует сигнал, описываемый скалярным разностным уравнением первого порядка

$$x(k) = ax(k-1) + bv(k), \quad (4.61)$$

где $a, b = \text{const.}$

$e(t)$ - белый шум.

В том случае, когда скалярное уравнение вида (4.61) имеет более высокий порядок, например, второй

$$x(k) = a_1x(k-1) + a_2x(k-2) + be(k),$$

его легко можно заменить векторным разностным уравнением вида

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ a_1 & a_2 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} + \begin{bmatrix} 0 \\ b \end{bmatrix} e(k),$$

где $x(k) = x_1(k)$,

$$x(k+1) = x_1(k+1) = x_2(k).$$

В общем случае для произвольного порядка уравнение марковского сигнала приводится к виду

$$x(k+1) = Ax(k) + Be(k), \quad (4.62)$$

где $x(k)$ – вектор размера $n \times 1$,

A – матрица размера $n \times n$,

B – вектор размера $n \times 1$.

Введение векторных случайных процессов требует определения их основных статистических характеристик.

По аналогии со скалярным случаем математическое ожидание векторного случайного процесса определяется как математическое ожидание каждой из его компонент

$$\bar{x} = M\{x(k)\} = \begin{bmatrix} \overline{x_1(k)} \\ \overline{x_2(k)} \\ \vdots \\ \overline{x_n(k)} \end{bmatrix}.$$

Взаимосвязь отдельных скалярных составляющих вектора $x(k)$ описывается соответствующей корреляционной матрицей

$$\begin{aligned} R_x(m) &= M\{[x(k) - \bar{x}][x(k+m) - \bar{x}]^T\} = \\ &= \begin{bmatrix} R_{x_1x_1}(m) & R_{x_1x_2}(m) & \dots & R_{x_1x_n}(m) \\ R_{x_2x_1}(m) & R_{x_2x_2}(m) & \dots & R_{x_2x_n}(m) \\ \vdots & \vdots & \ddots & \vdots \\ R_{x_nx_1}(m) & R_{x_nx_2}(m) & \dots & R_{x_nx_n}(m) \end{bmatrix}. \end{aligned}$$

Элементы, расположенные на главной диагонали приведенной матрицы представляют собой корреляционные функции компонент, а внедиагональные элементы – взаимные корреляционные функции компонент. (Не следует путать эту матричную форму корреляционной функции для векторного сигнала с ранее рассмотренной корреляционной матрицей для скалярного случайного сигнала).

Предельное увеличение порядка в приведенной марковской модели порождает следующую форму представления случайного процесса

$$x(k) = a_1x(k-1) + a_2x(k-2) + \dots + a_ix(k-i) + \dots + e(k) = \quad (4.63)$$

$$= \sum_{i=1}^{\infty} a_i x(k-i) + e(k).$$

Теоретически представление сигнала $x(k)$ из (4.64) возможно также с помощью другой бесконечной суммы

$$\begin{aligned} x(k) &= b_0 e(k) + b_1 e(k-1) + \dots + b_i e(k-i) + \dots = \\ &= \sum_{i=0}^{\infty} b_i e(k-i). \end{aligned} \quad (4.64)$$

Использование конечного числа в суммах данных бесконечных рядов (4.63), (4.64) приводят к трем наиболее распространенным моделям случайных процессов в виде временных рядов.

Авторегрессионная модель (АР-модель) имеет вид

$$x(k) = \sum_{i=1}^p a_i x(k-i) + e(k), \quad (4.65)$$

где p – порядок модели,

a_i – параметры модели.

Модель скользящего среднего (СС-модель) имеет вид

$$x(k) = \sum_{i=0}^q b_i e(k-i), \quad (4.66)$$

где q – порядок модели,

b_i – параметры модели.

Объединение АР и СС моделей (4.65) и (4.66) приводит к форме, получившей название авторегрессии скользящего среднего (АРСС), которая имеет следующий вид

$$x(k) = \sum_{i=1}^p a_i x(k-i) + \sum_{i=0}^q b_i e(k-i), \quad (4.67)$$

где p, q – порядки соответствующих частей модели,

a_i, b_i – параметры модели.

Все три вида моделей (4.65), (4.66), (4.67) широко используются при моделировании и анализе реальных случайных сигналов. Более подробно вопросы их использования мы рассмотрим в соответствующем разделе нашего курса.

Пример 4.1.

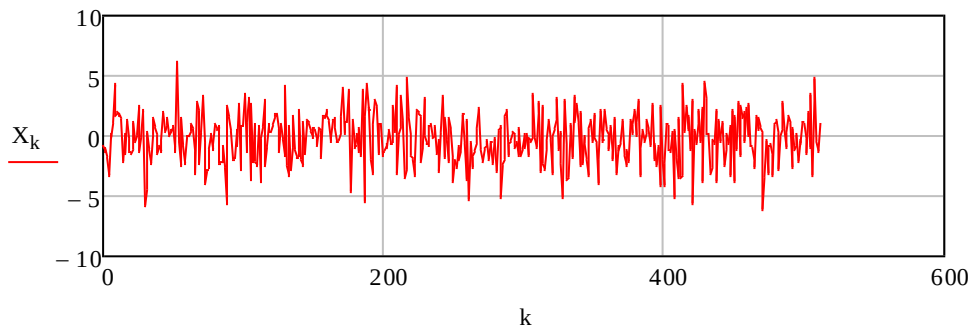
Продemonстрируем основные этапы определения корреляционной функции и спектральной плотности для дискретных сигналов с использованием программы MathCad.

1. Сформируем дискретный белый шум с нормальным законом распределения плотности вероятности.

```

m:=0
σ:=2
N:=512
X:=morm(N,m,σ)
k:=0..N-1

```

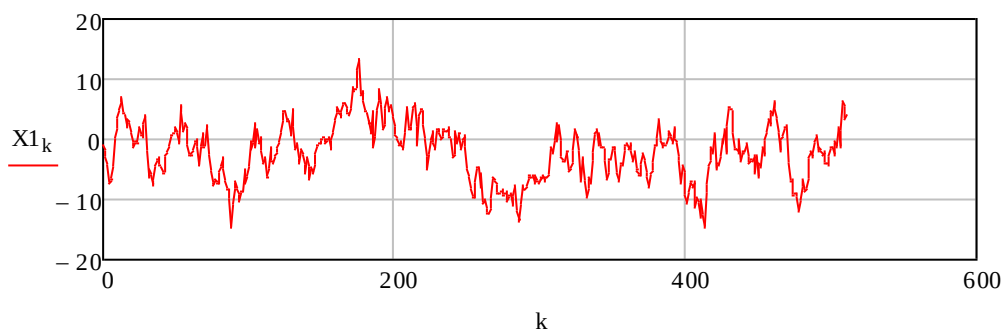


2. На основе данного сигнала сформируем второй случайный сигнал, связанный с исходным разностным уравнением первого порядка

```

a := 0.9
CV(X,a,N) :=
  Y_0 ← X_0
  for k ∈ 1..N - 1
    Y_k ← a · Y_{k-1} + X_k
  Y
X1 := CV(X,a,N)

```

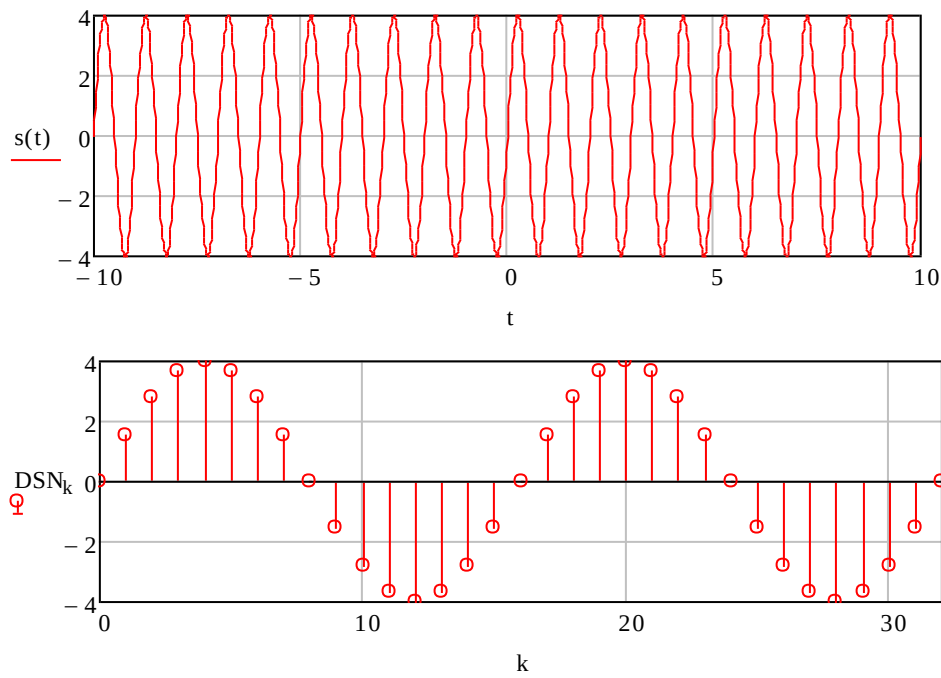


3. Сформируем гармонический сигнал и произведем дискретную выборку его наблюдений

```

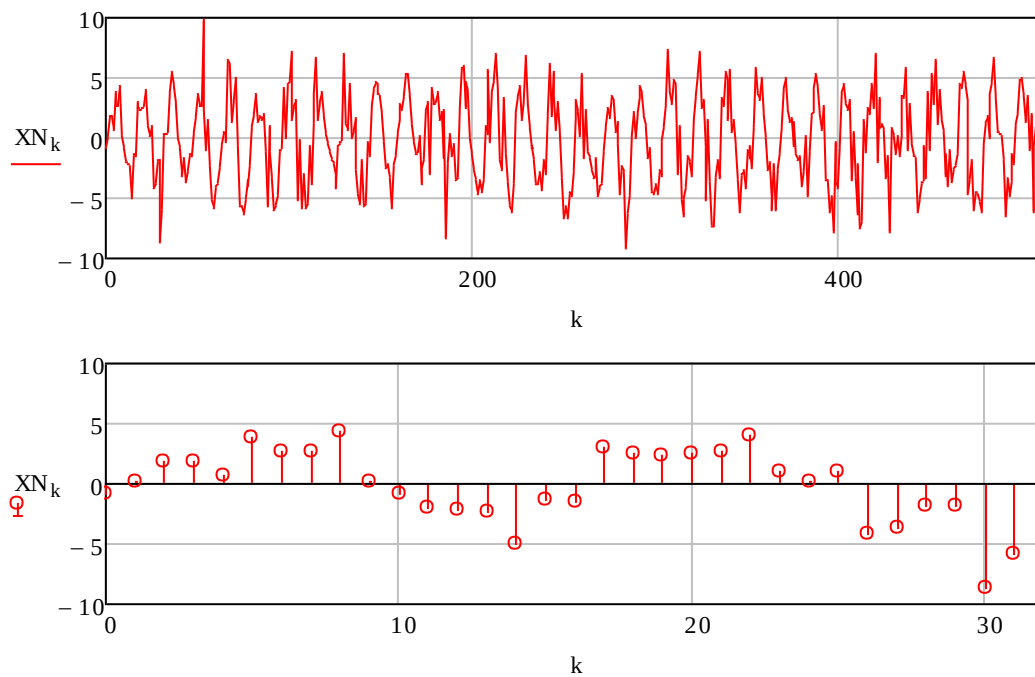
A := 4
f := 1
f0 := 1e
s(t) := A · sin(2π · f t)
T0 := 1/f0
k := 0..N - 1
n_k := T0 · k
DSN_k := s(n_k)

```



4. Суммируя ранее полученный случайный сигнал с гармоническим формируем суммарный сигнал, характеристики которого будем исследовать

$$XN_k := X_k + DSN_k$$



5. Произведем оценку корреляционных функций для всех сформированных сигналов на основе имеющегося конечного ряда из 512 наблюдений и построим графики найденных корреляционных функций.

```

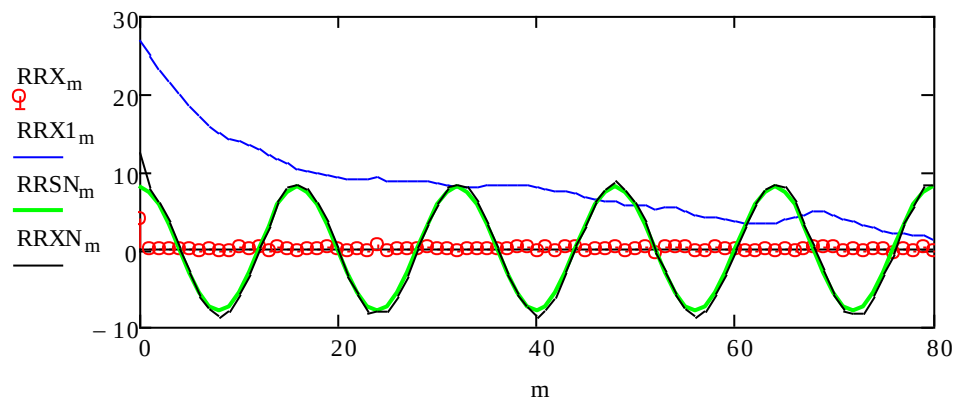
R(X,N,M) :=
  for k ∈ 0..M
  G_k ← 1 / (N - k) · ∑_{m=0}^{N-1-k} (X_m · X_{m+k})
  G

```

```

M := N - 1
RRX := R(X,N,M)
RRX1 := R(X1,N,M)
RRSN := R(DSN,N,M)
RRXN := R(XN,N,M)
SX := FFT(RRX)
SX1 := FFT(RRX1)
SSN := FFT(RRSN)
SXN := FFT(RRXN)
m := 0..80
n := 0..N/2

```

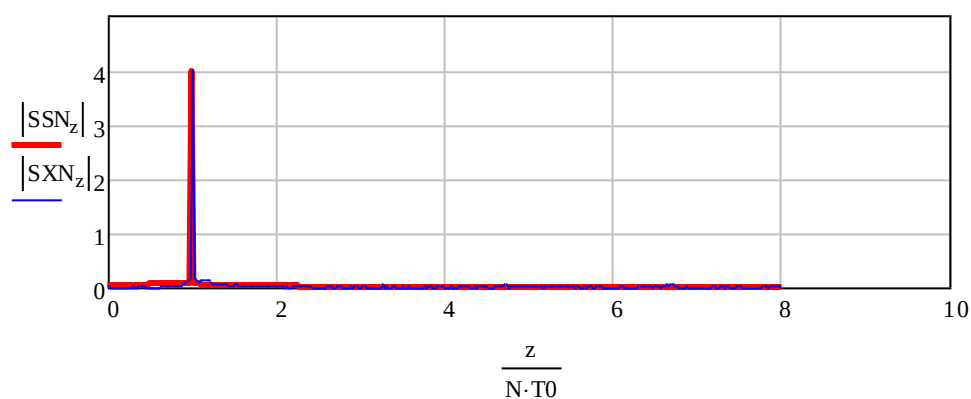
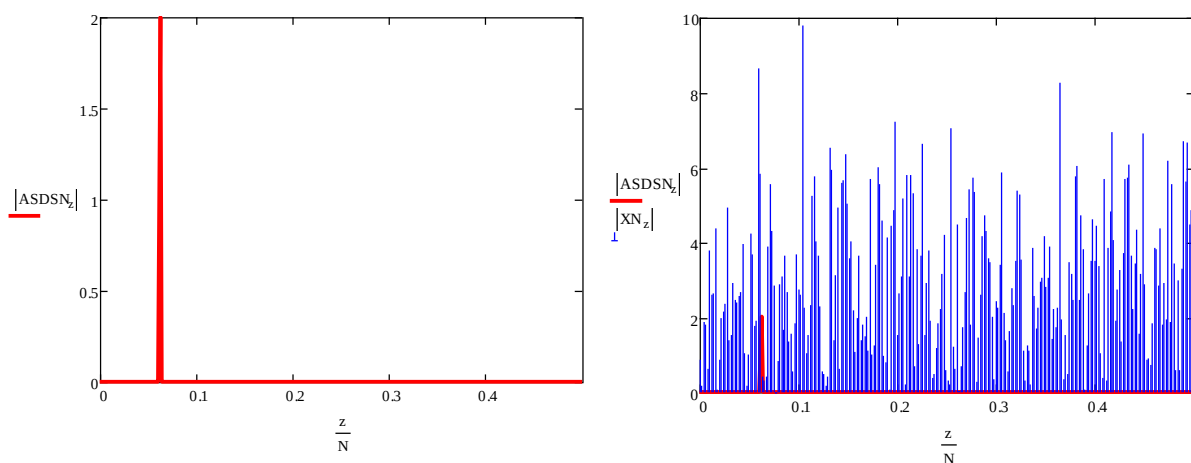


6. Теперь продемонстрируем разницу результатов, получающихся при использовании спектральной плотности мощности и попытки вычисления обычного амплитудного спектра случайного сигнала.

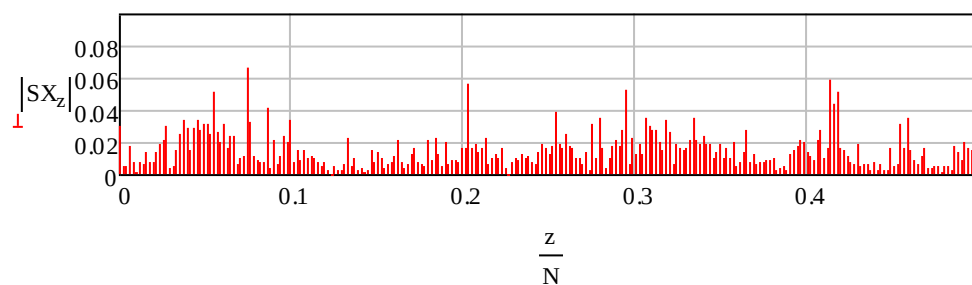
```

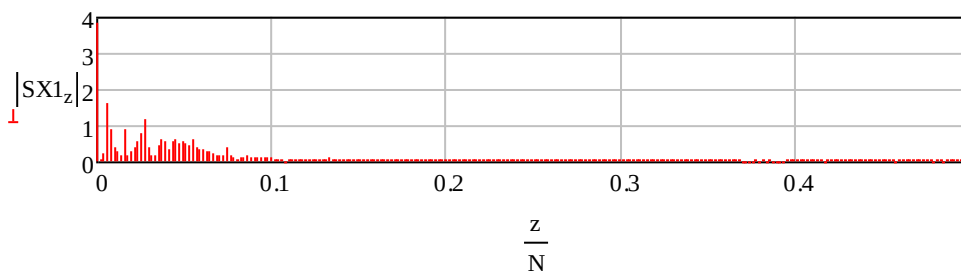
ASDSN := FFT(DSN)
ASXN := FFT(XN)

```



7. В результате нам достаточно точно удалось оценить частоту и мощность гармонического сигнала, наблюдаемого на фоне случайной помехи.
8. В заключение оценим качество использованных при моделировании случайных сигналов, сравнив оценки их спектральной плотности мощности

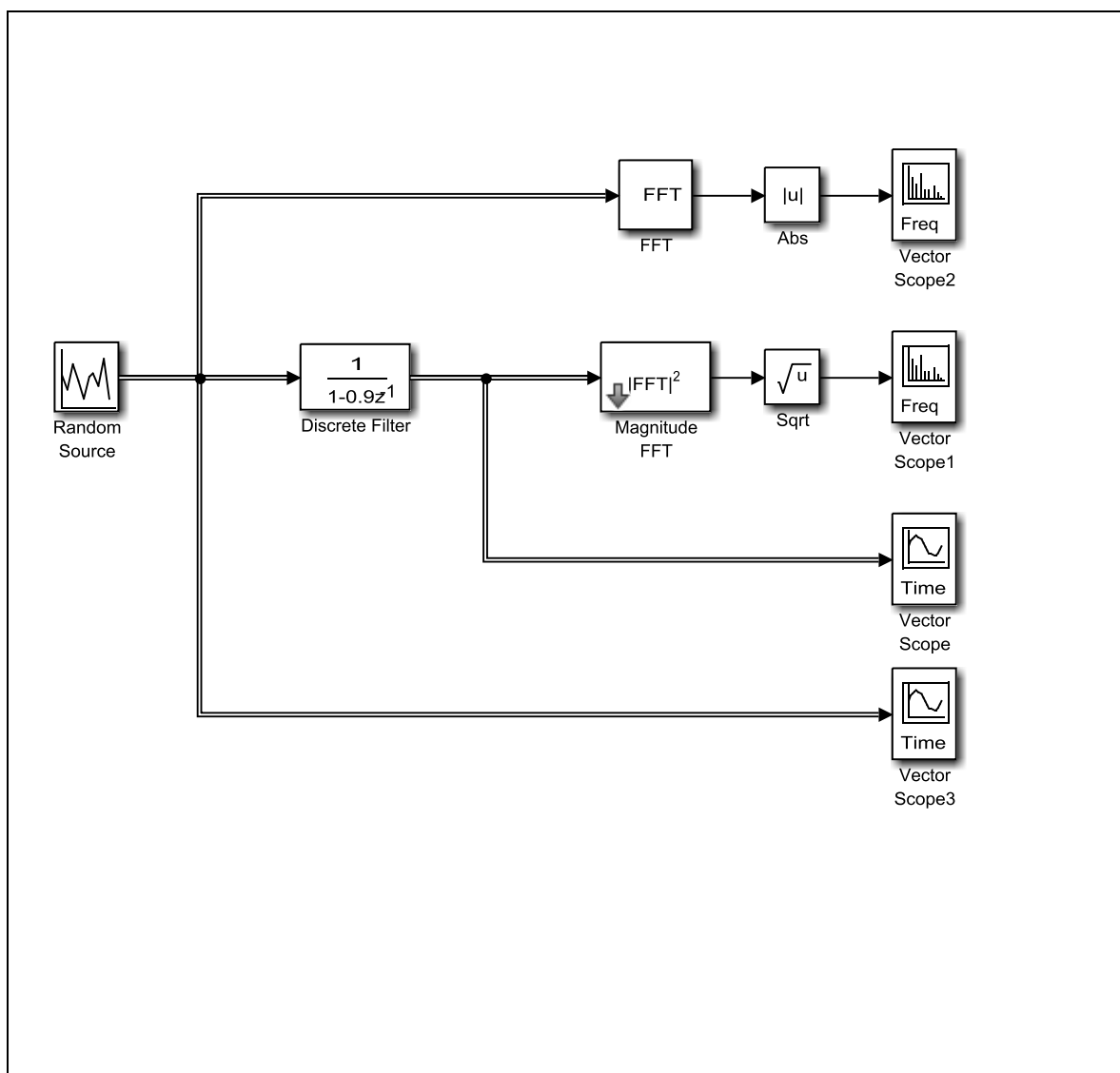




Полученные спектральные оценки, естественно, согласуются с ранее вычисленными корреляционными функциями этих шумов и достаточно хорошо аппроксимируют теоритические ожидаемые результаты.

Пример 4.2.

На рис. 4.7. представлена схема формирования случайного процесса из белого шума на основе авторегрессионной модели первого порядка с использованием программы Simulink (Matlab). На этом же рисунке приведены полученные в результате моделирования оценки спектральной плотности мощности исходного белого шума и сформированного АР процесса.



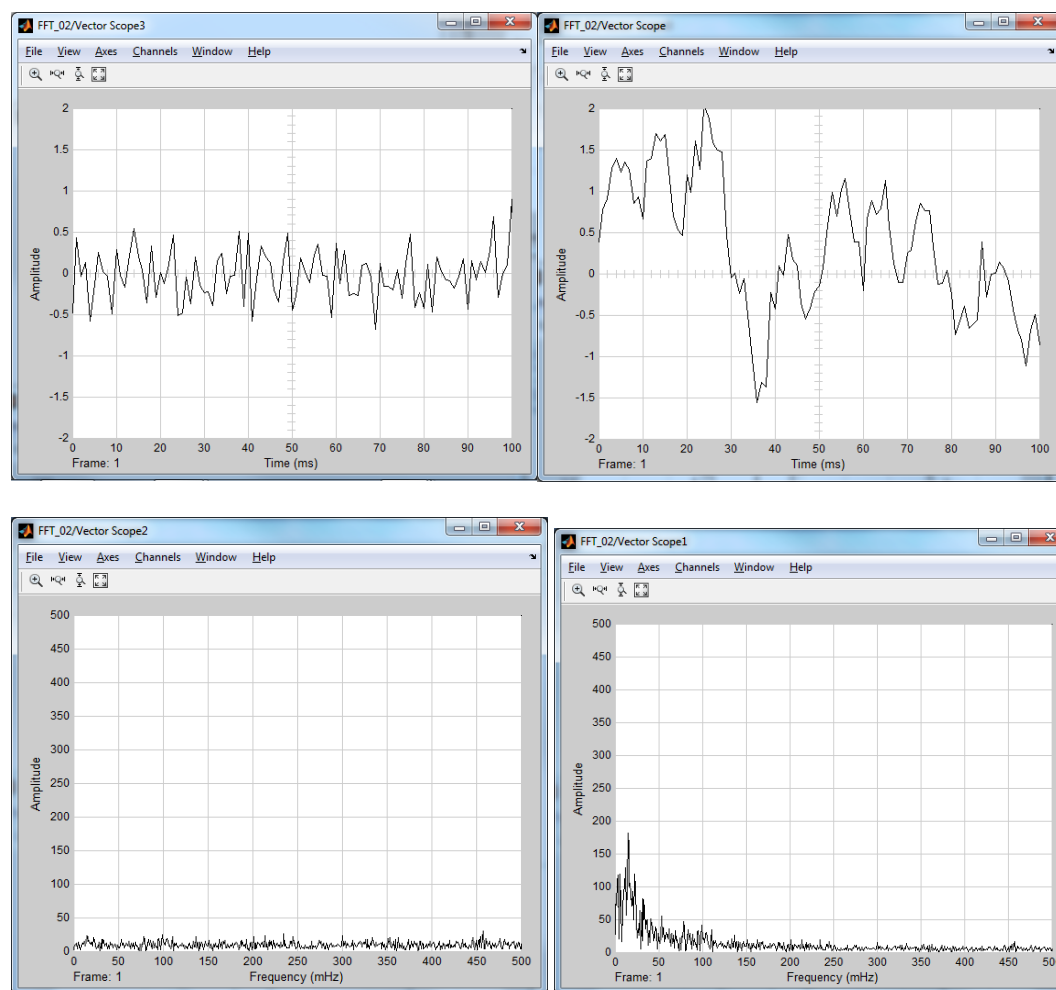


Рис.4.7. Формирование случайного сигнала на основе авторегрессионной модели

4.7. Вопросы для самоконтроля и повторения

1. Дайте определение стохастического процесса.
2. Что такое сечение случайного процесса?
3. Как связаны между собой функция распределения вероятности случайного процесса и его плотность вероятности?
4. Дайте определение математическому ожиданию случайного процесса.
5. Дайте определение дисперсии случайного процесса.
6. Что такое мода и медиана случайного процесса?
7. Дайте определение ковариационной функции случайного процесса.
8. Какой случайный процесс называется центрированным?
9. Какова связь корреляционной и ковариационной функций случайного процесса?

10. Дайте определение взаимной корреляционной функции двух случайных процессов.
11. Дайте определение стационарного случайного процесса.
12. Поясните разницу между стационарностью в широком и узком смысле.
13. Какой процесс называется эргодическим?
14. Как определяется спектральная плотность мощности случайного процесса?
15. Какова связь корреляционной функции и спектральной плотности мощности?
16. Что такое эффективная ширина спектра?
17. Что представляет собой случайный узкополосный сигнал?
18. Какой процесс называется марковским?
19. Запишите уравнение авторегрессионной модели случайного процесса.
20. Запишите уравнение модели скользящего среднего.
21. Запишите уравнение модели авторегрессии скользящего среднего.

ГЛАВА 5. Радиотехнические цепи с постоянными параметрами

5.1. Двухполюсники и четырехполюсники как базовые структурные элементы радиотехнических цепей

Радиотехнические цепи формируются путем соответствующего соединения отдельных элементов. Эти элементы и их соединения в простейшем случае имеют два подключаемых контакта, которые принято называть зажимами. Рассмотрение элемента цепи с точки зрения двух зажимов относит его к категории так называемых двухполюсников. В рамках такого рассмотрения принято анализировать соотношение тока, текущего в данном двухполюснике, в зависимости от напряжения, приложенного к его зажимам.

Если анализируемую цепь рассматривать как объект, имеющий большее количество зажимов, то ее называют **многополюсником**. Наиболее часто модели цепей и устройств представляют в виде четырехполюсников.

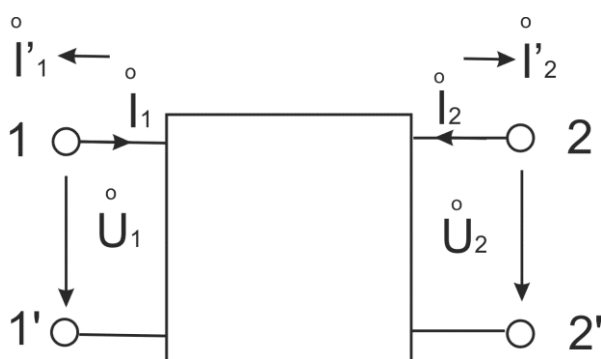


Рис.5.1. Обобщенная схема четырехполюсника.

Четырехполюсником (ЧП) называется участок электрической цепи (или устройства), рассматриваемый со стороны 4^х зажимов, два из которых являются входными, а два других – выходными. Общий вид четырехполюсника с обозначениями комплексных амплитуд токов и напряжений на зажимах представлен на рис. 5.1.

Двухполюсник тоже может быть рассмотрен с позиций четырех зажимов, как одноэлементный четырехполюсник, и вариантов такого рассмотрения существует два. Они оба представлены на рис. 5.2.

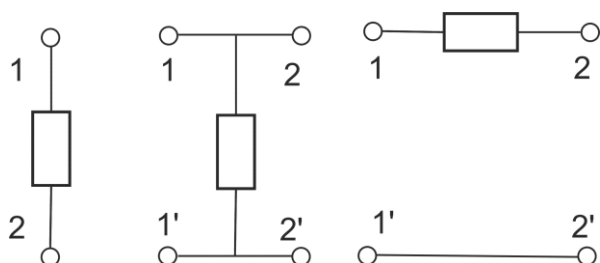


Рис.5.2. Одноэлементные двухполюсники и четырехполюсники.

ЧП могут быть подразделены на активные и пассивные.

Пассивные ЧП не содержат идеализированных активных элементов.

Активные ЧП включают в себя управляемые или неуправляемые идеализированные источники энергии.

Если выходы и входы ЧП, содержащего неуправляемые источники энергии, соединить между собой (закоротить), то хотя бы один из (входных или выходных) токов

будет отличен от нуля. Если входы и выходы такого ЧП остановить в режиме холостого хода, то напряжение между выходами или входами, не будет равняться нулю. Четырехполюсники, удовлетворяющие таким условиям, называются **автономными**.

Если в состав ЧП входят только идеализированные пассивные элементы или идеализированные пассивные элементы и управляемые источники, то токи короткого замыкания и напряжения холостого хода равны нулю. Такие ЧП называют **неавтономными**.

Таким образом, к неавтономным ЧП относятся все пассивные ЧП и те активные ЧП, которые не содержат неуправляемых источников тока или напряжения.

Кроме того ЧП могут быть подразделены на **взаимные (обратимые)** и **невзаимные (необратимые)**.

Для взаимного ЧП справедливо следующее:

- если независимый источник тока подключить к одной паре зажимов ЧП, то на другой паре зажимов возникает напряжение U ; этот же источник, подключенный ко второй паре зажимов, вызовет появление напряжения U на первой паре зажимов.
- если независимый источник напряжения подключить к первой паре зажимов, то он вызовет ток короткого замыкания I на второй паре, этот же источник, подключенный ко второй паре зажимов, вызовет ток короткого замыкания I на первой паре.

ЧП называется **симметричным**, если с помощью внешних изменений невозможно установить различие между входными и выходными зажимами, т.е. токи и напряжения цепи, к которой подключен симметричный ЧП, не изменятся, если пары зажимов 1-1' и 2-2' поменять местами.

В теории ЧП рассматривается только токи и напряжения на входах и выходах, а все реакции внутри ЧП остаются без внимания.

Токи и напряжения на входе и выходе ЧП могут быть связаны между собой системой уравнений. Вид этих уравнений зависит от того, какие две из четырех величин выбираются в виде независимых переменных. Так как $C_4^2 = 6$, то всего возможно 6 различных способов записи систем уравнений.

Рассмотрим некоторые из них.

Представление модели четырехполюсника в виде

$$\begin{cases} U_1^o = Z_{11}I_1^o + Z_{12}I_2^o \\ U_2^o = Z_{21}I_1^o + Z_{22}I_2^o \end{cases} \quad (5.1)$$

называют Z-формой.

Представление модели четырехполюсника в виде

$$\begin{cases} I_1^o = Y_{11}U_1^o + Y_{12}U_2^o \\ I_2^o = Y_{21}U_1^o + Y_{22}U_2^o \end{cases} \quad (5.2)$$

называют Y-формой.

Представление модели четырехполюсника в виде

$$\begin{cases} U_1^o = A_{11}U_2^o + A_{12}I_2^{o'} \\ I_1^o = A_{21}U_2^o + A_{22}I_2^{o'} \end{cases} \quad (5.3)$$

называют A-формой, которая часто встречается в двух отдельных вариантах

$$\begin{cases} U_1^o = AU_2^o + BI_2^{o'} \\ I_1^o = CU_2^o + DI_2^{o'} \end{cases} \quad (5.4)$$

A-форма при прямом включении и

$$\begin{cases} U_2^\circ = AU_1^\circ + BI_1^{\circ'} \\ I_2^\circ = CU_1^\circ + DI_1^{\circ'} \end{cases} \quad (5.5)$$

А-форма при обратном включении.

Коэффициенты ЧП, входящие в уравнения (5.1) –(5.5) для каждого данного ЧП суть величины, зависящие только от частоты.

Если ЧП является взаимным, то из 4^х коэффициентов 3 являются независимыми, а 4-ый находится из уравнения связи

$$Z_{12} = Z_{21} \text{ — для } Z \text{ — формы,} \quad (5.6a)$$

$$Y_{12} = Y_{21} \text{ — для } Y \text{ — формы,} \quad (5.6b)$$

$$AD - BC = 1 \text{ — для } A \text{ — формы.} \quad (5.6c)$$

Если четырехполюсник является симметричным, то кроме того справедливы следующие соотношения

$$Z_{11} = Z_{22}, \quad Y_{11} = Y_{22}, \quad A = D. \quad (5.7)$$

Физический смысл некоторых из коэффициентов четырехполюсника легко определить из приведенных выше систем уравнений (5.1) –(5.7).

Так, например,

$$Z_{11} = \left. \frac{U_1^\circ}{I_1^\circ} \right|_{I_2^\circ=0} = Z_{\text{вх } \text{хх}},$$

то есть входное сопротивление на зажимах 1-1' в режиме холостого хода на зажимах 2-2'.

$$Y_{11} = \left. \frac{I_1^\circ}{U_1^\circ} \right|_{U_2^\circ=0} = Y_{\text{вх } \text{кз}},$$

то есть входная проводимость на зажимах 1-1' в режиме короткого замыкания на зажимах 2-2'. При этом очевидно, что $Z_{11} \neq 1/Y_{11}$.

Аналогично для А-формы можно определить смысл коэффициента А

$$A = \left. \frac{U_1^\circ}{U_2^\circ} \right|_{I_2^\circ=0} = \frac{1}{K_{\text{ухх}}},$$

где $K_{\text{ухх}}$ – коэффициент передачи по напряжению (в режиме холостого хода).

Коэффициенты четырехполюсника легко определить экспериментально.

Первое измерение проводится в режиме холостого хода ($I_2^{\circ'} = 0$), для которого из системы уравнений

$$\begin{cases} U_1^\circ = AU_2^\circ + BI_2^{\circ'} \\ I_1^\circ = CU_2^\circ + DI_2^{\circ'} \end{cases}$$

следует, что

$$Z_{1\text{вх хх}} = Z_{1\text{х}} = \frac{U_1^\circ}{I_1^\circ} = \frac{A}{C}.$$

Второе измерение проводится при коротком замыкании ($U_2^\circ = 0$), аналогично получая

$$Z_{1\text{вх кз}} = Z_{1\text{к}} = \frac{U_1^\circ}{I_1^\circ} = \frac{B}{D}.$$

Следующие два измерения проводятся при обратном включении четырехполюсника

$$\begin{cases} U_2^\circ = AU_1^\circ + BI_1^{\circ'} \\ I_2^\circ = CU_1^\circ + DI_1^{\circ'} \end{cases}$$

откуда аналогично предыдущему получают уравнения для входных сопротивлений при холостом ходе и коротком замыкании

$$Z_{2\text{вх хх}} = Z_{2\text{х}} = \frac{U_2^\circ}{I_2^\circ} = \frac{D}{C}$$

$$Z_{2\text{вх кз}} = Z_{2\text{к}} = \frac{B}{A}.$$

Поскольку между коэффициентами четырехполюсника существует связь $AD - BC = 1$, то достаточно взять любые 3 из 4-х полученных в ходе опытов уравнений и определить коэффициенты. Например,

$$\begin{aligned} Z_{1\text{х}} - Z_{1\text{к}} &= \frac{A}{C} - \frac{B}{D} = \frac{1}{CD}, \\ \frac{Z_{2\text{х}}}{Z_{1\text{х}} - Z_{1\text{к}}} &= \frac{\frac{D}{C}}{\frac{1}{CD}} = D^2. \end{aligned}$$

Дальнейшие вычисления очевидны.

Если ЧП симметричный ($A = D$), то для экспериментального определения его коэффициентов требуется только два опыта.

Так как соотношения между токами и напряжениями любого неавтономного линейного ЧП задаются системой из 2^x уравнений, содержащей не более 4^x независимых параметров, то такой ЧП может быть представлен идеализированной схемой, имеющей не более 4^x элементов. Для взаимных ЧП соответственно 3 коэффициента – 3 элемента. Таких схем существует две: Т-образная и П-образная. Вид канонических схем замещения неавтономных четырехполюсников представлен на рис.5.3.

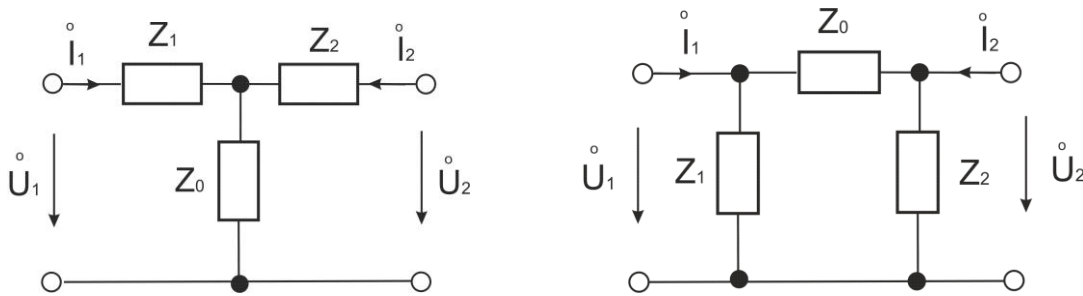


Рис.5.3. Канонические схемы замещения четырехполюсника.

Пример 5.1.

Получим уравнения, связывающие между собой коэффициенты четырехполюсника и параметры канонической Т-образной схемы замещения Z_0, Z_1, Z_2 .

Применяя к схеме замещения правила Кирхгофа получаем следующие уравнения

$$\begin{cases} I_0^\circ = I_1^\circ + I_2^\circ \\ U_1^\circ = I_1^\circ Z_1 + I_0^\circ Z_0 \\ U_2^\circ = I_2^\circ Z_2 + I_0^\circ Z_0 \end{cases}$$

откуда

$$\begin{aligned} U_1^\circ &= I_1^\circ Z_1 + U_2^\circ - I_2^\circ Z_2, \\ I_1^\circ &= -I_2^\circ + U_2^\circ \frac{1}{Z_0} - I_2^\circ \frac{Z_2}{Z_0} = \frac{1}{Z_0} U_2^\circ - \left(1 + \frac{Z_2}{Z_0}\right) I_2^\circ. \end{aligned}$$

Принимая во внимание, что $I_2^\circ = -I_2'^\circ$, получаем

$$C = \frac{1}{Z_0}, \quad D = 1 + \frac{Z_2}{Z_0}.$$

С другой стороны

$$U^{\circ}_1 = I^{\circ}_1 Z_1 + U^{\circ}_2 - I^{\circ}_2 Z_2 = \left(\frac{1}{Z_0} U^{\circ}_2 - \left(1 + \frac{Z_2}{Z_0} \right) I^{\circ}_2 \right) Z_1 + U^{\circ}_2 - I^{\circ}_2 Z_2,$$

что после несложных преобразований дает

$$U^{\circ}_1 = \left(1 + \frac{Z_1}{Z_0} \right) U^{\circ}_2 - \left(Z_1 + Z_2 + \frac{Z_1 Z_2}{Z_0} \right) I^{\circ}_2.$$

Следовательно,

$$A = 1 + \frac{Z_1}{Z_0}, \quad B = Z_1 + Z_2 + \frac{Z_1 Z_2}{Z_0}.$$

Для симметричного четырехполюсника

$$A = B \rightarrow Z_1 = Z_2.$$

Если известны параметры четырехполюсника A, B, C, D , то можно найти соответствующие комплексные сопротивления Z_0, Z_1, Z_2 канонической Т-образной схемы замещения по следующим уравнениям:

$$Z_0 = \frac{1}{C}, \quad Z_1 = \frac{A-1}{C}, \quad Z_2 = \frac{D-1}{C}.$$

Задание 5.1.

Самостоятельно для П-образной схемы замещений четырехполюсника проверить справедливость следующих выражений

$$Z_0 = B, \quad Y_1 = \frac{D-1}{B}, \quad Y_2 = \frac{A-1}{B}.$$

и

$$A = 1 + \frac{Z_0}{Z_2}, \quad B = Z_0, \quad C = \frac{1}{Z_1} + \frac{1}{Z_2} + \frac{Z_0}{Z_1 Z_2}, \quad D = 1 + \frac{Z_0}{Z_1}$$

5.2.Преобразование Лапласа и его основные свойства

Рассмотрим основные свойства преобразования Лапласа, которое будет использоваться нами при изучении радиотехнических цепей с позиций линейных динамических систем, то есть систем, процессы в которых описываются линейными дифференциальными уравнениями.

Преобразованием Лапласа называют соотношение

$$X(s) = \int_0^{\infty} x(t) e^{-st} dt, \quad (5.8)$$

которое ставит в соответствие функции $x(t)$ вещественного аргумента функцию $X(s)$ комплексной переменной $s = \sigma + j\omega$.

При этом вводятся следующие наименования:

$x(t)$ - оригинал,

$X(s)$ - изображение (или изображение по Лапласу).

Краткая форма связи (или соответствия) оригинала и соответствия (5.8) записывается как

$$x(t) \Leftrightarrow X(s) \quad \text{или} \quad X(s) \Leftrightarrow x(t).$$

Иногда для этих же целей используется символическая запись

$$X(s) = \mathcal{L}\{x(t)\},$$

где $\mathcal{L}$ - оператор Лапласа.

Функция $x(t)$, подвергаясь преобразованию Лапласа должна удовлетворять ряду свойств, а именно:

1. $x(t)$ определена и кусочно-дифференцируема по всей положительной полуоси $t \in [0, \infty)$,
2. $x(t) = 0$, при $t < 0$,
3. $\exists M, c > 0 : |x(t)| \leq Me^{ct}$ при $0 \leq t \leq \infty$.

Обратный переход от изображения $X(s)$ к функции-оригиналу $x(t)$ осуществляется в соответствии с уравнением

$$x(t) = \frac{1}{2\pi j} \int_{\sigma_0 - j\infty}^{\sigma_0 + j\infty} X(s)e^{st} ds, \quad (5.9)$$

называемым обратным преобразованием Лапласа. Интеграл в обратном преобразовании Лапласа берется вдоль любой прямой $Re(s) = \sigma_0 > c$.

Символическая запись выражения (5.9) имеет вид

$$x(t) = \mathcal{L}^{-1}\{X(s)\},$$

где $\mathcal{L}^{-1}$ - обратный оператор Лапласа.

Рассмотрим примеры изображений Лапласа для некоторых, часто встречающихся в радиотехнике сигналов.

- | | |
|-----------------------------|---|
| 1. $x(t) = a = const,$ | $X(s) = \mathcal{L}\{x(t)\} = a/s,$ |
| 2. $x(t) = e^{-at},$ | $X(s) = \mathcal{L}\{x(t)\} = 1/(s + a),$ |
| 3. $x(t) = e^{-j\omega t},$ | $X(s) = \mathcal{L}\{x(t)\} = 1/(s + j\omega),$ |
| 4. $x(t) = \sin(\omega t),$ | $X(s) = \mathcal{L}\{x(t)\} = \omega/(s^2 + \omega^2),$ |
| 5. $x(t) = \cos(\omega t),$ | $X(s) = \mathcal{L}\{x(t)\} = s/(s^2 + \omega^2)$ |

Основные свойства преобразования Лапласа.

1. Свойство линейности.

Для любых скаляров $\alpha, \beta = const$ преобразование Лапласа от линейной комбинации функций-оригиналов равной такой же линейной комбинации их преобразований Лапласа

$$\mathcal{L}\{\alpha x_1(t) + \beta x_2(t)\} = \alpha \mathcal{L}\{x_1(t)\} + \beta \mathcal{L}\{x_2(t)\}. \quad (5.10)$$

2. Дифференцирование оригинала

$$\mathcal{L}\left\{\frac{d^n}{dt^n}x(t)\right\} = s^n X(s) - s^{n-1}x(0) - s^{n-2}\frac{d}{dt}x(0) - \dots - \frac{d^{n-1}}{dt^{n-1}}x(0), \quad (5.11)$$

где

$$\frac{d^k}{dt^k}x(0) = \lim_{t \rightarrow +0} \frac{d^k}{dt^k}x(t), \text{ при } k = 0, 1, 2, \dots, n-1.$$

При нулевых начальных условиях данное выражение существенно упрощается

$$\mathcal{L}\left\{\frac{d^n}{dt^n}x(t)\right\} = s^n X(s),$$

то есть n -кратное дифференцирование оригинала в данном случае соответствует умножению изображения на s^n .

3. Интегрирование оригинала

$$\mathcal{L}\left\{\int_0^t x(t)dt\right\} = \frac{X(s)}{s}. \quad (5.12)$$

Таким образом, интегрирование функции-оригинала соответствует делению изображения на s .

4. Теорема о запаздывании.

Для любого положительного числа τ справедливо следующее

$$\mathcal{L}\{x(t - \tau)\} = e^{-s\tau} \mathcal{L}\{x(t)\} = e^{-s\tau} X(s). \quad (5.13)$$

5. Теорема о свертке (теорема об умножении изображений)

Если заданы сигналы $x_1(t)$ и $x_2(t)$, которым соответствуют изображения

$$X_1(s) = \mathcal{L}\{x_1(t)\}, \quad X_2(s) = \mathcal{L}\{x_2(t)\},$$

то свертке этих сигналов во временной области соответствует произведение их изображений, то есть

$$\begin{aligned} X_1(s)X_2(s) &\Leftrightarrow x_1(t) * x_2(t) \\ X_1(s)X_2(s) &\Leftrightarrow \int_0^t x_1(\tau) x_2(t - \tau) d\tau = \int_0^t x_1(t - \tau) x_2(\tau) d\tau. \end{aligned} \quad (5.14)$$

6. Теорема о предельных значениях

Если функция $x(t)$ является оригиналом для изображения $X(s)$, то

$$x(0) = \lim_{s \rightarrow \infty} sX(s) \quad (5.15)$$

и при существовании предела $x(\infty) = \lim_{t \rightarrow \infty} x(t)$

$$x(\infty) = \lim_{s \rightarrow 0} sX(s). \quad (5.16)$$

7. Теорема разложения.

Если изображение $X(s)$ является дробно-рациональной функцией, то есть

$$X(s) = \frac{A(s)}{B(s)},$$

причем степень $A(s)$ меньше степени $B(s)$, то оригиналом является умноженная на $\mathbf{1}(t)$ функция

$$x(t) = \sum_{k=1}^l \frac{1}{(n_k - 1)!} \lim_{s \rightarrow s_k} \frac{d^{n_k - 1}}{ds^{n_k - 1}} [X(s)(s - s_k)^{n_k} e^{st}], \quad (5.17)$$

где s_k - корни уравнения $B(s) = 0$,

k - кратность корней,

l - число различных корней.

Если все корни простые, то формула разложения существенно упрощается и принимает вид

$$x(t) = \sum_{k=1}^n \frac{A(s_k)}{B'(s_k)} e^{s_k t}, \quad (5.18)$$

где n - степень полинома $B(s)$,

$$B'(s_k) = \left. \frac{d}{ds} B(s) \right|_{s = s_k}.$$

Рассмотрим применение свойств преобразования Лапласа (5.10) – (5.18) на простых примерах.

Пример 5.2

Докажем справедливость уравнения теоремы разложения для случая простых корней. Если все корни знаменателя изображения Лапласа являются простыми, то

$$X(s) = \frac{A(s)}{B(s)} = \sum_{k=1}^n \frac{c_k}{s - s_k},$$

где $c_k = \text{const}$,

s_k – корни уравнения $B(s) = 0$.

Представим $X(s)$ в виде

$$X(s) = \frac{A(s)}{B(s)} = \frac{c_1}{s - s_1} + \sum_{k=2}^n \frac{c_k}{s - s_k},$$

тогда

$$\lim_{s \rightarrow s_1} \frac{A(s)}{B(s)} (s - s_1) = c_1 + \lim_{s \rightarrow s_k} \sum_{k=2}^n \frac{c_k}{s - s_k} (s - s_1) = c_1.$$

$$c_1 = \lim_{s \rightarrow s_k} \frac{A(s) \frac{d}{ds} (s - s_1)}{\frac{d}{ds} B(s)} = \frac{A(s_1)}{B'(s_1)}$$

Пример 5.3.

Найдем изображение Лапласа для функции, заданной следующим выражением

$$x(t) = 1, t \geq 0.$$

В соответствии с определением преобразования Лапласа запишем

$$X(s) = \int_0^{\infty} x(t) e^{-st} dt = \int_0^{\infty} 1 e^{-st} dt = -\frac{1}{s} e^{-st} \Big|_0^{\infty} = \frac{1}{s}.$$

Пример 5.4.

Найдем изображение Лапласа для функции, заданной следующим выражением

$$x(t) = e^{-at}, t \geq 0.$$

В соответствии с определением преобразования Лапласа запишем

$$X(s) = \int_0^{\infty} x(t) e^{-st} dt = \int_0^{\infty} e^{-at} e^{-st} dt = -\frac{1}{s+a} e^{-(s+a)t} \Big|_0^{\infty} = \frac{1}{s+a}.$$

Пример 5.5.

Найдем изображение Лапласа для функции, заданной следующим выражением

$$x(t) = \begin{cases} 1, & 0 \leq t \leq T \\ 0, & t > T \end{cases}$$

В соответствии с определением преобразования Лапласа запишем

$$X(s) = \int_0^{\infty} x(t) e^{-st} dt = \int_0^T 1 e^{-st} dt = \int_0^T e^{-st} dt = -\frac{1}{s} e^{-st} \Big|_0^T = \frac{1}{s} (1 - e^{-sT}).$$

Пример 5.6.

Найдем изображение Лапласа для функции, заданной следующим выражением

$$x(t) = \sin \omega t, t \geq 0$$

В соответствии с определением преобразования Лапласа запишем

$$X(s) = \int_0^{\infty} x(t) e^{-st} dt = \int_0^T \frac{1}{2j} (e^{j\omega t} + e^{-j\omega t}) dt = \frac{1}{2j} \left(\frac{1}{s - j\omega} + \frac{1}{s + j\omega} \right) = \frac{\omega}{s^2 + \omega^2}.$$

5.3. Радиотехнические цепи как линейные системы.

5.3.1. Передаточные функции

Для описания линейных систем широко используются линейные дифференциальные уравнения. В общем виде, если на вход линейной системы с постоянными во времени параметрами поступает входной сигнал $u(t)$, и при этом на выходе системы наблюдается выходной сигнал $y(t)$, то связь между ними описывается линейным дифференциальным уравнением вида

$$\begin{aligned} a_0 \frac{d^n}{dt^n} y(t) + a_1 \frac{d^{n-1}}{dt^{n-1}} y(t) + \dots + a_n y(t) = \\ = b_0 \frac{d^m}{dt^m} u(t) + b_1 \frac{d^{m-1}}{dt^{m-1}} u(t) + \dots + b_m u(t). \end{aligned} \quad (5.19)$$

Введя для операции дифференцирования обозначение

$$p \equiv \frac{d}{dt},$$

запишем приведенное выше дифференциальное уравнение (5.19) в более компактном виде

$$(a_0 p^n + a_1 p^{n-1} + \dots + a_n) y(t) = (b_0 p^m + b_1 p^{m-1} + \dots + b_m) u(t). \quad (5.20)$$

При записи и преобразовании дифференциальных уравнений оператор " p " можно рассматривать как алгебраический сомножитель, а выражение " $py(t)$ ", как произведение, не обладающее свойством коммутативности.

Введем в уравнении (5.20) обозначения

$$\begin{aligned} A(p) &= a_0 p^n + a_1 p^{n-1} + \dots + a_n, \\ B(p) &= b_0 p^m + b_1 p^{m-1} + \dots + b_m, \end{aligned}$$

где $A(p)$ - собственный оператор,

$B(p)$ - оператор воздействия,

Тогда исходное уравнение (5.19) можно записать еще короче

$$A(p)y(t) = B(p)u(t). \quad (5.21)$$

Отношение оператора воздействия к собственному оператору называют передаточной функцией в операторной форме

$$W(p) = \frac{B(p)}{A(p)} = \frac{b_0 p^m + b_1 p^{m-1} + \dots + b_m}{a_0 p^n + a_1 p^{n-1} + \dots + a_n}. \quad (5.22)$$

Наряду с передаточной функцией в операторной форме (5.22) широко используется передаточная функция в форме изображений Лапласа.

Передаточной функцией в форме изображений Лапласа называют отношение изображения выходной величины к изображению входной величины при нулевых начальных условиях.

Вернемся к исходному дифференциальному уравнению (5.19) и перейдем в обеих частях этого уравнения к изображениям Лапласа

$$\begin{aligned} \mathcal{L}\left\{a_0 \frac{d^n}{dt^n} y(t) + a_1 \frac{d^{n-1}}{dt^{n-1}} y(t) + \dots + a_n y(t)\right\} \\ = \mathcal{L}\left\{b_0 \frac{d^m}{dt^m} u(t) + b_1 \frac{d^{m-1}}{dt^{m-1}} u(t) + \dots + b_m u(t)\right\}. \end{aligned} \quad (5.23)$$

Воспользовавшись свойствами линейности и дифференцирования оригинала, применительно к уравнению (5.23) получаем

$$(a_0 s^n + a_1 s^{n-1} + \dots + a_n) Y(s) = (b_0 s^m + b_1 s^{m-1} + \dots + b_m) U(s), \quad (5.24)$$

где $Y(s) = \mathcal{L}\{y(t)\}$ - изображение по Лапласу функции $y(t)$ выходного сигнала, $U(s) = \mathcal{L}\{u(t)\}$ - изображение по Лапласу функции $u(t)$ входного сигнала.

Тогда, исходя из (5.24), для передаточной функции в форме изображений Лапласа можно записать

$$W(s) = \frac{Y(s)}{U(s)} = \frac{b_0 s^m + b_1 s^{m-1} + \dots + b_m}{a_0 s^n + a_1 s^{n-1} + \dots + a_n}. \quad (5.25)$$

Структурная схема, соответствующая уравнению (5.25) представлена на рис.5.4.

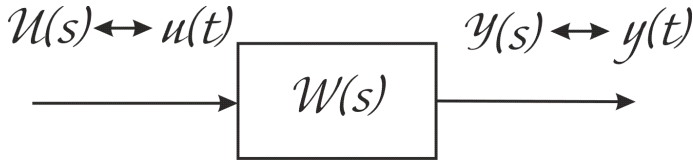


Рис.5.4. Общая структурная схема линейной системы

При введении понятия передаточной функции не было наложено никаких ограничений на физический вид входного и выходного сигнала. Поэтому в качестве таковых могут рассматриваться любые физические процессы и их характеристики, свойственные конкретной системе.

Если в качестве линейной системы рассматривать четырехполюсник, то в качестве входного сигнала могут быть приняты напряжение или ток на первичных зажимах, а в качестве выходных - напряжение или ток на вторичных зажимах. Таким образом, теоретически для четырехполюсника можно ввести 4 различных передаточных функции: по напряжению $\frac{U_2(s)}{U_1(s)}$, току $\frac{I_2(s)}{I_1(s)}$, передаточному сопротивлению $\frac{U_2(s)}{I_1(s)}$ или передаточной проводимости $\frac{I_2(s)}{U_1(s)}$.

Все четыре передаточные функции для четырехполюсника можно определить только в случае подключения к нему ненулевой и не бесконечной нагрузки (при этом параметры нагрузки входят в передаточную функцию).

Сравнивая полученные выражения для передаточных функций $W(p)$ и $W(s)$ видно, что они совпадают с точностью до замены переменных. Однако это сходство только внешнее и имеет место лишь для линейных систем с постоянными во времени параметрами.

Пример5.7.

Найдем передаточную функцию для цепи, изображенной на рис.5.5.

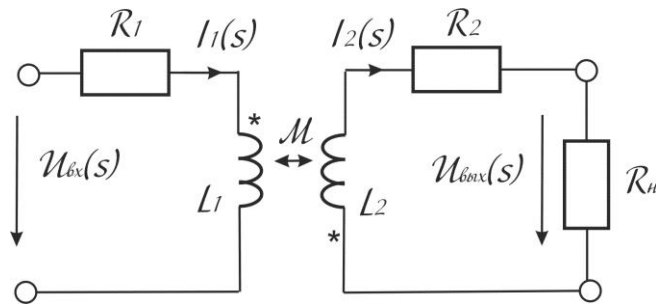


Рис. 5.5. Схема цепи

В операторной форме для двух обозначенных на схеме контуров для нулевых начальных условий запишем следующее

$$\begin{aligned} U_{\text{вх}}(s) &= R_1 I_1(s) + sL_1 I_1(s) + sM I_2(s) \\ 0 &= sM I_1(s) + sL_2 I_2(s) + R_2 I_2(s) + R_n I_2(s) \\ U_{\text{вых}}(s) &= R_n I_2(s). \end{aligned}$$

Тогда для передаточной функции по напряжению

$$W(s) = \frac{U_{\text{вых}}(s)}{U_{\text{вх}}(s)}$$

можно получить

$$W(s) = \frac{sMR_n}{(L_1 L_2 - M^2)s^2 + (R_1 L_2 + (R_2 + R_n)L_1)s + (R_1(R_2 + R_n))}$$

или в сокращенной записи

$$W(s) = \frac{sb_0}{a_0 s^2 + a_1 s + a_2}$$

5.3.2. Частотные характеристики линейных систем

Подобно представлению сигнала в частотной области с помощью спектральной функции, можно ввести частотные характеристики линейной системы с постоянными параметрами. Для этого произведем в уравнении для передаточной функции (5.25) замену переменных $s = j\omega$:

$$W(j\omega) = \frac{b_0(j\omega)^m + b_1(j\omega)^{m-1} + \dots + b_m}{a_0(j\omega)^n + a_1(j\omega)^{n-1} + \dots + a_n}. \quad (5.26)$$

Получившаяся в результате замены переменных комплекснозначная функция от действительного аргумента ω , который называют частотой, носит название частотной передаточной функции (или комплексного коэффициента передачи).

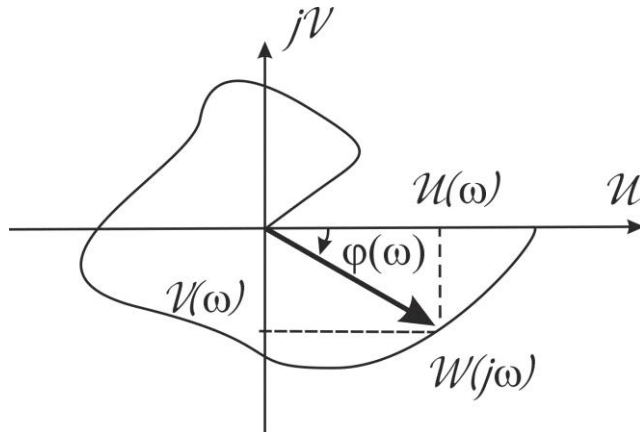
Полученную частотную передаточную функцию (5.26) можно представить в ином виде

$$W(j\omega) = U(\omega) + jV(\omega) = A(\omega)e^{j\varphi(\omega)}, \quad (5.27)$$

где $A(\omega) = \sqrt{U^2(\omega) + V^2(\omega)}$,

$$\varphi(\omega) = \arctg\left(\frac{V(\omega)}{U(\omega)}\right) + k\pi, \quad k = 0, \pm 1, \pm 2, \dots$$

На комплексной плоскости частотная передаточная функция $W(j\omega)$ (иногда называемая амплитудно-фазовой функцией) определяет вектор $\overrightarrow{OC}$, длина которого равна $A(\omega)$, а угол, образуемый с положительной вещественной осью $\varphi(\omega)$, как изображено на рис.5.6.



Кривую, которую описывает конец этого вектора при изменении частоты ω от нуля до бесконечности, называют амплитудно-фазовой частотной характеристикой (сокращенно АФЧХ).

Функции $U(\omega) = \text{Re}[W(j\omega)]$, $V(\omega) = \text{Im}[W(j\omega)]$ называют вещественной и мнимой частотными функциями соответственно, а их графики вещественной и мнимой частотными характеристиками.

Рис.5.6. Амплитудно-фазовая частотная характеристика

Функции $A(\omega) = |W(j\omega)|$, $\varphi(\omega) = \arg(W(j\omega))$ из (5.27) носят названия амплитудной и фазовой частотных функций, а их графики соответствующих частотных характеристик.

Кроме перечисленных характеристик достаточно широко используются полулогарифмические и логарифмические частотные характеристики, получаемые из исходных при переходе к логарифмическому масштабу по одной или обеим осям соответственно.

Функцию

$$L(\omega) = 20 \lg(A(\omega)) = 20 \lg(|W(j\omega)|), \quad (5.28)$$

которую называют логарифмической амплитудной частотной функцией. График $L(\omega)$ в зависимости от логарифма частоты $\lg \omega$ соответственно имеет название логарифмической амплитудной частотной характеристики. График зависимости $\varphi(\omega)$ от логарифма частоты $\lg \omega$ называют логарифмической фазовой частотной характеристикой.

Единицей измерения $L(\omega)$ из (5.28) является децибел (дБ). Данная единица измерения заимствована из акустики. Усиление сигнала, при котором мощность возрастает в 10^q раз, считается равным q Б. Поскольку мощность сигнала пропорциональна квадрату амплитуды, то увеличение амплитуды в K раз соответствует усилению $20 \lg K$ дБ.

Ось ординат при построении логарифмических характеристик проводят через произвольную точку, а не через $\omega = 0$, так как этой частоте соответствует бесконечно удаленная точка $\lg \omega \rightarrow \infty$. Единицей измерения величины $\lg \omega$ является декада, то есть частотный интервал, соответствующий десятикратному изменению частоты.

Иногда при построении логарифмических частотных характеристик для оси частот используют не десятичный логарифм, а логарифм по основанию 2. В этом случае вместо частотного интервала декада используют интервал – октава, соответствующий двукратному изменению частоты.

По причине такого широкого многообразия в выборе способов изображения частотных характеристик следует во избежание недоразумений обращать особое внимание на то, какие именно величины откладываются по осям графиков.

Рассмотрим теперь, каков физический смысл имеют введенные частотные характеристики.

Допустим, что на вход линейной системы с передаточной функцией

$$W(s) = \frac{b_0 s^m + b_1 s^{m-1} + \dots + b_m}{a_0 s^n + a_1 s^{n-1} + \dots + a_n}$$

подан сигнал вида

$$u(t) = U_m \cos \omega t.$$

Для ответа на вопрос, каков при этом будет сигнал на выходе системы, требуется решить дифференциальное уравнение

$$(a_0 p^n + a_1 p^{n-1} + \dots + a_n) y(t) = (b_0 p^m + b_1 p^{m-1} + \dots + b_m) u(t).$$

Общее решение данного уравнения имеет вид

$$y(t) = y_c(t) + y_b(t),$$

где $y_c(t)$ - общее решение однородного дифференциального уравнения, определяющее «свободную» составляющую движения системы (переходный процесс),

$y_b(t)$ - частное решение неоднородного дифференциального уравнения, которое определяет так называемое «вынужденное» решение.

Нас в данном контексте будет интересовать вид выходного сигнала в установившемся режиме, то есть после того как окончится переходный процесс. (Сейчас мы рассматриваем только системы, для которых $y_c(t) \xrightarrow{t \rightarrow \infty} 0$).

Для нахождения $y_b(t)$ в соответствии с формулой Эйлера представим входной сигнал в виде

$$u(t) = U_m \cos \omega t = U_m \left(\frac{e^{j\omega t} + e^{-j\omega t}}{2} \right) = u_1(t) + u_2(t),$$

$$\text{где } u_1(t) = \frac{U_m}{2} e^{j\omega t}, \quad u_2(t) = \frac{U_m}{2} e^{-j\omega t}.$$

Тогда в силу принципа суперпозиции искомое решение также будет состоять из двух компонент

$$y(t) = y_1(t) + y_2(t),$$

где y_1 – решение, получаемое при входном сигнале u_1 ,

y_2 – решение, получаемое при входном сигнале u_2 .

Заметим, что

$$p u_1(t) = p \frac{U_m}{2} e^{j\omega t} = \frac{U_m}{2} p e^{j\omega t} = j\omega u_1(t)$$

и

$$p^n u_1(t) = (j\omega)^n u_1(t).$$

Частное решение для уравнения

$$(a_0 p^n + a_1 p^{n-1} + \dots + a_n) y_1(t) = (b_0 p^m + b_1 p^{m-1} + \dots + b_m) u_1(t)$$

будем искать в виде

$$y_1(t) = A_1 u_1(t) = A_1 \frac{U_m}{2} e^{j\omega t}.$$

Подставляя это выражение в уравнение (5.19)

$$(a_0 (j\omega)^n + a_1 (j\omega)^{n-1} + \dots + a_n) A_1 u_1 = (b_0 (j\omega)^m + b_1 (j\omega)^{m-1} + \dots + b_m) u_1$$

откуда

$$A_1 = \frac{b_0 (j\omega)^m + b_1 (j\omega)^{m-1} + \dots + b_m}{a_0 (j\omega)^n + a_1 (j\omega)^{n-1} + \dots + a_n}.$$

Сравнивая полученное выражение с формулой (5.26), определяющей частотную передаточную функцию, видно, что

$$A_1 = W(j\omega) = A(\omega) e^{j\varphi(\omega)}.$$

Следовательно,

$$y_1(t) = A(\omega)e^{j\varphi(\omega)} \frac{U_m}{2} e^{j\omega t} = A(\omega) \frac{U_m}{2} e^{j(\omega t + \varphi(\omega))}.$$

Аналогично

$$y_2(t) = A(\omega) \frac{U_m}{2} e^{-j(\omega t + \varphi(\omega))}.$$

Тогда

$$\begin{aligned} y(t) = y_1(t) + y_2(t) &= A(\omega) \frac{U_m}{2} e^{j(\omega t + \varphi(\omega))} + A(\omega) \frac{U_m}{2} e^{-j(\omega t + \varphi(\omega))} \\ &= A(\omega) U_m \cos(\omega t + \varphi(\omega)). \end{aligned} \quad (5.29)$$

Таким образом, из уравнения (5.29) видно, что при подаче на вход устойчивой линейной системы с постоянными параметрами гармонического сигнала на ее выходе в установившемся режиме будет наблюдаться выходной сигнал, изменяющийся по гармоническому закону с той же частотой, но с другими значениями амплитуды и фазы. При этом отношение выходной и входной амплитуд равно модулю, а сдвиг фаз - аргументу частотной передаточной функции, описывающей данную линейную систему.

5.3.3. Временные характеристики линейных цепей

Рассмотренные выше частотные функции и характеристики используются при анализе и синтезе сигналов и систем в частотной области. Для анализа во временной области широко используются временные характеристики – переходная и импульсная (весовая) функции системы.

Переходной функцией системы называют функцию, описывающую изменение выходного сигнала системы, при подаче на ее вход единичного ступенчатого воздействия (единичного скачка) при нулевых начальных условиях.

Наиболее часто для обозначения переходной функции используют $h(t)$. График переходной функции носит название переходной или разгонной характеристики.

Импульсной переходной (или весовой) функцией системы называют функцию, описывающую изменения выходного сигнала системы, возникающего при подаче на вход системы единичного импульса при нулевых начальных условиях. Для обозначения весовой функции мы будем использовать $w(t)$. График весовой функции называют импульсной характеристикой системы.

Напомним (см. (1.42)), что функция единичного скачка определяется как

$$\mathbf{1}(t) = \begin{cases} 0, & \text{если } t < 0 \\ \frac{1}{2}, & \text{если } t = 0 \\ 1, & \text{если } t > 0. \end{cases}$$

Значение функции в нулевой момент времени иногда считают неопределенным или приравнивают единице.

Единичный импульс (см. (1.40)) задается следующими выражениями

$$\delta(t) = \begin{cases} \infty, & \text{при } t = 0 \\ 0, & \text{при } t \neq 0 \end{cases}$$

и одновременно

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

Производные от единичного импульса определяются в виде

$$\int_{-\infty}^{\infty} \dot{\delta}(t) f(t) dt = -\dot{f}(0), \quad \int_{-\infty}^{\infty} \ddot{\delta}(t) f(t) dt = \ddot{f}(0), \quad \int_{-\infty}^{\infty} \dddot{\delta}(t) f(t) dt = -\dddot{f}(0) \text{ и т. д.}$$

Используя эти соотношения легко найти преобразование Лапласа для дельта-функции и ее производных

$$\mathcal{L}\{\delta(t)\} = 1, \quad \mathcal{L}\{\dot{\delta}(t)\} = s, \quad \mathcal{L}\{\ddot{\delta}(t)\} = s^2 \text{ и т. д.} \quad (5.30)$$

Если линейная система описывается передаточной функцией $W(s)$, то связь между изображениями входного и выходного сигналов определяется как

$$Y(s) = W(s)U(s). \quad (5.31)$$

Для определения переходной функции надо задать входной сигнал в виде единичного скачка, а для получения весовой функции – в виде единичного импульса.

Принимая во внимание, что в соответствии с (5.30) $\mathcal{L}\{\delta(t)\} = 1$ и $\mathcal{L}\{1(t)\} = \frac{1}{s}$, получаем для весовой функции

$$\mathcal{L}\{w(t)\} = W(s) \quad \text{или} \quad w(t) = \mathcal{L}^{-1}\{W(s)\} \quad (5.32)$$

и для переходной функции

$$\mathcal{L}\{h(t)\} = \frac{1}{s} W(s) \quad \text{или} \quad h(t) = \mathcal{L}^{-1}\left\{\frac{W(s)}{s}\right\}. \quad (5.33)$$

При этом связь между весовой и переходной функциями находится из соотношения

$$w(t) = \frac{d}{dt} h(t). \quad (5.34)$$

Переходная и импульсная функции (5.32) – (5.34) исчерпывающе описывают линейную систему при нулевых начальных условиях. С их помощью по известному входному сигналу произвольной формы $u(t)$ можно однозначно определить выходной сигнал $y(t)$. Действительно, применяя теорему о свертке для преобразования Лапласа к уравнению (5.31) получим

$$y(t) = \int_0^t w(t-\tau)u(\tau)d\tau = \int_0^t w(\tau)u(t-\tau)d\tau. \quad (5.35)$$

На практике вместо уравнения (5.35) часто используют способ вычисления выходного сигнала системы по переходной характеристике с помощью, так называемого интеграла наложения (или интеграла Дюамеля)

$$y(t) = u(t_0)h(t-t_0) + \int_{t_0}^t \frac{du(\tau)}{d\tau} h(t-t_0)dt. \quad (5.36)$$

С помощью интеграла Дюамеля (5.36) можно определять реакцию системы на заданный входной сигнал и в том случае, когда входной сигнал представляет собой кусочно-непрерывную функцию времени. В этом случае интервал интегрирования следует разбить на несколько промежутков в соответствии с интервалами непрерывности входного сигнала и учесть реакцию системы на конечные скачки в точках разрыва.

Пример 5.8.

Воспользуемся передаточной функцией, полученной в примере 5.7.

Рассчитаем соответствующие ей частотные и временные характеристики, воспользовавшись программой MATLAB. Результаты моделирования приведены на рис. 5.7.

```
M=0.01;           % задание исходных параметров цепи
L1=0.1;
L2=0.1;
R1=100;
R2=100;
RN=1000;
a0=L1*L2-M^2      % расчет коэффициентов ПФ
a1=R1*L1+(R2+RN)*L2
a2=R1*(R1+RN)
b0=M*RN
num=[b0 0];       % задание параметров числителя ПФ
den=[a0 a1 a2];    % задание параметров знаменателя ПФ
W=tf(num,den)      % формирование модели системы в tf-форме
subplot(2,2,1); bode(W); grid on ; % построение диаграммы Бode
subplot(2,2,2); nyquist(W); grid on ; % вычисление годографа Найквиста
subplot(2,2,3); step(W); grid on ; % определение переходной функции
subplot(2,2,4); impulse(W); grid on ; % определение импульсной переходной функции
```

Визуально проконтролируем параметры моделируемой передаточной функции

```
a0 = 0.0099
a1 = 120
a2 = 110000
b0 = 10
W =
```

10 s

0.0099 s² + 120 s + 110000

Continuous-time transfer function.

Результаты моделирования представим в графическом виде:

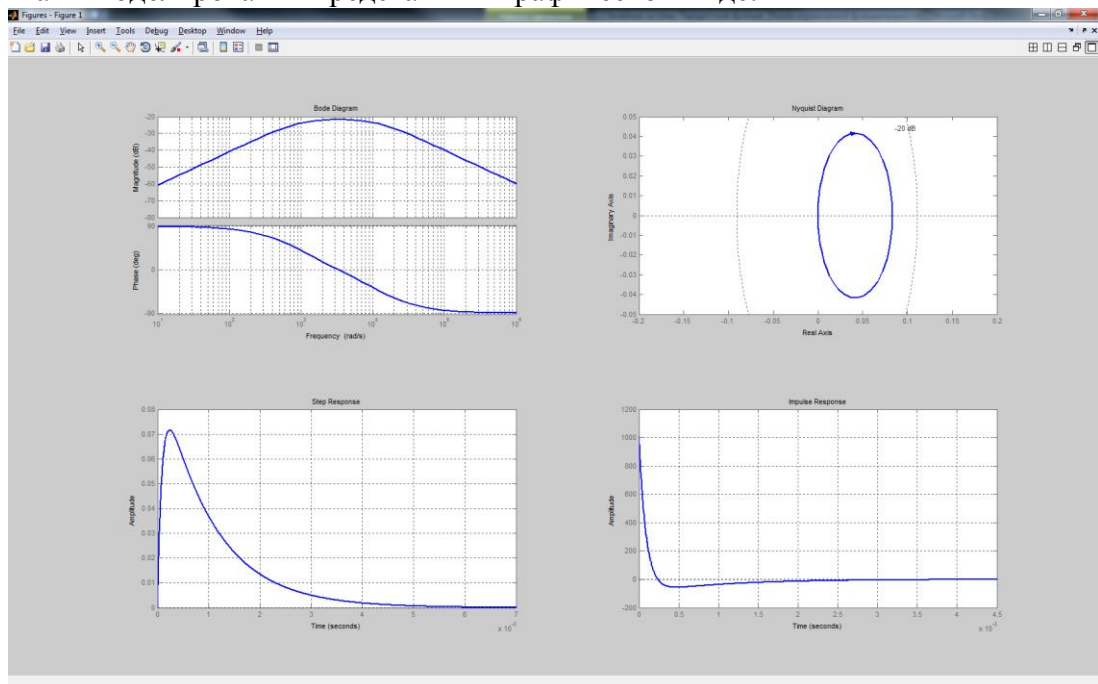


Рис.5.7. Нахождение частотных и временных характеристик цепи по заданной передаточной функции

Пример.5.9.

Рассмотрим процесс применения интеграла Дюамеля для определения реакции линейной цепи на сложное входное воздействие с использованием программы MathCad. (См. рис. 5.8.)

Выберем цепь с импульсной характеристикой вида

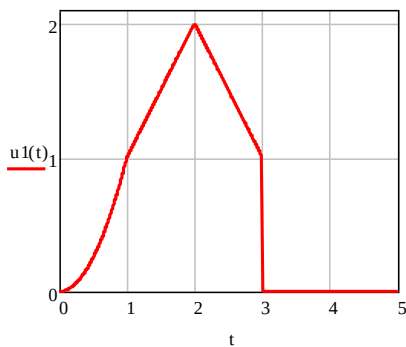
$$g(t) := \frac{1}{\tau} e^{-\frac{t}{\tau}}$$

$$\tau := 1$$

Предположим, что входной сигнал задается уравнением

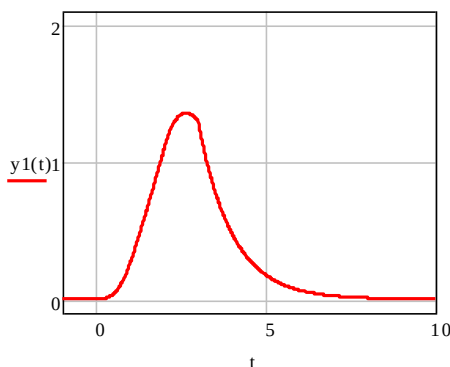
$$u_1(t) := \begin{cases} t^2 & \text{if } 0 \leq t \leq 1 \\ t & \text{if } 1 < t < 2 \\ [-(t-2) + 2] & \text{if } 2 < t < 3 \\ 0 & \text{otherwise} \end{cases}$$

График входного сигнала имеет вид



Для определения выходного сигнала воспользуемся уравнением

$$y_1(t) := \int_0^t u_1(\tau) g(t - \tau) d\tau$$



В результате получаем вид выходного сигнала

В качестве дополнительного упражнения рекомендуем проделать аналитические преобразования для данного примера и сравнить свои результаты с полученным в программе графиком выходного сигнала.

Рис.5.8. Нахождение выходного сигнала цепи с помощью интеграла наложения.

5.4. Формы представления передаточной функции

5.4.1. Нули и полюсы

Для любой линейной системы с постоянными параметрами передаточная функция системы имеет вид дробно-рациональной функции, то есть отношения двух полиномов от переменной s :

$$W(s) = \frac{b_0 s^m + b_1 s^{m-1} + \dots + b_m}{a_0 s^n + a_1 s^{n-1} + \dots + a_n}.$$

В соответствии с теоремой Безу любой полином может быть разложен на множители, соответствующие его корням. С учетом этого свойства передаточную функцию $W(s)$ (при $n \geq m$) можно привести к следующему виду

$$W(s) = K \frac{(s - z_m)(s - z_{m-1}) \dots (s - z_1)}{(s - p_n)(s - p_{n-1}) \dots (s - p_1)}, \quad (5.37)$$

где $K = \frac{b_m}{a_n}$ – постоянный коэффициент, называемый коэффициентом усиления,

p_n – корни знаменателя передаточной функции, называемые *полюсами*,

z_m – корни числителя передаточной функции, называемые *нулями*.

Для передаточных функций с вещественными значениями коэффициентов нули и полюсы являются либо вещественными, либо комплексно-сопряженными, а коэффициент усиления всегда вещественный.

Таким образом, свойства передаточной функции полностью определяется набором полюсов и нулей (5.27), а их расположение на комплексной плоскости позволяет получить важную качественную и количественную информацию о свойствах системы.

5.4.2. Диаграмма полюсов и нулей

Графическое изображение расположения полюсов и нулей передаточной функции на комплексной s - плоскости называется диаграммой полюсов-нулей. На рис. 5.9. приведен пример расположения нулей и полюсов для передаточной функции

$$W(s) = \frac{s + 0.5}{s^2 - 2s + 2} = \frac{s + 0.5}{(s - (1 - j))(s - (1 + j))},$$

имеющей один нуль в точке $z_1 = -0.5$ и пару комплексно-сопряженных полюсов $p_1 = 1 - j$ и $p_2 = 1 + j$.

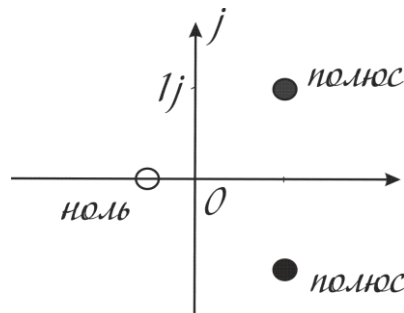


Рис.5.9. Диаграмма полюсов-нулей

Представление передаточной функции в виде полюсов и нулей оказывается также весьма полезным при построении логарифмических частотных характеристик систем.

Положим для начала, что в системе присутствуют только вещественные полюса и нули, отличные от $s = 0$.

Тогда выражение для передаточной функции можно представить в виде

$$W(s) = K(0) \frac{(1 + \tau_m s)(1 + \tau_{m-1} s) \dots (1 + \tau_1 s)}{(1 + T_n s)(1 + T_{n-1} s) \dots (1 + T_1 s)},$$

где τ_m - постоянная времени, соответствующая нулю передаточной функции z_m ,
 T_n - постоянная времени, соответствующая полюсу передаточной функции p_n .

Переходя к модулю частотной передаточной функции $|W(j\omega)|$ и прологарифмировав это выражение, получаем

$$L(\omega) = 20 \lg |W(j\omega)| = 20 \lg K(0) + \\ + 20 \lg |1 + j\tau_m \omega| + 20 \lg |1 + j\tau_{m-1} \omega| + \dots + 20 \lg |1 + j\tau_1 \omega| - \\ - 20 \lg |1 + jT_m \omega| + 20 \lg |1 + jT_{m-1} \omega| + \dots + 20 \lg |1 + jT_1 \omega|$$

и

$$\varphi(\omega) = \arg(W(j\omega)) = \\ = \arg(1 + j\tau_m \omega) + \arg(1 + j\tau_{m-1} \omega) + \dots + \arg(1 + j\tau_1 \omega) - \\ - \arg(1 + jT_m \omega) - \arg(1 + jT_{m-1} \omega) - \dots - \arg(1 + jT_1 \omega).$$

Таким образом, если нам известны амплитудные и фазовые характеристики отдельных членов числителя $(1 + j\tau_m \omega)$ и знаменателя $(1 + jT_n \omega)$, то итоговые характеристики системы находятся простым суммированием. (В зарубежной литературе этот метод нахождения частотных характеристик часто называют методом Боде, а получающиеся в результате графики – диаграммами Боде).

Полученная логарифмическая амплитудная характеристика имеет три вида возможных прямолинейных асимптот: горизонтальную – соответствующую члену $20 \lg K(0)$, и две наклонные с крутизной $+20$ дБ/дек и -20 дБ/дек, соответствующие членам $(1 + j\tau_m \omega)$ и $1 + jT_n \omega$ соответственно. (Наклон 20 дБ/дек соответствует 6 дБ/октаву). Наибольшие отклонения от асимптот происходят в точках $\omega = 1/\tau$ и $\omega = 1/T$ соответственно.

Теперь рассмотрим, что происходит при наличии в передаточной функции комплексно-сопряженных нулей или полюсов. Например, если передаточная функция имеет пару комплексно-сопряженных полюсов

$$W(s) = \frac{K}{(s - (a - jb))(s - (a + jb))} = \frac{K}{s^2 - 2as + a^2 + b^2},$$

что приводится к виду

$$W(s) = \frac{K}{T^2 s^2 + 2\xi T s + 1},$$

где $0 < \xi < 1$ - коэффициент колебательности.

Тогда частотная передаточная функция примет вид

$$W(j\omega) = \frac{K}{(1 - T^2 \omega^2) + j2\xi \omega T}.$$

Следовательно, в данном случае наклонная асимптота, будет иметь крутизну -40 дБ/дек (или -12 дБ/окт), а величина отклонения от асимптоты в точке $\omega = 1/T$, будет зависеть от коэффициента ξ .

5.4.3. Полюсы и вычеты

Помимо представления передаточной функции в виде произведения членов (5.37), соответствующих нулям и полюсам, возможно представление в виде суммы элементарных дробей. Так в случае простых полюсов передаточная функция степень числителя которой меньше степени знаменателя представляется в виде

$$W(s) = \frac{k_1}{s - p_1} + \frac{k_2}{s - p_2} + \dots + \frac{k_n}{s - p_n}, \quad (5.38)$$

где k_n - коэффициенты, называемые вычетами.

Представление передаточной функции в виде (5.38) чрезвычайно удобно для нахождения импульсной (весовой) функции системы.

Действительно, если коэффициенты разложения k_n найдены, то импульсная функция находится сразу в виде

$$w(t) = k_1 e^{p_1 t} + k_2 e^{p_2 t} + \dots + k_n e^{p_n t}, \quad t > 0.$$

В случае равенства степеней числителя и знаменателя передаточной функции в разложение добавляется постоянный член k_0 , а в случае кратного корня p_i с кратностью l вместо одного слагаемого, соответствующего этому корню появляются l слагаемых вида

$$\frac{k_{i1}}{s - p_i} + \frac{k_{i2}}{(s - p_i)^2} + \dots + \frac{k_{il}}{(s - p_i)^l}.$$

Соответствующим образом изменится и слагаемое в выражении для импульсной функции

$$k_{i1} e^{p_i t} + k_{i2} t e^{p_i t} + \dots + \frac{t^{m-1}}{(m-1)!} k_{in} e^{p_i t}.$$

Для случая простых корней вычеты k_n находятся по простой формуле

$$k_n = [X(s)(s - s_k)]|_{s=s_k}.$$

Применение данного метода продемонстрируем на примерах.

Пример 5.10.

Допустим, что изображение Лапласа задано в виде следующей дробно-рациональной функции

$$X(s) = \frac{s + 2}{s^2 + s}.$$

Произведем необходимые преобразования для приведения данного выражения к сумме простых дробей

$$X(s) = \frac{s + 2}{s^2 + s} = \frac{s + 2}{s(s + 1)} = \frac{k_1}{s} + \frac{k_2}{s + 1}.$$

Искомые коэффициенты определим как

$$k_1 = [X(s)(s - s_1)]|_{s=s_1} = \frac{s + 2}{s + 1}|_{s=s_1=0} = 2$$

$$k_2 = [X(s)(s - s_2)]|_{s=s_2} = \frac{s + 2}{s}|_{s=s_2=-1} = -1$$

В результате для изображения получаем

$$X(s) = \frac{2}{s} - \frac{1}{s + 2}.$$

Тогда для функции оригинала легко записать

$$x(t) = 2 - e^{-t}.$$

Пример 5.11.

Если полюса функции-изображения являются комплексно-сопряженными, то общая идея преобразования остается той же.

Например, пусть

$$X(s) = \frac{s + 1}{s^2 + 1}.$$

То есть мы имеем случай двух различных, но комплексно сопряженных корней

$$X(s) = \frac{s + 1}{s^2 + 1} = \frac{s + 1}{(s + j)(s - j)} = \frac{k_1}{s + j} + \frac{k_2}{s - j}.$$

Аналогично предыдущему примеру найдем искомые коэффициенты разложения

$$k_1 = [X(s)(s - s_1)]|_{s=s_1} = \frac{s + 1}{s - j}|_{s=s_1=-j} = \frac{-j + 1}{-2j} = \frac{1}{\sqrt{2}} e^{j\pi/4}$$

$$k_2 = [X(s)(s - s_2)]|_{s=s_2} = \frac{s+1}{s+j}\bigg|_{s=s_2=j} = \frac{j+1}{2j} = \frac{1}{\sqrt{2}} e^{-j\pi/4}.$$

Объединяя два последних выражения, и переходя к функции-оригиналу получаем

$$x(t) = \frac{1}{\sqrt{2}} e^{j\pi/4} e^{-jt} + \frac{1}{\sqrt{2}} e^{-j\pi/4} e^{jt} = \sqrt{2} \cos\left(t - \frac{\pi}{4}\right).$$

5.5. Связь передаточной функции с коэффициентами четырехполюсника

В предыдущих разделах мы выяснили, что частотная передаточная функция может быть получена из передаточной функции в форме Лапласа путем подстановки $s = j\omega$. Теперь определим связь частотной передаточной функции с параметрами четырехполюсника.

Обобщенная схема четырехполюсника с комплексной нагрузкой приведена на рис. 5.10.

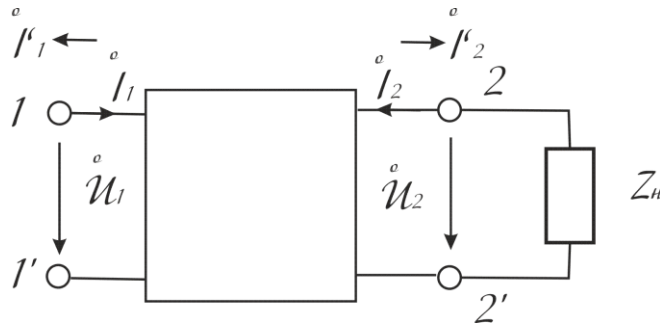


Рис. 5.10. Четырехполюсник с комплексной нагрузкой

Напомним, что в А-форме четырехполюсник описывается системой уравнений

$$\begin{cases} U_1^o = AU_2^o + BI_2^{o'} \\ I_1^o = CU_2^o + DI_2^{o'} \\ U_2^o = AU_1^o + BI_1^{o'} \\ I_2^o = CU_1^o + DI_1^{o'} \end{cases}$$

и в случае если $Z_H \neq 0$ и $Z_H \neq \infty$ можно определить все четыре возможных частотных передаточных функции

$$\begin{aligned} W_U(j\omega) &= \frac{U_2^o}{U_1^o} = \frac{U_2^o}{AU_2^o + BI_2^{o'}} = \frac{I_2^{o'} Z_H}{AI_2^{o'} Z_H + BI_2^{o'}} = \frac{Z_H(j\omega)}{A(j\omega)Z_H(j\omega) + B(j\omega)} \\ W_I(j\omega) &= \frac{I_2^{o'}}{I_1^o} = \frac{I_2^{o'}}{CU_2^o + DI_2^{o'}} = \frac{1}{C(j\omega)Z_H(j\omega) + D(j\omega)} \\ W_Z(j\omega) &= \frac{U_2^o}{I_1^o} = \frac{U_2^o}{CU_2^o + DI_2^{o'}} = \frac{Z_H(j\omega)}{C(j\omega)Z_H(j\omega) + D(j\omega)} \\ W_Y(j\omega) &= \frac{I_2^{o'}}{U_1^o} = \frac{I_2^{o'}}{AU_2^o + BI_2^{o'}} = \frac{1}{A(j\omega)Z_H(j\omega) + B(j\omega)}. \end{aligned}$$

В режиме холостого хода ($Z_H = \infty$) отличны от нуля только две передаточные функции

$$W_{U0}(j\omega) = \frac{1}{A(j\omega)} \text{ и } W_{Z0}(j\omega) = \frac{1}{C(j\omega)}.$$

В режиме короткого замыкания ($Z_H = 0$) от нуля отличны две другие передаточные функции

$$W_{IK}(j\omega) = \frac{1}{D(j\omega)} \text{ и } W_{YK}(j\omega) = \frac{1}{B(j\omega)}.$$

Пример.5.12.

Найдем частотную передаточную функцию по напряжению в режиме холостого хода для четырехполосника, схема которого приведена на рис. 5.11.

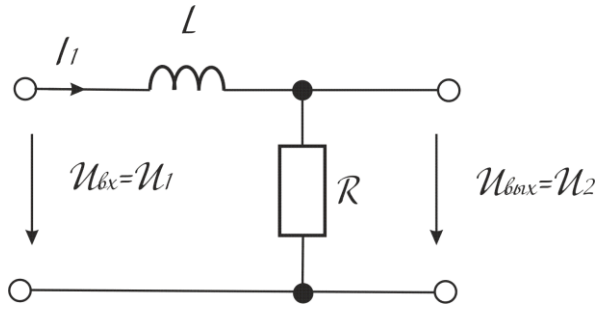


Рис.5.11. Схема исследуемого четырехполосника

Используя операторный метод описания цепей, для данного случая запишем следующие уравнения

$$\begin{aligned} U_1(s) &= I_1(s)(sL + R), \\ U_2(s) &= I_1(s)R \end{aligned}$$

откуда получаем

$$W_U(s) = \frac{R}{sL + R} = \frac{1}{s \frac{L}{R} + 1} = \frac{1}{\tau s + 1}.$$

Следовательно,

$$W_U(j\omega) = \frac{1}{j\omega\tau + 1}.$$

С другой стороны $W_U(j\omega)$ в режиме холостого хода должна удовлетворять полученному ранее при изучении четырехполосников соотношению

$$W_{U0}(j\omega) = \frac{1}{A(j\omega)},$$

где A – коэффициент четырехполосника, описываемого системой уравнений

$$\begin{cases} U_1^o = AU_2^o + BI_2^{o'} \\ I_1^o = CU_2^o + DI_2^{o'} \end{cases}$$

Проверим, действительно ли это так.

$$A = \left. \frac{U_1^o}{U_2^o} \right|_{I_2^{o'}=0} = \frac{(Z_L + Z_R)I_1^o}{Z_R I_1^o} = \frac{j\omega L + R}{R}.$$

Откуда

$$W_{U0}(j\omega) = \frac{1}{A(j\omega)} = \frac{1}{j\omega\tau + 1},$$

где $\tau = \frac{L}{R}$.

В результате, как и ожидалось, получено точно такое же уравнение, как при использовании операторного метода анализа цепей с последующим переходом к частотному представлению передаточной функции.

5.6. Пространство состояний

Метод описания цепей (как линейных систем) в переменных состояния основан на понятии состояния системы. Состояние динамической системы описывается совокупностью физических переменных $x_1(t)$, $x_2(t), \dots, x_n(t)$, характеризующих

поведение системы в будущем при условии, если известно состояние в исходный момент времени и приложенные к системе воздействия.

Начнем рассмотрение описания системы в пространстве состояний с простого примера.

Пусть система описывается передаточной функцией вида

$$W(s) = \frac{b}{a_0 s^n + a_1 s^{n-1} + \dots + a_n},$$

которая соответствует дифференциальному уравнению

$$a_0 \frac{d^n}{dt^n} y(t) + a_1 \frac{d^{n-1}}{dt^{n-1}} y(t) + \dots + a_n y(t) = u(t).$$

Введем новые переменные

$$\begin{aligned} x_1(t) &= y(t), \\ x_2(t) &= \frac{dx_1}{dt} = x_1'(t), \\ &\quad \text{---} \text{---} \text{---} \text{---} \\ x_n(t) &= \frac{dx_{n-1}}{dt} = x_{n-1}'(t). \end{aligned}$$

Тогда переписывая уравнения относительно первых производных, получаем

$$\begin{cases} x_1'(t) = x_2 \\ x_2'(t) = x_3 \\ \quad \text{---} \text{---} \text{---} \text{---} \\ x_n'(t) = -\frac{a_n}{a_0} x_1(t) - \frac{a_{n-1}}{a_0} x_2(t) - \dots - \frac{a_n}{a_0} x_n(t) + \frac{b}{a_0} u(t), \end{cases}$$

и

$$y(t) = x_1(t).$$

В векторно-матричном виде эти уравнения запишутся как

$$\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}u(t), \tag{5.40}$$

$$y(t) = \mathbf{C}\mathbf{x}(t),$$

где

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \\ -a_n/a_0 & -a_{n-1}/a_0 & -a_{n-2}/a_0 & \dots & -a_1/a_0 \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \dots \\ b \\ a_0 \end{bmatrix} \quad \mathbf{C} = [1 \quad 0 \quad 0 \quad \dots \quad 0].$$

Форма представления модели (5.40) носит название формы фазовой переменной.

Структурная схема данного представления модели цепи приведена на рис.5.12.

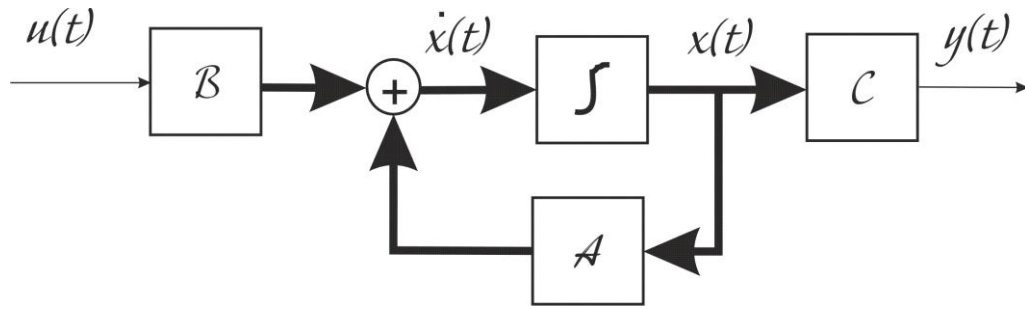


Рис.5.12. Структурная схема модели цепи в пространстве состояний

В общем случае, когда степень полинома числителя больше нуля представление системы в форме пространства состояний имеет вид

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}u(t), \quad (5.41)$$

$$y(t) = \mathbf{C}\mathbf{x}(t) + Du(t),$$

где

$y(t)$ – выходной сигнал,

$u(t)$ – входной сигнал,

$\mathbf{x}(t)$ – n -мерный вектор состояния,

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} - \text{квадратная матрица размерности } n \times n,$$

$$\mathbf{B} = \begin{bmatrix} b_1 \\ b_2 \\ \dots \\ b_n \end{bmatrix} - \text{вектор-столбец размерности } n \times 1,$$

$$\mathbf{C} = [c_1 \quad c_2 \quad \dots \quad c_n] - \text{вектор-строка размерности } 1 \times n,$$

D – скаляр, принимающий значение отличное от нуля только в системах порядок числителя и знаменателя, передаточных функций которых равны друг другу.

Переход от представления в пространстве состояний к представлению в виде передаточной функции системы производится путем взятия преобразования Лапласа от уравнений (5.41), что дает следующее

$$W(s) = D + \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B},$$

$$\text{где } \mathbf{I} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{bmatrix} - \text{единичная матрица размера } n \times n.$$

Структурная схема представления передачи сигнала системой, описываемой переменными состояния и передаточной функцией приведены на рис.5.13.

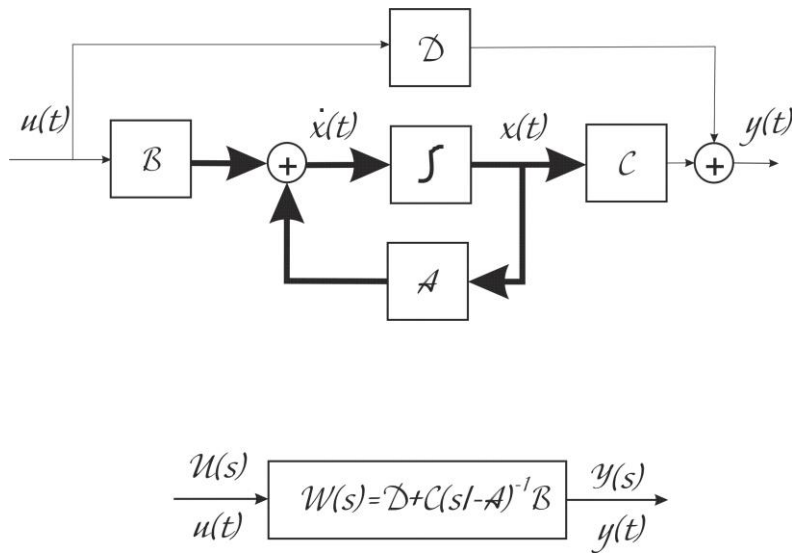


Рис.5.13. Представление модели линейной системы с помощью переменных состояния и передаточной функции

Переход от описания в виде передаточной функции к переменным состояния не является однозначным, то есть для одной и той же передаточной функции могут быть построены различные (в общем случае бесконечно много) представления в пространстве состояний, получаемых друг из друга произвольным линейным преобразованием, которому соответствует невырожденная квадратная матрица той же размерности, что и у знаменателя передаточной функции. Введя для такого преобразования обозначение M , получаем

$$A_M = MAM^{-1},$$

$$B_M = MB,$$

$$C_M = CM^{-1},$$

$$D_M = D.$$

По причине неоднозначности перехода от описания в виде передаточной функции к переменным состояния, с целью некоторого однообразия стараются использовать те или иные унифицированные представления, называемые каноническими представлениями, например, диагональную форму, горизонтально сопровождающую форму (к ней в частности относится рассмотренная нами форма фазовой переменной), вертикально сопровождающую форму и ряд других.

Дальнейшее обобщение такого способа описания динамических систем происходит при исследовании систем, имеющих несколько входных сигналов и несколько выходных. Общая структура модели остается такой же, как было описано, с тем исключением, что входные и выходные сигналы становятся векторными.

Обобщенная схема системы со многими входами и выходами приведена на рис.5.14.

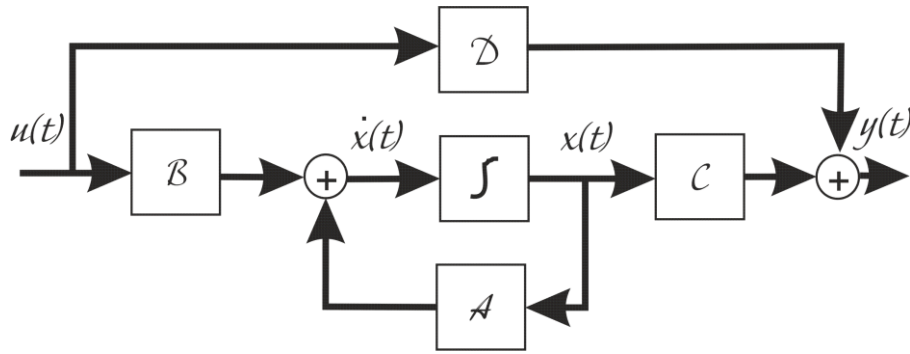


Рис.5.14. Обобщенная структурная схема модели многомерной линейной системы

5.7. Фундаментальная матрица системы и некоторые способы ее вычисления

Рассмотрим способы решения полученного в предыдущем разделе векторного уравнения состояния линейной системы с постоянными параметрами

$$\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}u(t). \quad (5.42)$$

Начнем со скалярного случая ($n = 1$)

$$x'(t) = ax(t) + bu(t).$$

Применяя к данному уравнению преобразование Лапласа, получим

$$sX(s) - x(0) = aX(s) + bU(s),$$

откуда следует

$$X(s) = \frac{x(0)}{s - a} + \frac{b}{s - a}U(s).$$

Взяв обратное преобразование Лапласа от полученного уравнения, имеем

$$x(t) = e^{at}x(0) + \int_0^t e^{a(t-\tau)}bu(\tau)d\tau.$$

Для векторного случая ($n > 1$) из (5.42) получаем

$$s\mathbf{X}(s) - \mathbf{x}(0) = \mathbf{A}\mathbf{X}(s) + \mathbf{B}U(s)$$

или

$$\mathbf{X}(s) = (s\mathbf{I} - \mathbf{A})^{-1}\mathbf{x}(0) + (s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}U(s), \quad (5.43)$$

где $\mathbf{I} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{bmatrix}$ - единичная матрица размерности $n \times n$.

По аналогии с рассмотренным скалярным случаем для (5.43) получим следующее выражение

$$\mathbf{x}(t) = e^{\mathbf{A}t}\mathbf{x}(0) + \int_0^t e^{\mathbf{A}(t-\tau)}\mathbf{B}u(\tau)d\tau.$$

В полученное уравнение входит матричная экспонента $e^{\mathbf{A}t}$, которая может быть представлена в виде ряда

$$e^{\mathbf{A}t} = \mathbf{I} + \mathbf{A}t + \frac{\mathbf{A}^2t^2}{2!} + \dots + \frac{\mathbf{A}^nt^n}{n!} + \dots,$$

сходящегося для всех конечных значениях t .

При этом функцию

$$\Phi(t) = e^{At} = \begin{bmatrix} \phi_{11} & \dots & \phi_{1n} \\ \dots & \dots & \dots \\ \phi_{n1} & \dots & \phi_{nn} \end{bmatrix} \quad (5.44)$$

называют переходной или фундаментальной матрицей. Заметим, что фундаментальная матрица удовлетворяет однородному уравнению состояния системы при $\mathbf{u}(t) = 0$, $\Phi(0) = I$.

Так как уравнение (5.42) представляет собой систему линейных неоднородных дифференциальных уравнений, то его общее решение представляет собой сумму общего решения соответствующей однородной системы и частного решения неоднородной системы. Первое решение часто называют свободным, а второе – вынужденным. Тогда можно записать

$$x(t) = x_{\text{св}}(t) + x_{\text{в}}(t) = \Phi(t)x(0) + \int_0^t \Phi(t - \tau) \mathbf{B} \mathbf{u}(\tau) d\tau. \quad (5.45)$$

Введенная переходная матрица (5.44) состояния линейной динамической системы с постоянными параметрами имеет следующие свойства

1. $\Phi(t_1 - t_2)\Phi(t_2 - t_3) = \Phi(t_1 - t_3)$ для $\forall t_1, t_2, t_3$.
2. $\Phi(0) = I$.
3. $\Phi(t)$ невырожденная матрица для конечномерной матрицы $\mathbf{A}$.
4. $\Phi^{-1}(t) = \Phi(-t)$.

Переходная матрица состояния (5.44), входящая в искомое решение вида (5.45) может быть вычислена различными методами, некоторые из которых мы сейчас рассмотрим.

1. Метод разложения в бесконечный степенной ряд.

Матрица $\Phi(t)$ представляется в виде степенного ряда.

$$\Phi(t) = e^{At} = I + \frac{At}{1} + \frac{A^2 t^2}{2!} + \dots$$

Члены этого выражения могут вычисляться итеративно.

Например, каждый член ряда равен:

$$\frac{A^k t^k}{k!}$$

а $(k + 1)$ -ый член ряда равен

$$\frac{A^{k+1} t^{k+1}}{(k + 1)!}$$

Сравнивая эти выражения, можно записать

$$[(k + 1) - \text{й}] = \frac{A^1 t^1}{(k + 1)} [k - \text{ый}]$$

где $k = 0, 1, 2, \dots$

Обычно при вычислении длины ряда проводится проверка на сходимость, и итерации могут прекращаться после N членов.

2. Метод разложения в ограниченный степенной ряд (теорема Кэли-Гамильтона).

Теорема Кэли-Гамильтона утверждает, что каждая квадратная матрица должна удовлетворять своему характеристическому уравнению.

Напомним, что **характеристическим уравнением** матриц A , является уравнение, получаемое приравниванием к нулю определителя матрицы $\lambda I - A$, т.е.:

$$|\lambda I - A| = 0$$

При этом корни характеристического уравнения называются **собственными значениями** матрицы A .

Собственные значения обладают следующими важными свойствами:

1. Если коэффициенты характеристического уравнения действительны, то собственные значения либо действительны, либо образуют комплексно-сопряженные пары.
2. След матрицы (сумма элементов главной диагонали) равен сумме собственных значений матрицы.

$$tr(A) = \sum_{i=1}^n a_{ii} = \lambda_1 + \lambda_2 + \dots + \lambda_n$$

3. Определитель матрицы A равен произведению собственных значений.

$$\det(A) = \prod_{i=1}^n \lambda_i$$

4. Если матрица A невырождена и имеет собственные значения $\lambda_i, i = 1, 2, \dots, n$, то $\frac{1}{\lambda_i}, i = 1, 2, \dots, n$ - собственные значения матрицы A^{-1} .

5. Если λi – собственное значение матрицы A , то оно же является собственным значением матрицы A' (или A^T) (транспонированной к A).

6. Если A – действительная симметрическая матрица, то все ее собственные значения действительны.

7. Для квадратных матриц A и B справедливо следующее

$$|\lambda I - AB| = |\lambda I - BA|$$

Для иллюстрации применения теоремы Кэли-Гамильтона предположим, что матрица A имеет вид

$$A = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}$$

Тогда характеристическое уравнение для этой матрицы представляется как:

$$|\lambda I - A| = \det \begin{bmatrix} \lambda - 3 & -2 \\ -2 & \lambda - 3 \end{bmatrix}$$

Откуда

$$(\lambda - 3)^2 - 4 = 0$$

$$\lambda^2 - 6\lambda + 5 = 0$$

Применяя теорему Кэли-Гамильтона, получаем матричное управление

$$\mathbf{A}^2 - 6\mathbf{A} + 5\mathbf{I} = 0,$$

из которого имеем

$$\mathbf{A}^2 = 6\mathbf{A} - 5\mathbf{I}.$$

Поэтому

$$\mathbf{A}^3 = \mathbf{A}^2\mathbf{A} = (6\mathbf{A} - 5\mathbf{I})\mathbf{A} = 6\mathbf{A}^2 - 5\mathbf{A} = 6(6\mathbf{A} - 5\mathbf{I}) - 5\mathbf{A} = 31\mathbf{A} - 30\mathbf{I}$$

Аналогичным способом может быть вычислена любая степень матрицы $\mathbf{A}$.

5.8. Однонаправленные цепи и цепи с обратной связью

До сих пор при рассмотрении уравнений четырехполюсников предполагалось, что они обладают двумя основными свойствами.

1. выходная (управляемая) величина пропорциональна входной (управляющей) величине.

2. входная величина не зависит от выходной.

Четырехполюсники, обладающие такими свойствами называются **однонаправленными**. Они обеспечивают передачу энергии только в одном направлении (от входа к выходу); передача энергии в обратном направлении в однонаправленных ЧП полностью отсутствует.

При соединении нескольких четырехполюсников в единую сеть ситуация может коренным образом измениться.

Рассмотрим сначала принципиально возможные способы соединения четырехполюсников в единую цепь и способы получения уравнений для описания такой составной цепи.

Составным называется четырехполюсник, который может быть представлен как соединение нескольких более простых (элементарных) четырехполюсников. Если при соединении четырехполюсников не происходит изменения соотношений между напряжениями и токами на зажимах, то такие соединения называют регулярными.

Последовательное соединение четырехполюсников. Иногда такое соединение называют последовательно-последовательным, подчеркивая, что последовательно соединены как входные, так и выходные зажимы четырехполюсников.

Схема последовательно-последовательного соединения представлена на рис.5.15.

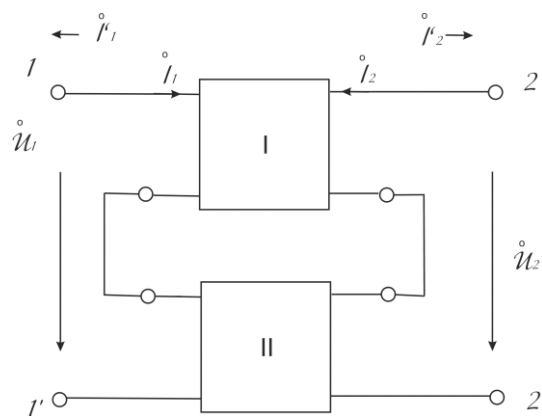


Рис.5.15. Последовательно-последовательное соединение четырехполюсников

Параллельное соединение четырехполюсников иногда называют параллельно-параллельным, акцентируя внимание на способах соединения входных и выходных зажимов.

Для описания такого типа соединений удобно использовать Z-форму представления четырехполюсников и переходить к составному векторному описанию соединения:

$$\begin{bmatrix} U_1^{\circ} \\ U_2^{\circ} \end{bmatrix} = ([Z_1] + [Z_2]) \begin{bmatrix} I_1^{\circ} \\ I_2^{\circ} \end{bmatrix} = [Z] \begin{bmatrix} I_1^{\circ} \\ I_2^{\circ} \end{bmatrix}. \quad (5.46)$$

Доказать справедливость уравнения (5.46) рекомендуется провести самостоятельно в качестве дополнительного упражнения.

Схема параллельно-параллельного соединения четырехполюсника приведена на рис.5.16.

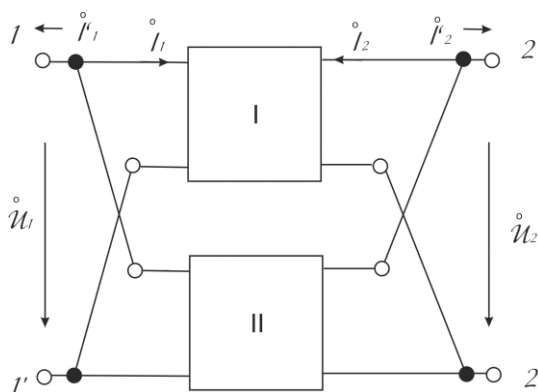


Рис.5.16. Параллельно-параллельное соединение четырехполюсника

Для описания такого типа соединений удобно использовать Y-форму представления четырехполюсников и переходить к составному векторному описанию соединения:

$$\begin{bmatrix} I_1^{\circ} \\ I_2^{\circ} \end{bmatrix} = ([Y_1] + [Y_2]) \begin{bmatrix} U_1^{\circ} \\ U_2^{\circ} \end{bmatrix} = [Y] \begin{bmatrix} U_1^{\circ} \\ U_2^{\circ} \end{bmatrix}. \quad (5.47)$$

Доказать справедливость уравнения (5.47) рекомендуется провести самостоятельно в качестве дополнительного упражнения.

Одним из наиболее распространенных типов соединений четырехполюсников является каскадное соединение, при котором выходные зажимы первого четырехполюсника соединяется с входными зажимами второго.

Схема каскадного соединения четырехполюсников приведена на рис.5.17.

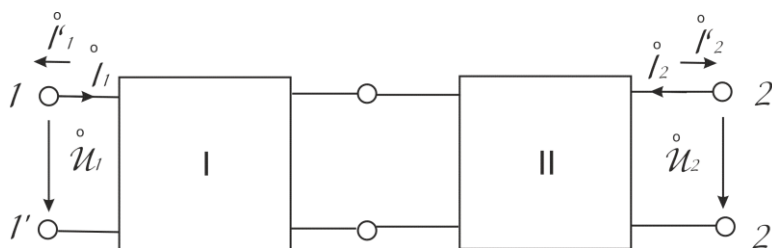


Рис.5.17. Каскадное соединение четырехполюсников

Для описания такого типа соединений удобно использовать Y-форму представления четырехполюсников и переходить к составному векторному описанию соединения:

$$\begin{bmatrix} U_1^o \\ I_1^o \end{bmatrix} = ([A_1][A_2]) \begin{bmatrix} U_2^o \\ I_2^o \end{bmatrix} = [A] \begin{bmatrix} U_2^o \\ I_2^o \end{bmatrix}. \quad (5.48)$$

Доказать справедливость уравнения (5.48) рекомендуется провести самостоятельно в качестве дополнительного упражнения.

Все реально существующие устройства не являются полностью однонаправленными.

Невзаимные ЧП, в которых наряду с преимущественной передачей энергии в одном направлении имеет место также передача энергии в противоположном направлении, называется **ЧП с обратной связью**.

Обратная связь может быть **внутренней** (следствие не идеальности ЧП) и **внешней** (в случае подключения к ЧП внешней цепи обратной связи).

В зависимости от частотных свойств цепи обратной связи (ОС) различают частотно-зависимую ОС и частотно-независимую ОС.

Последовательно-параллельная схема соединения двух ЧП с образованием ОС по напряжению приведена на рис.5.18.

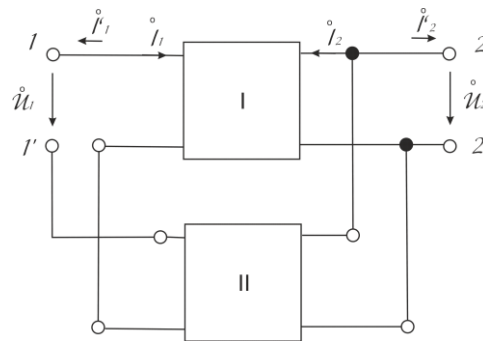


Рис.5.18. Соединение четырехполюсников с использованием обратной связи

Структурная схема соединения четырехполюсников по принципу обратной связи приведена на рис.5.19.

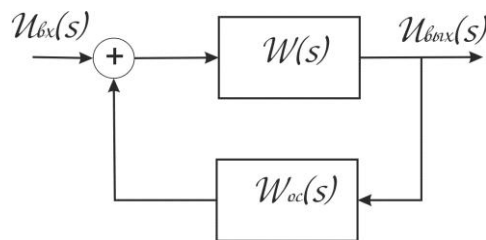


Рис.5.19. Структурная схема соединения элементов цепей по принципу обратной связи

Для упрощения анализа будем считать, что вспомогательный ЧП обратной связи, также как и основной, является однонаправленным.

В этом случае

$$U_{\text{вх}\Sigma}(s) = U_{\text{вх}}(s) + U_{\text{ос}}(s) = U_{\text{вх}}(s) + W_{\text{ос}}(s)U_{\text{вых}}(s)$$

$$U_{\text{вых}}(s) = W(s)U_{\text{вх}}(s).$$

Тогда очевидно, что

$$U_{\text{вых}}(s) = W(s)(U_{\text{вх}}(s) + W_{\text{ос}}(s)U_{\text{вых}}(s))$$

откуда

$$W_{\Sigma}(s) = \frac{W(s)}{1 - W(s)W_{\text{ос}}(s)} \quad (5.49)$$

или для частотной передаточной функции

$$W_{\Sigma}(j\omega) = \frac{W(j\omega)}{1 - W(j\omega)W_{\text{ос}}(j\omega)} \quad (5.50)$$

Из полученных уравнений (5.49), (5.50) видно, что модуль $|W_{\Sigma}(j\omega)|$ может оказаться как больше $|W(j\omega)|$, так и меньше его, что определяется знаком алгебраической суммы на рис.5.19. В первом случае ОС называется **положительной**, во втором - **отрицательной**.

Во многих практических случаях

$$|W(j\omega)W_{\text{ос}}(j\omega)| \gg 1$$

|
тогда

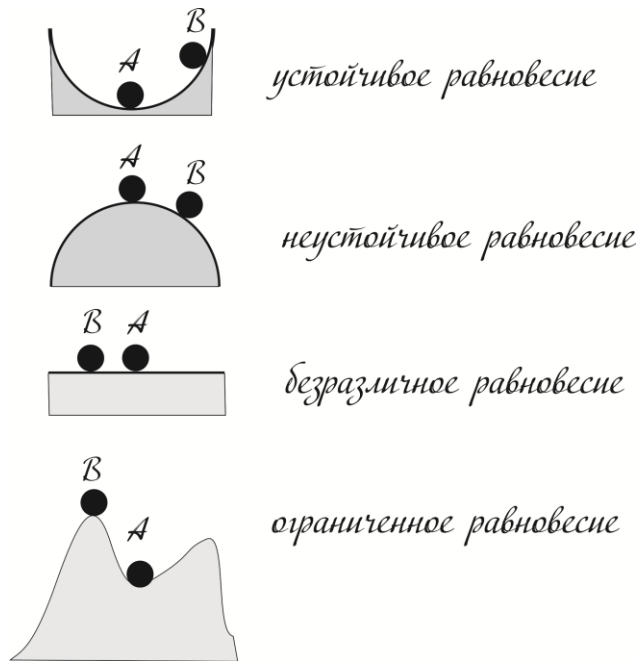
$$W_{\Sigma}(j\omega) \approx - \frac{1}{W_{\text{ос}}(j\omega)}.$$

5.9. Общие сведения об устойчивости линейных цепей

На любую физическую систему всегда действуют различные внешние возмущения, которые могут нарушить ее нормальную работу. Правильно спроектированная система должна устойчиво работать всех внешних возмущениях.

В простейшем случае понятие устойчивости системы связано со способностью ее возвращаться (с определенной точностью) в состояние равновесия после исчезновения внешних сил, которые вывели ее из этого равновесия.

На рис.5.20. приведена наглядная иллюстрация возможных вариантов равновесия на примере механической системы.



В приведенном примере вопрос об устойчивости решается наглядно и довольно просто. Однако в общем случае далеко не всегда ясно, при каких условиях равновесное положение системы будет устойчивым.

Дифференциальное уравнение линейной системы имеет вид

Рис.520. Примеры равновесных состояний системы

$$a_0 \frac{d^n}{dt^n} y(t) + a_1 \frac{d^{n-1}}{dt^{n-1}} y(t) + \dots + a_n y(t) = b_0 \frac{d^m}{dt^m} u(t) + b_1 \frac{d^{m-1}}{dt^{m-1}} u(t) + \dots + b_m u(t).$$

Для ответа на вопрос, каков при этом будет сигнал на выходе системы, требуется решить дифференциальное уравнение

$$(a_0 p^n + a_1 p^{n-1} + \dots + a_n) y(t) = (b_0 p^m + b_1 p^{m-1} + \dots + b_m) u(t).$$

Общее решение данного уравнения имеет вид

$$y(t) = y_c(t) + y_b(t),$$

где $y_c(t)$ - общее решение однородного дифференциального уравнения, определяющее «свободную» составляющую движения системы (переходный процесс),

$y_b(t)$ - частное решение неоднородного дифференциального уравнения, которое определяет так называемое «вынужденное» решение.

Свободные колебания находятся решением однородного уравнения

$$(a_0 p^n + a_1 p^{n-1} + \dots + a_n) y_c(t) = 0.$$

Поскольку полезной в данном случае является составляющая выходного сигнала, обусловленная воздействием на цепь входного сигнала, то естественным условие устойчивости цепи будет

$$\lim_{t \rightarrow \infty} y_c(t) = 0.$$

В случае если все корни характеристического уравнения простые, то выполнение условия асимптотического затухания свободной составляющей очевидно: корни должны иметь отрицательную вещественную часть. В случае наличия кратных корней λ_k , как известно в решении появляются слагаемые вида $C_k t^{k-1} e^{\lambda_k t}$, однако при отрицательности вещественной части корня эти составляющие также являются затухающими.

Так как рассмотренное характеристическое уравнение соответствует знаменателю передаточной функции линейной цепи, то условие устойчивости можно сформулировать следующим образом: для устойчивости линейной цепи, имеющей передаточную функцию $W(s)$, необходимо, чтобы полюсы этой передаточной функции располагались в левой полуплоскости комплексной переменной s .

Эффект возрастания величины собственных колебаний в электрических цепях возможен либо при использовании активных элементов, либо при соединении нескольких каскадов по схеме с обратной связью.

Современная вычислительная техника позволяет сравнительно просто вычислять корни любого заданного характеристического уравнения цепи. Однако иногда бывает крайне полезным иметь возможность оценить устойчивость системы без решения характеристического уравнения. Поэтому важное значение приобретают правила, которые позволяют определить устойчивость системы без вычисления корней. Эти правила называют **критериями устойчивости**.

Критерии устойчивости могут быть разделены на **алгебраические** и **частотные**. С математической точки зрения все критерии устойчивости эквивалентны, однако в зависимости от конкретной задачи проводить исследование устойчивости может оказаться более просто тем или другим методом.

В задачи нашего курса не входит подробное изучение критериев устойчивости цепей, поэтому мы ограничимся только краткими замечаниями по этому поводу.

Из алгебраических критериев устойчивости наиболее широкое распространение получили критерии устойчивости Рауса и Гурвица.

При этом следует отметить, что необходимым условием устойчивости линейной системы является положительность всех коэффициентов ее характеристического уравнения (для систем первого и второго порядка это необходимое условие одновременно является и достаточным).

Для систем более высокого порядка критерий Гурвица основан на построении определителя Гурвица в виде

$$\Delta_n = \begin{vmatrix} a_1 & a_3 & a_5 & \cdot & \cdot & \cdot \\ a_0 & a_2 & a_4 & \cdot & \cdot & \cdot \\ 0 & a_1 & a_3 & \cdot & \cdot & \cdot \\ 0 & a_0 & a_2 & a_4 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & a_n \end{vmatrix}. \quad (5.51)$$

Отчерчивая в главном определителе Гурвица (5.51) дополнительные миноры можно получить определители Гурвица низшего порядка.

Критерий устойчивости Гурвица гласит, что для того чтобы линейная система была устойчива необходимо и достаточно, чтобы все определители Гурвица имели знаки, одинаковые со знаком первого коэффициента характеристического уравнения. То есть были положительными при $a_0 > 0$

$$\Delta_1 > 0, \Delta_2 > 0, \dots, \Delta_n > 0. \quad (5.52)$$

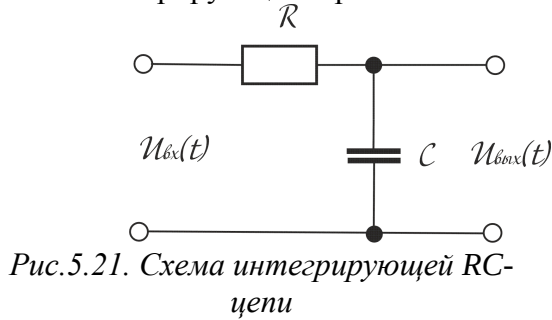
Критерий устойчивости (5.52) является алгебраическим критерием. Частотные критерии устойчивости позволяют судить об устойчивости систем по виду их частотных характеристик. При этом некоторые из них, как например, критерий Михайлова позволяет судить об устойчивости непосредственно анализируемой системы, а другие, как например, критерий устойчивости Найквиста, позволяет судить об устойчивости системы, которая получится при замыкании исследуемой системы единичной обратной связью.

5.10. Частотные и временные характеристики простейших RC и RL цепей

Простейшие цепи, состоящие только из резисторов и конденсаторов, весьма широко применяются в радиотехнике для преобразования сигналов. Поэтому на их

характеристиках мы остановимся более подробно на примере цепей состоящих всего из двух элементов.

Схема интегрирующей простейшей RC-цепи представлена на рис.5.21.



Используя операторный метод записи уравнений цепи, получим

$$W(s) = \frac{\frac{1}{sC}}{\frac{1}{sC} + R} = \frac{1}{1 + sRC} = \frac{1}{1 + \tau s},$$

где $\tau = RC$ – постоянная времени цепи.

Воспользовавшись формулами преобразования Лапласа для импульсной характеристики цепи легко получить

$$g(t) = \mathcal{L}^{-1}\{W(s)\} = \frac{1}{\tau} \mathcal{L}^{-1}\left\{\frac{1}{s + \frac{1}{\tau}}\right\} = \frac{1}{\tau} e^{-t/\tau},$$

что с учетом области определения импульсной характеристики дает

$$g(t) = \begin{cases} \frac{1}{\tau} e^{-t/\tau}, & t \geq 0, \\ 0, & t < 0. \end{cases}$$

Аналогично получаем переходную характеристику цепи

$$h(t) = \mathcal{L}^{-1}\{W(s)/s\} = 1 - e^{-t/\tau}.$$

или с учетом области определения

$$h(t) = \begin{cases} 1 - e^{-t/\tau}, & t \geq 0, \\ 0, & t < 0. \end{cases}$$

Графики импульсной и переходной характеристик данной цепи представлены на рис.5.22.

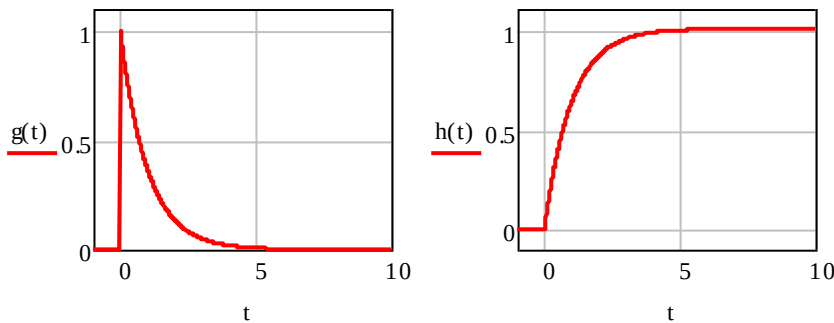


Рис.5.22. Импульсная и переходная характеристики цепи

Произведя замену переменных ($s = j\omega$) в передаточной функции получаем частотную передаточную функцию (комплексный коэффициент передачи), что позволяет определить соответствующие амплитудно-частотную и фазочастотную характеристики исследуемой цепи.

$$W(j\omega) = \frac{1}{j\omega\tau + 1} = \frac{1}{\sqrt{1 + \tau^2\omega^2}} e^{-j\arctg(\tau\omega)},$$

то есть

$$A(\omega) = \frac{1}{\sqrt{1 + \tau^2\omega^2}},$$

$$\varphi(\omega) = -\arctg(\tau\omega).$$

Графики полученных характеристик представлены на рис. 5.23.

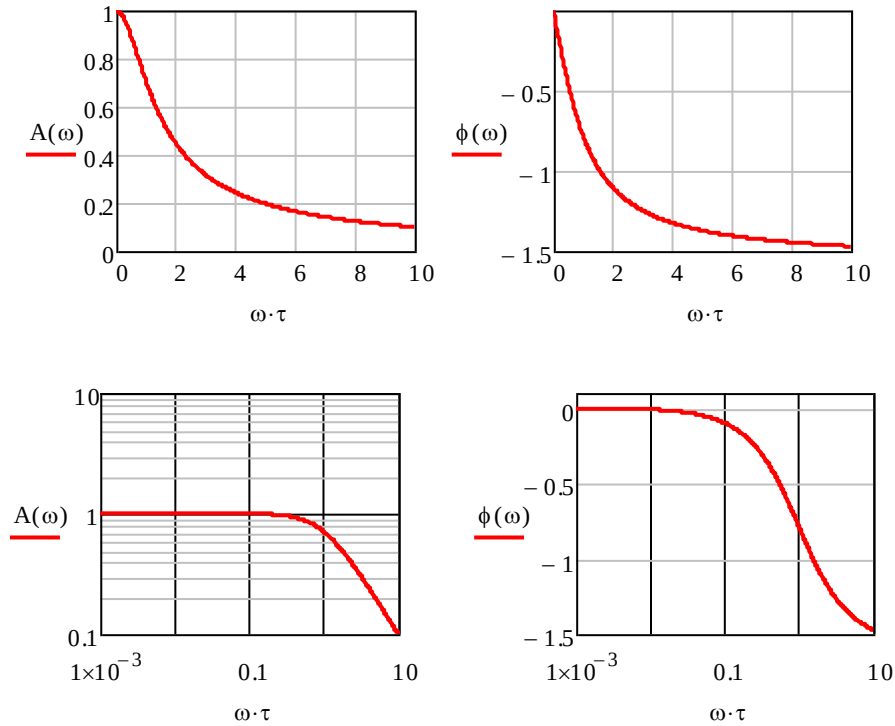
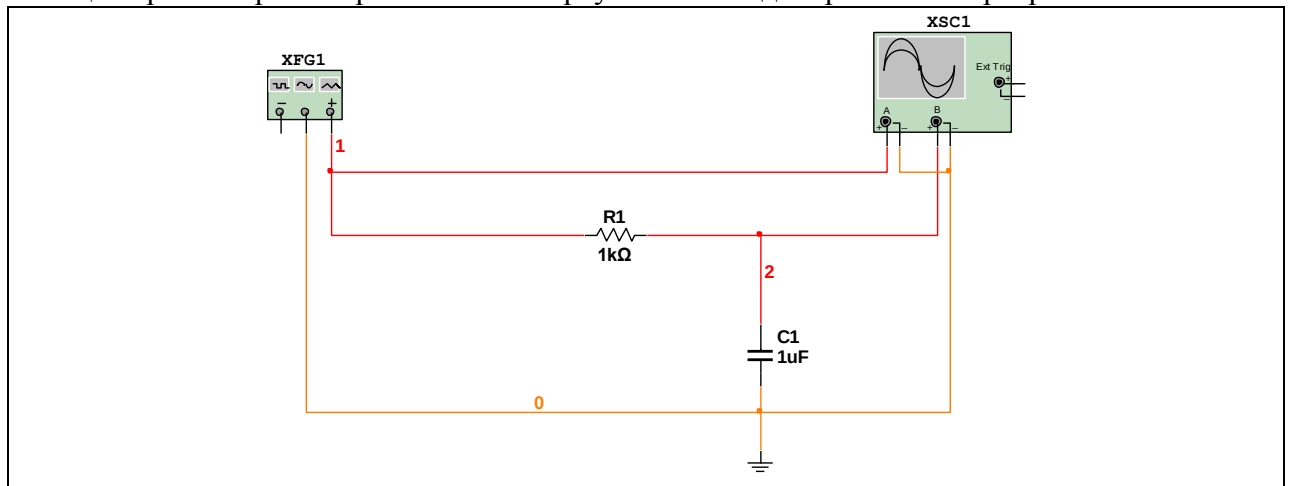


Рис.5.23. Амплитудно-частотная и фазо-частотная характеристики цепи, построенные в линейном и логарифмическом масштабе.

Пример 5.13.

Прохождение импульсного сигнала (меандра) через простейшую интегрирующую RC-цепь рассмотрим с применением виртуального моделирования в программе Multisim.



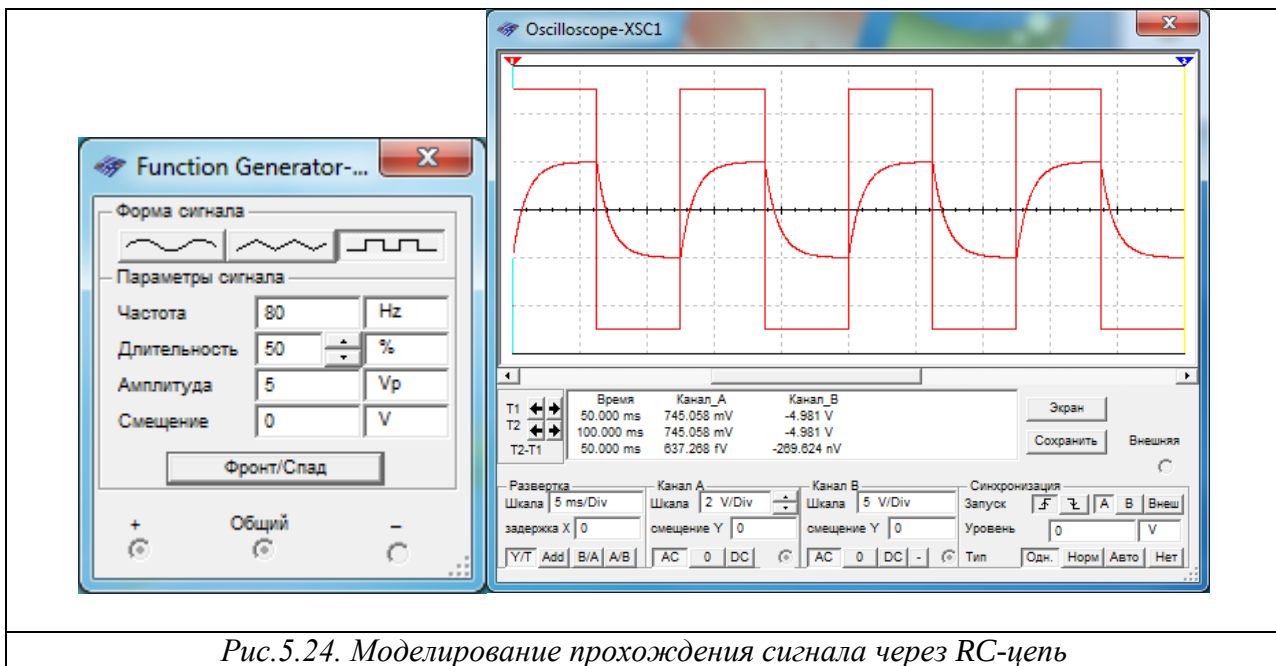


Рис.5.24. Моделирование прохождения сигнала через RC-цепь

Если резистор и емкость соединить как показано на рис.5.25, то получившаяся цепь будет обладать принципиально иными характеристиками.

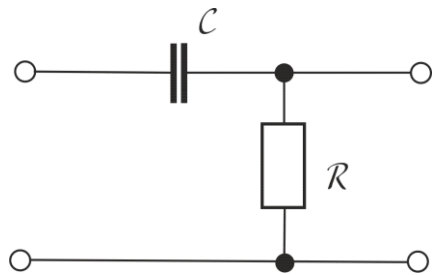


Рис.5.25. Дифференцирующая RC-цепь

Передаточная функция для данной цепи принимает вид

$$W(s) = \frac{R}{\frac{1}{sC} + R} = \frac{sRC}{1 + sRC} = \frac{s\tau}{1 + s\tau},$$

где $\tau = RC$ – постоянная времени цепи. Для частотной передаточной функции (комплексного коэффициента передачи) запишем

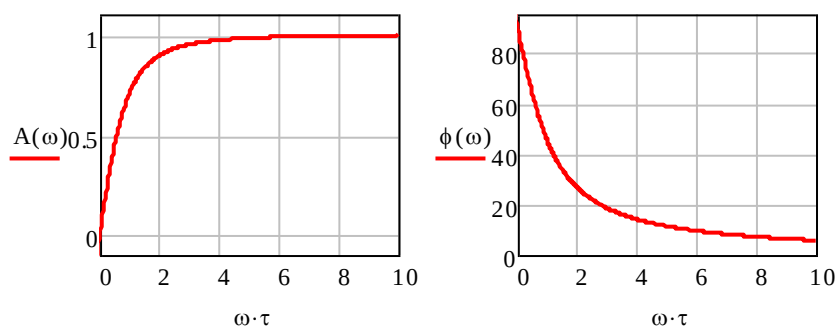
$$W(j\omega) = \frac{j\omega\tau}{1 + j\omega\tau} = A(\omega)e^{j\varphi(\omega)},$$

где

$$A(\omega) = \frac{\omega\tau}{\sqrt{1 + \tau^2\omega^2}},$$

$$\varphi(\omega) = \begin{cases} \frac{\pi}{2} - \arctg(\tau\omega), & \omega > 0 \\ -\frac{\pi}{2} + \arctg(\tau\omega), & \omega < 0 \end{cases}.$$

Графики соответствующих частотных характеристик приведены на рис.5.26.



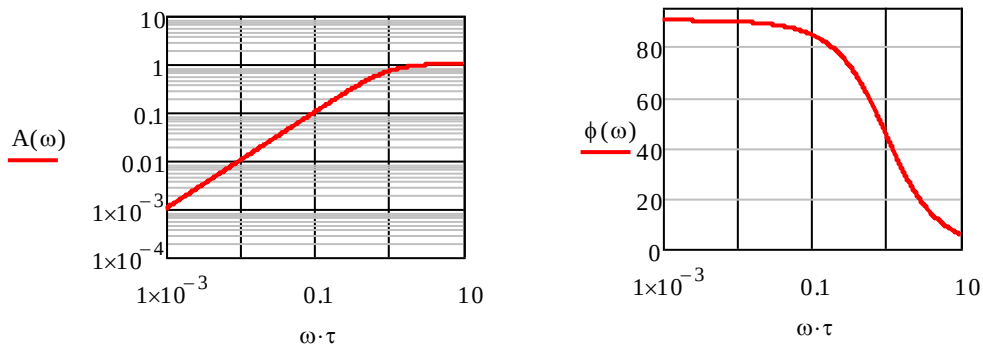


Рис.5.26. Частотные характеристики дифференцирующей RC-цепи в линейном и логарифмическом масштабе.

Для импульсной и переходной характеристик данной цепи нетрудно получить

$$g(t) = \begin{cases} \delta(t) - \frac{1}{\tau} e^{-t/\tau}, & t \geq 0, \\ 0, & t < 0 \end{cases}$$

и

$$h(t) = \begin{cases} 1 - (1 - e^{-t/\tau}), & t \geq 0, \\ 0, & t < 0. \end{cases}$$

Соответствующие графики приведены на рис.5.27.

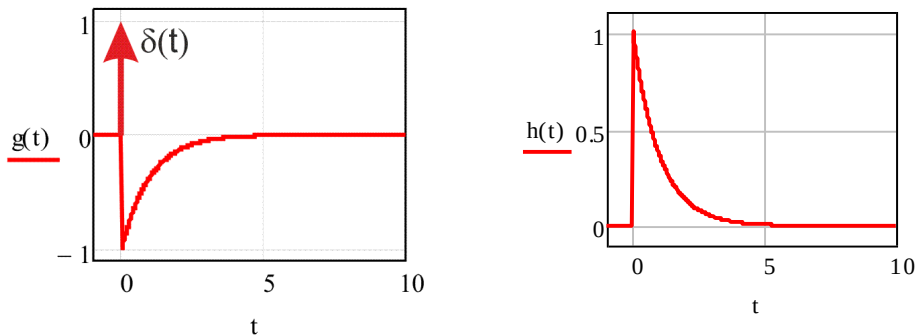


Рис.5.27. Импульсная и переходная характеристики дифференцирующей RC-цепи.

Пример 5.14.

Прохождение импульсного сигнала (меандра) через простейшую дифференцирующую RC-цепь рассмотрим с применением виртуального моделирования в программе Multisim.

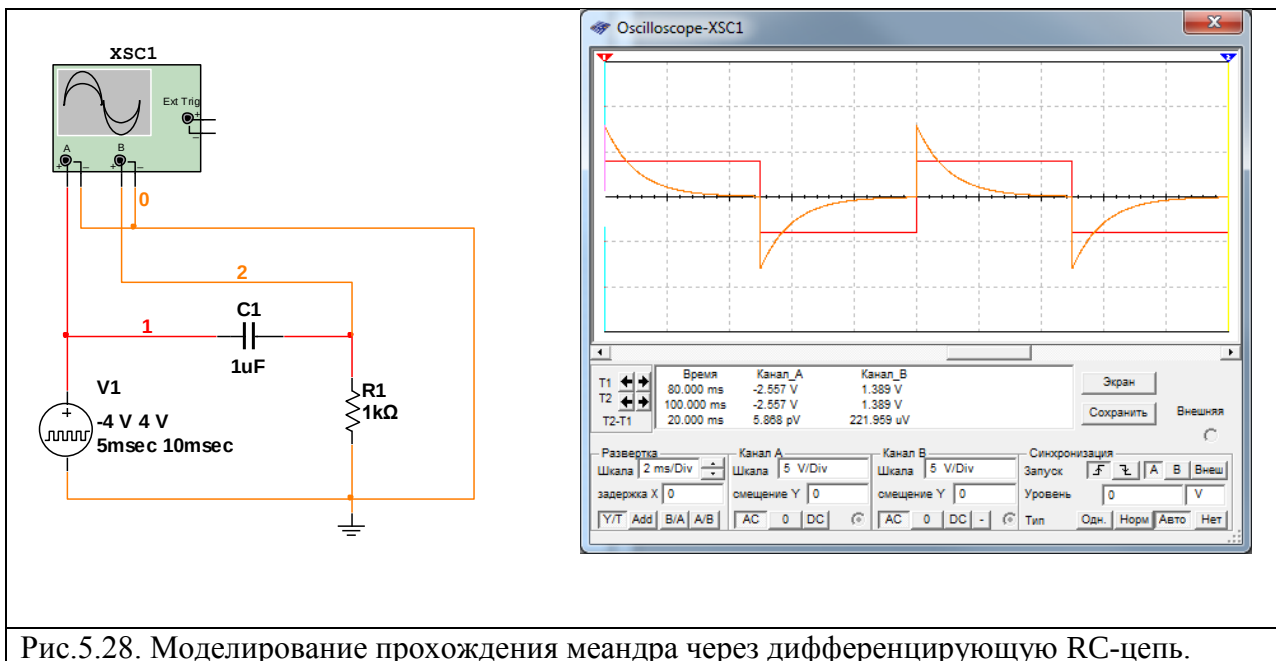


Рис.5.28. Моделирование прохождения меандра через дифференцирующую RC-цепь.

Аналогичные характеристики можно получить при использовании RL-цепей. Схема простейшей RL-цепи приведена на рис.5.29.

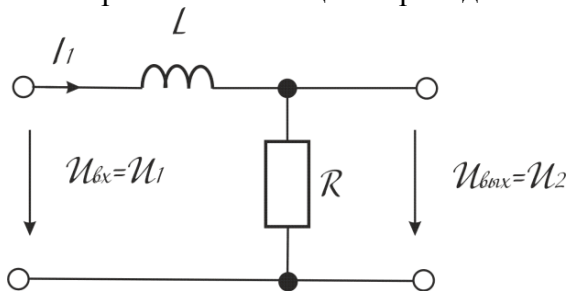


Рис.5.29. Схема простейшей RL-цепи

Передаточная функция данной цепи легко определяется из выражения

$$W(s) = \frac{R}{R + sL} = \frac{1}{1 + s\frac{L}{R}} = \frac{1}{1 + s\tau},$$

где $\tau = \frac{L}{R}$ – постоянная времени RL-цепи.

Данное выражение полностью соответствует передаточной функции интегрирующей RC-цепи, полученной ранее. Поэтому и все характеристики цепи будут схожими с поправкой на уравнения, задающие значение постоянной времени[^]

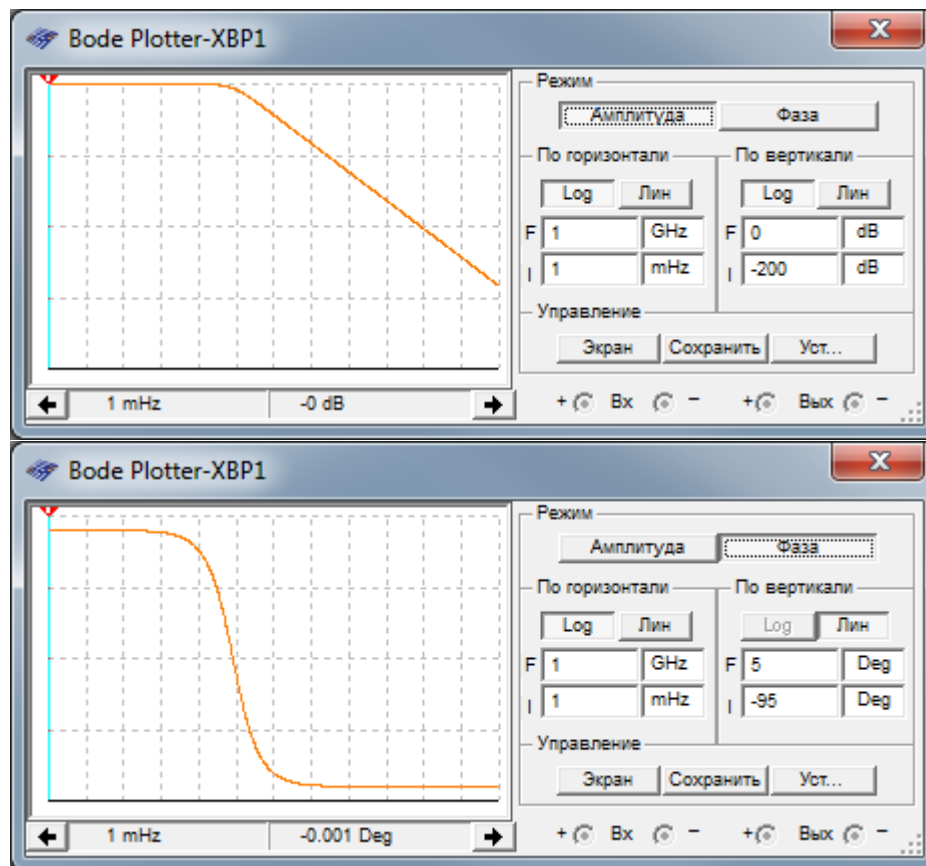
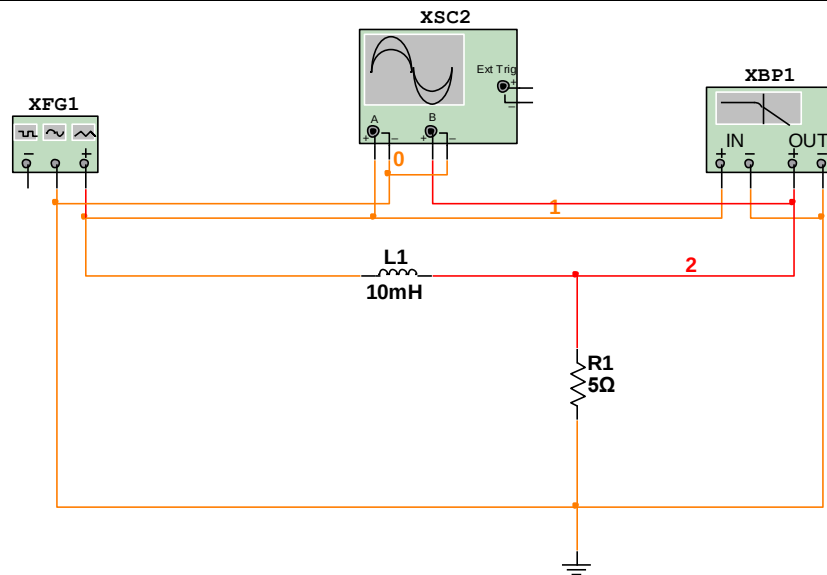
$$\tau_{RC} = RC,$$

$$\tau_{RL} = \frac{L}{R}.$$

Если в приведенной схеме поменять местами резистор и индуктивность, то обнаружится аналогичная связь с характеристиками дифференцирующей RC-цепи.

Пример 5.15.

Процесс экспериментального определения частотных характеристик RL-цепи, а также прохождение импульсного сигнала (меандра) через простейшую интегрирующую RL-цепь рассмотрим с применением виртуального моделирования в программе Multisim (см.рис.5.30)



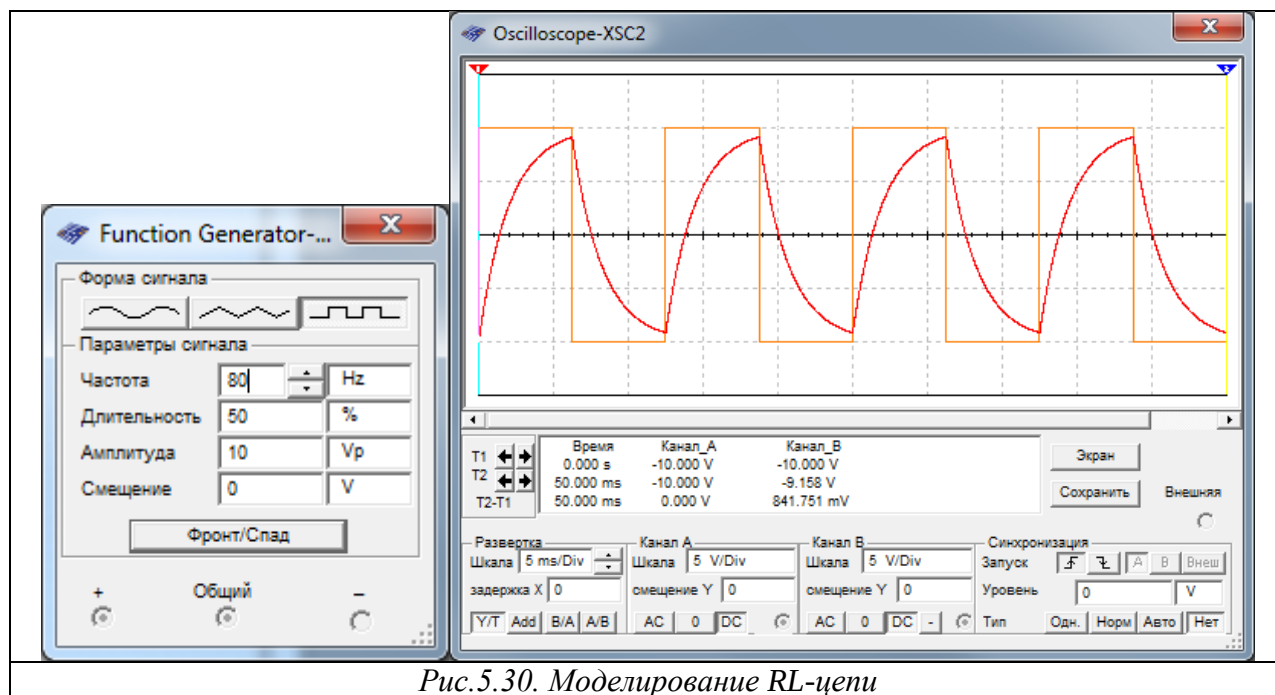


Рис.5.30. Моделирование RL-цепи

5.11. Изолированный колебательный контур и его характеристики

5.11.1. Последовательный колебательный контур

В предыдущем разделе мы рассматривали характеристики цепей, содержащих один реактивный элемент. Если цепь содержит и емкостной, и индуктивный элементы, то в такой цепи может возникнуть явление резонанса, которое широко используется для выделения или задержания определенной полосы частот сигнала.

Под резонансом в электрической цепи понимают любой режим работы цепи, содержащей емкости и индуктивности, при котором ее входное сопротивление имеет чисто резистивный характер, и, следовательно, **сдвиг фаз между током и напряжением на входе цепи равен нулю**.

Отметим, что данное определение является фазовым критерием резонанса. Кроме фазового критерия так же существует и амплитудный. Согласно которому явление резкого возрастания амплитуды отклика цепи при приближении частоты внешнего воздействия к определенному значению называется резонансом. Амплитудный критерий заимствован из механики, где он широко используется.

Частоты, соответствующие фазовому критерию и амплитудному критерию, вообще говоря, могут и не совпадать, поэтому для устранения неоднозначности в электрических цепях принято использовать фазовый критерий.

В теории цепей различают резонанс напряжений (последовательный резонанс) и резонанс токов (резонанс в параллельной цепи).

Начнем рассмотрение с простого случая последовательного соединения резистора, индуктивности и емкости, как показано на рис.5.31.

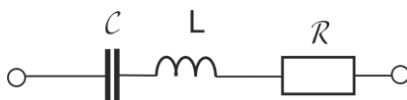


Рис.5.31. Последовательный колебательный контур

Представленная цепь является двухполюсником и для нее можно определить комплексное сопротивление как сумму комплексных сопротивлений элементов:

$$Z(\omega) = R + j\omega L + \frac{1}{j\omega C} = R + j\frac{\omega^2 LC - 1}{\omega C}.$$

В соответствии с определением резонанса

$$\operatorname{Im}(Z(\omega))|_{\omega=\omega_0} = 0,$$

где ω_0 – резонансная частота.

Для резонансной частоты, таким образом, нетрудно получить

$$\omega_0 L = \frac{1}{\omega_0 C}.$$

Откуда имеем

$$\omega_0 = \frac{1}{\sqrt{LC}}. \quad (5.53)$$

При этом очевидно, что на резонансной частоте (5.53) сопротивление емкости и индуктивности равны между собой.

$$Z_C|_{\omega=\omega_0} = |x_C|_{\omega=\omega_0} = \frac{1}{\omega_0 C} = \rho \quad (5.54)$$

$$Z_L|_{\omega=\omega_0} = |x_L|_{\omega=\omega_0} = \omega_0 L = \rho, \quad (5.55)$$

где $\rho = \sqrt{\frac{L}{C}}$ – характеристическое сопротивление контура.

На резонансной частоте

$$I_m = \frac{U_m}{R}$$

и напряжение на реактивных элементах в соответствии с (5.54), (5.55) определяется в виде

$$U_{C_m}|_{\omega_0} = U_{L_m}|_{\omega_0} = \rho I_m.$$

Отношение амплитуды напряжения на реактивном элементе контура к амплитуде напряжения на контуре на резонансной частоте называется **добротностью контура**.

$$Q = \frac{U_{C_m}}{U_m}|_{\omega=\omega_0} = \frac{U_{L_m}}{U_m}|_{\omega=\omega_0} = \frac{\rho}{R} = \frac{1}{R} \sqrt{\frac{L}{C}} \quad (5.56)$$

Иначе добротность (5.56) контура можно представить в виде

$$Q = \frac{1}{R\omega_0 C} = \frac{\omega_0 L}{R}. \quad (5.57)$$

Используя выражение, связывающее резонансную частоту с параметрами колебательного контура, исходное уравнение для комплексного сопротивления Z можно записать иначе:

$$Z(\omega) = R + j \frac{\omega^2 LC - 1}{\omega C} = R + j \frac{\frac{\omega^2}{\omega_0^2} - 1}{\omega C} = R + j \frac{\omega^2 - \omega_0^2}{\omega_0^2 \omega C} = R + j \frac{(\omega + \omega_0)(\omega - \omega_0)}{\omega_0^2 \omega C}.$$

Наибольший интерес представляет поведение цепи вблизи точки резонанса, тогда можно использовать следующие приближенные выражения

$$\begin{aligned}\omega + \omega_0 &\approx 2\omega_0, \\ \omega_0^2 \omega &\approx \omega_0^3.\end{aligned}$$

С учетом этих допущений получаем

$$Z(\omega) \approx R + j \frac{2\omega_0(\omega - \omega_0)}{\omega_0^3 C} = R(1 + j2 \frac{(\omega - \omega_0)}{\omega_0} \frac{1}{\omega_0 RC}).$$

Теперь, вводя безразмерную величину, называемую относительной расстройкой и характеризующую отклонение частоты от резонансной

$$\varepsilon = \frac{(\omega - \omega_0)}{\omega_0}, \quad (5.58)$$

для комплексного сопротивления цепи с учетом (5.56), (5.57), (5.58) можно записать

$$Z(\omega) = R(1 + j2Q\varepsilon),$$

где Q – добротность контура.

Записывая комплексное сопротивление контура в экспоненциальном виде получаем

$$Z(\omega) = |Z(\omega)|e^{j\arg(Z(\omega))},$$

где $|Z(\omega)| = R\sqrt{1 + 4Q^2\varepsilon^2}$,

$\arg(Z(\omega)) = \arctg(2Q\varepsilon)$.

Графики полученных характеристик представлены на рис.5.32.

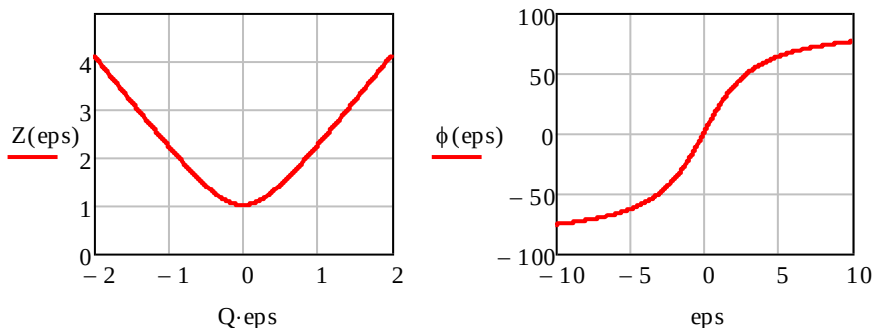


Рис.5.32. Частотная зависимость модуля и фазы комплексного сопротивления контура вблизи точки резонанса

Резонансной характеристикой последовательного колебательного контура принято называть отношение комплексной амплитуды тока на заданной частоте к значению тока в точке резонанса

$$K_p(\omega) = \frac{I_m^o(\omega)}{I_m^o(\omega_0)} = \frac{Z(\omega_0)}{Z(\omega)} = \frac{R}{R(1 + j2Q\varepsilon)} = \frac{1}{1 + j2Q\varepsilon}.$$

Соответственно модуль этого выражения определяется как

$$|K_p(\omega)| = \frac{1}{\sqrt{1 + 4Q^2\varepsilon^2}}. \quad (5.59)$$

График зависимости соответствующий уравнению (5.59) представлен на рис.5.32.

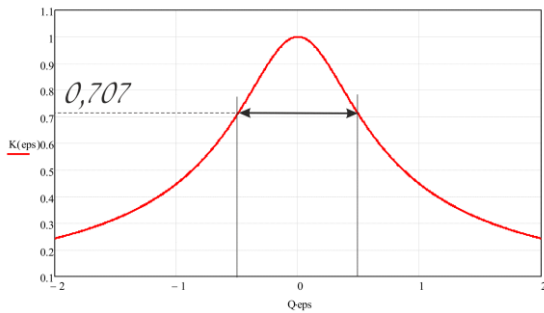


Рис.5.32. Модуль резонансной характеристики последовательного колебательного контура

В радиотехнике принято определять полосу пропускания цепи по уровню $\frac{1}{\sqrt{2}} \approx 0,707$.

Данному уровню соответствует условие $1 + 4Q^2\varepsilon^2 = 2$, откуда

$$\varepsilon_{0,707} = \pm \frac{1}{2Q},$$

то есть полоса пропускания удовлетворяет условию

$$\left| \frac{\omega - \omega_0}{\omega_0} \right| \leq \frac{1}{2Q}.$$

Понятно, что ширина полосы пропускания в этом случае будет равна

$$\Delta\omega = \frac{\omega_0}{Q}.$$

5.11.2. Четырехполюсники на основе последовательного колебательного контура

В большинстве реальных ситуаций колебательный контур используется не как рассмотренный двухполюсник, а как четырехполюсник. Схем построения четырехполюсника на основе последовательного колебательного контура может быть несколько, и их характеристики будут несколько отличаться друг от друга. Простейших вариант построения четырехполюсника на базе последовательного колебательного контура приведен на рис.5.33.

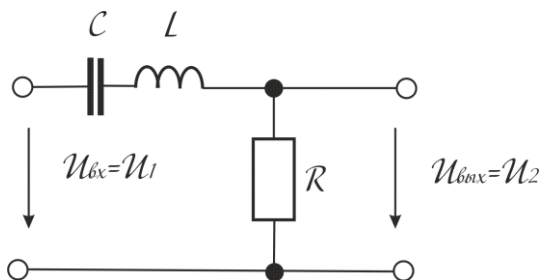


Рис.5.33. Четырехполюсник на основе последовательного колебательного контура.

Передаточная функция данного четырехполюсника может быть записана следующим образом

$$W(s) = \frac{R}{R + sL + \frac{1}{sC}} = \frac{sRC}{s^2LC + sRC + 1}$$

и соответствующая частотная передаточная функция (комплексный коэффициент передачи) определится в виде

$$W(j\omega) = \frac{j\omega RC}{(1 - \omega^2 LC) + j\omega RC}.$$

В качестве дополнительного упражнения рекомендуем произвести вывод уравнения частотной характеристики данного четырехполюсника в зависимости от переменных $Q, \Delta\omega, \omega_0$ по аналогии с действиями проделанными ранее.

Пример 5.16.

Проведем моделирование процедуры определения частотных характеристик четырехполюсника, созданного на основе последовательного колебательного контура с помощью программы Multisim.

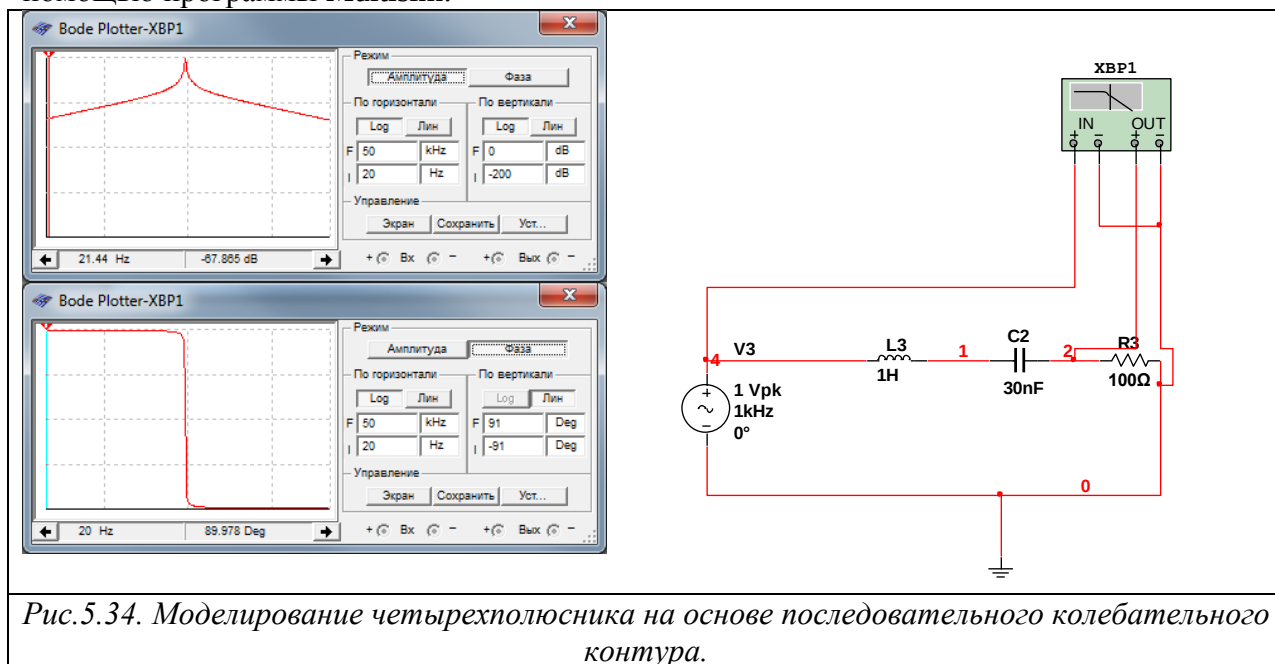


Рис.5.34. Моделирование четырехполюсника на основе последовательного колебательного контура.

5.11.3. Параллельный колебательный контур

Параллельный колебательный контур образуется путем параллельного соединения резистора, конденсатора и катушки индуктивности, как показано на рис.5.35.

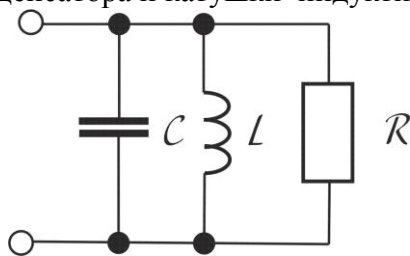


Рис.5.35. Схема параллельного колебательного контура

При рассмотрении параллельного колебательного контура как двухполюсника входной сигнал считается заданным источником тока, а выходным сигналом является напряжение на зажимах данной цепи.

Резонансная частота находится из равенства токов на реактивных элементах, что эквивалентно равенству их реактивных сопротивлений

$$\text{Im}[Y] \Big|_{\omega = \omega_p} = 0$$

$$I_m \left[g + j \left(\omega C - \frac{1}{\omega L} \right) \right] = 0$$

$$\omega_p C - \frac{1}{\omega_p L} = 0$$

$$\omega_p = \frac{1}{\sqrt{LC}}. \quad (5.60)$$

Из приведенных уравнений видно, что резонансная частота параллельного контура ω_p (5.60) равна резонансной частоте последовательного контура (5.53)

$$\omega_0 = \frac{1}{\sqrt{LC}}$$

Аналогично способу, использованному при рассмотрении последовательного колебательного контура, для параллельного колебательного контура можно получить

$$Y(\omega) = \frac{1}{R} (1 + jQ_1 \varepsilon),$$

где $\varepsilon = \frac{(\omega - \omega_0)}{\omega_0}$ – относительная расстройка,

$$Q_1 = \omega_0 RC = \frac{R}{\omega_0 L} = \frac{R}{\frac{L}{\sqrt{C}}} = \frac{R}{\rho} - \text{добротность параллельного контура.}$$

Заметим, что добротность параллельного контура связана с параметрами элементов контура иным соотношением, нежели добротность последовательного контура, что не является неожиданным, так как от рассмотрения комплексного сопротивления мы перешли к рассмотрению комплексной проводимости.

Изменение способа определения добротности в данном случае позволяет согласовать и сохранить справедливость основных уравнений, связывающих частотные параметры контура, добротность и относительную расстройку:

$$\Delta\omega = \frac{\omega_0}{Q_1} \text{ и } \varepsilon_{0,707} = \pm \frac{1}{2Q_1}.$$

5.11.4. Прохождение амплитудно-модулированного сигнала через колебательный контур

Предположим, что к зажимам последовательного колебательного контура приложено напряжение представляющее собой АМ-сигнал

$$x(t) = A_0 (1 + M \cos(\Omega t + \theta_{\Omega})) \cos[\omega_0 t + \theta_0].$$

При этом считаем, что резонансная частота контура и частота несущего колебания АМ-сигнала совпадают.

Ранее мы получили выражение для комплексного сопротивления последовательного колебательного контура для различных значений частот

$$Z(\omega) = |Z(\omega)| e^{j \arg(Z(\omega))},$$

$$\text{где } |Z(\omega)| = R \sqrt{1 + 4Q^2 \varepsilon^2},$$

$$\arg(Z(\omega)) = \arctg(2Q\varepsilon).$$

При рассмотрении амплитудной модуляции мы выяснили, что спектр модулированного сигнала представляет собой три гармонические составляющие соответствующие частотам ω_0 , $\omega_0 - \Omega$, $\omega_0 + \Omega$, как показано на рис. 1.61.

Поскольку ток в последовательном контуре может быть найден по закону Ома

$$I_m^{\circ} = \frac{U_m^{\circ}}{Z(\omega)},$$

то в данном случае требуется определить значение комплексного сопротивления цепи для трех значений частоты: ω_0 , $\omega_0 - \Omega$, $\omega_0 + \Omega$.

Для значения резонансной частоты имеем

$$Z(\omega_0) = R.$$

Для двух боковых частот из уравнения

$$Z(\omega) = R(1 + j2Q\varepsilon),$$

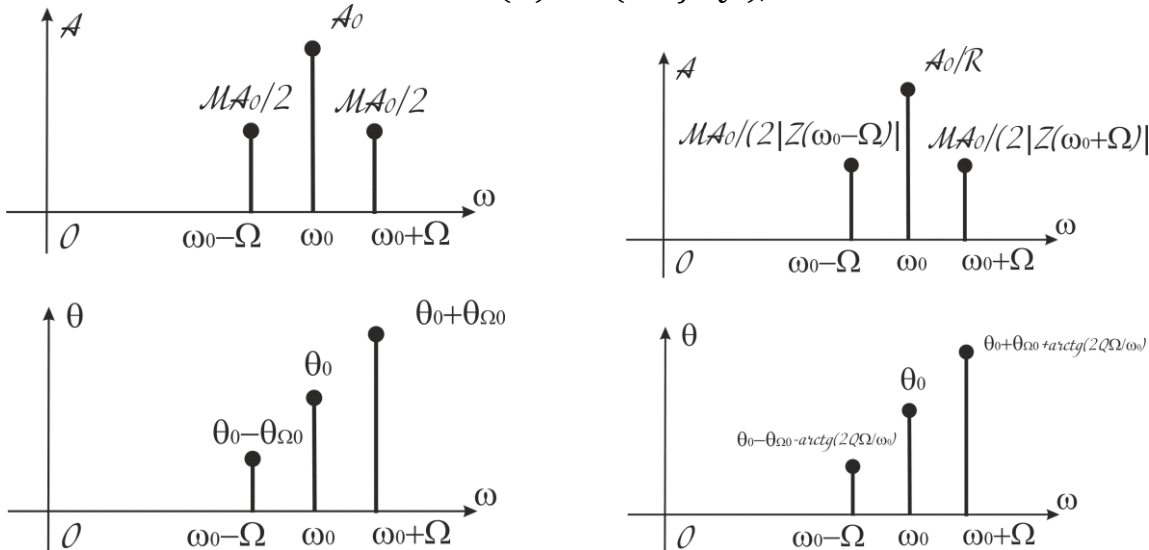


Рис.5.36. Спектр сигнала с тональной модуляцией

получаем

$$Z(\omega_0 \pm \Omega) = R \left(1 \pm j2Q \frac{\Omega}{\omega_0} \right).$$

Спектры получаемого тока в колебательном контуре представлены на рис.5.36, что показывает сохранения типа сигнала (тональный АМ-сигнал). При этом глубина модуляции тока может быть определена по формуле

$$M_I = \frac{M}{\sqrt{1 + \frac{4Q^2\Omega^2}{\omega_0^2}}} = \frac{M}{\sqrt{1 + \frac{4L^2\Omega^2}{R^2}}}.$$

Справедливость данной формулы предлагаем проверить самостоятельно в качестве дополнительного упражнения.

В случае если резонансная частота контура не совпадает с частотой несущего колебания, то процесс анализа становится более трудоемким, однако имеющиеся частотные характеристики контура позволяют произвести необходимые расчеты при любом соотношении указанных частот.

Пример 5.17.

Проведем моделирование процесса прохождения сигнала в виде меандра через четырехполосник, построенный на основе последовательного колебательного контура с помощью программы Multisim. Результаты моделирования представлены на рис. 5.37.

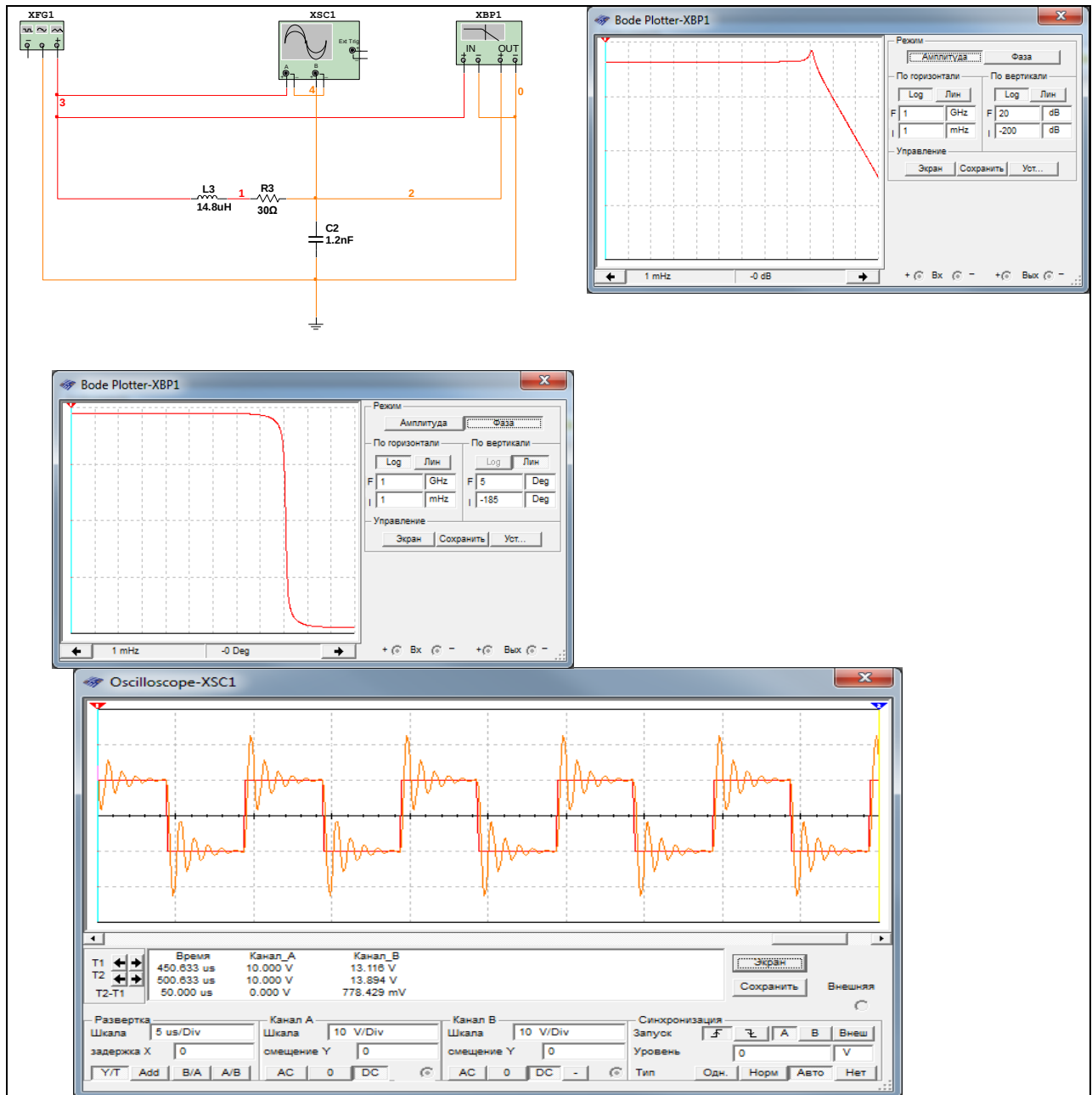


Рис.5.37. Моделирование процесса прохождения меандра через четырехполюсник, созданный на основе последовательного колебательного контура.

5.12. Электрические фильтры

5.12.1. Общие понятия и классификация

Электрическим фильтром называется четырехполюсник, предназначенный для пропускания или задержания сигнала определенной полосы частот.

Диапазон частот, пропускаемых фильтром, называется **полосой пропускания**, остальная область частот – полосой задержания.

Возможны различные способы классификации фильтров:

1. По расположению полосы пропускания фильтры подразделяются на следующие группы:

- 1) фильтры низких частот (полоса пропускания от 0 до ω_c ; ω_c – частота среза).

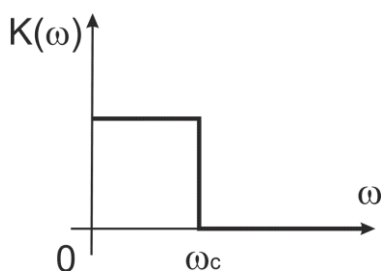


Рис.5.38. Характеристика фильтра нижних частот

- 2) фильтры верхних частот (полоса пропускания от ω_c до ∞).

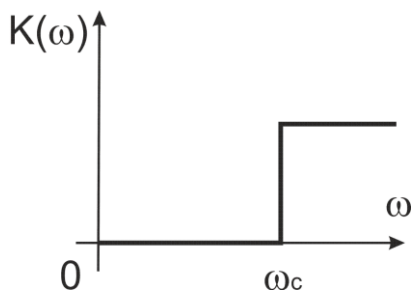


Рис.5.39. Характеристика фильтра верхних частот

- 3) полосовые фильтры (полоса пропускания от ω_{c1} до ω_{c2}).

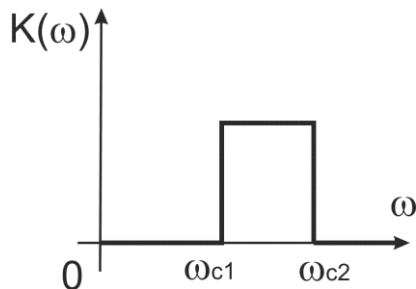


Рис.5.40. Характеристика полосового фильтра

- 4) заграждающие (режекторные) фильтры (полоса задержания от ω_{c1} до ω_{c2}).

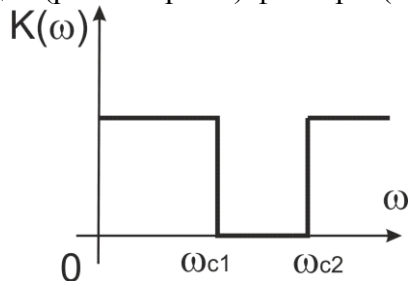


Рис.5.41. Характеристика заграждающего (режекторного) фильтра

2. По типу элементов, входящих в состав фильтра электрические фильтры подразделяются на:

1. реактивные (LC - фильтры)
2. резистивно-емкостные (RC - фильтры)
3. активные (ARC - фильтры)

4. пьезоэлектрические
 5. цифровые
 3. По типу схемы фильтра : - Т; П; Г; мостовые.
 4. По виду математических функций, аппроксимирующих частотные характеристики фильтра (фильтры Чебышева, Баттерворта, Бесселя и др.)
- Реальные АЧХ не имеют резких фронтов спада частотных характеристик, как показано на рис.5.42.

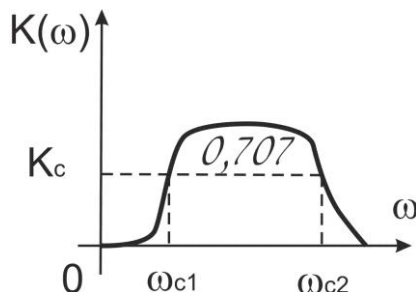


Рис.5.42. Частотная характеристика реального фильтра

Величину K_c обычно выбирают равной $\frac{1}{\sqrt{2}} \approx 0,707$ и частоты соответствующие этому значению принимаются за соответствующие частоты среза анализируемой цепи.

5.12.2. Реактивные фильтры

В настоящее время на практике реактивные фильтры применяются не часто, хотя в течение длительного времени в различных радиоэлектронных устройствах в основном использовался именно этот тип фильтров.

Мы ограничимся рассмотрением простейшей схемы реактивного Т-образного фильтра в режиме холостого хода и при подключении к активной нагрузке.

На рис.5.43. приведены схемы соответствующие обоим рассматриваемым вариантам.

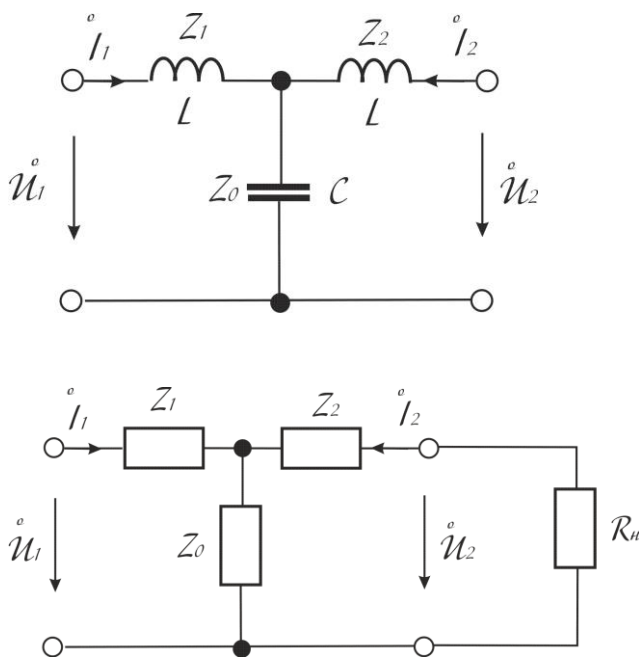


Рис.5.43. Реактивный фильтр в режиме холостого хода и с активной нагрузкой

Для анализа режима холостого хода воспользуемся ранее полученными уравнениями, связывающими между собой передаточные функции и коэффициенты четырехполюсника.

$$W_{U0} = \frac{1}{A(\omega)}.$$

Для симметричного Т-образного четырехполюсника ранее мы получали уравнение, связывающее его коэффициенты с параметрами элементов, формирующих каноническую форму четырехполюсника:

$$A(\omega) = 1 + \frac{Z_1}{Z_0},$$

что в рассматриваемом случае дает

$$A(\omega) = 1 + \frac{j\omega L}{\frac{1}{j\omega C}} = 1 - \omega^2 LC.$$

Соответственно для частотной передаточной функции получаем

$$W_{U0}(j\omega) = \frac{1}{1 - \omega^2 LC} = \frac{1}{1 - \omega^2 \tau^2},$$

где $\tau = \sqrt{LC}$.

В результате мы получили нереальный случай с точкой разрыва 2-го рода в амплитудной частотной характеристике, как показано на рис.5.44.

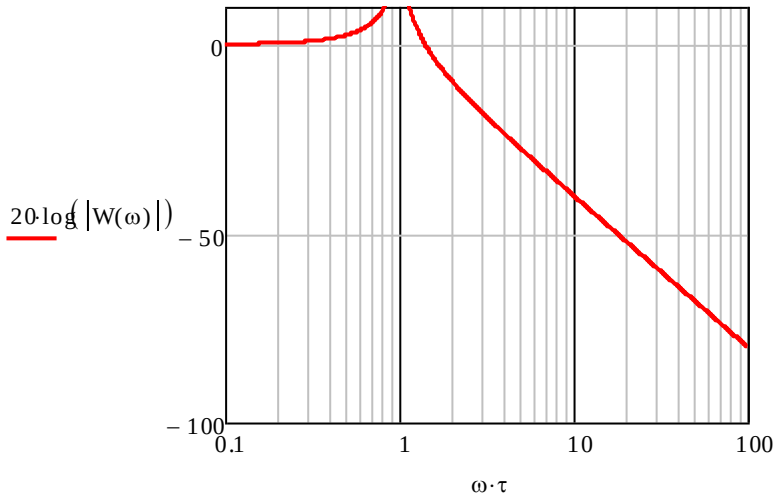


Рис.5.44. Нереальная амплитудная частотная характеристика с разрывом 2-го рода.

Для получения реальных характеристик требуется учесть влияние подключаемой к фильтру нагрузки.

Если $Z_H = R_H \neq 0$, частотная передаточная функция по напряжению для анализируемого четырехполюсника может быть получена из известного нам уравнения

$$W_U(j\omega) = \frac{R_H}{A(j\omega)R_H + B(j\omega)} = \frac{1}{A(j\omega) + B(j\omega)/R_H}.$$

Коэффициенты четырехполюсника, представленного в канонической Т-форме, связаны с комплексными его элементами уравнениями

$$A(j\omega) = 1 + \frac{Z_1}{Z_0} = 1 - \omega^2 LC,$$

$$B(j\omega) = Z_1 + Z_2 + \frac{Z_1 Z_2}{Z_0},$$

что при $Z_1 = Z_2$ дает

$$B(j\omega) = Z_1 \left(2 + \frac{Z_1}{Z_0} \right) = j\omega L(2 - \omega^2 LC).$$

Тогда

$$W_U(j\omega) = \frac{1}{(1 - \omega^2 LC) + j\omega L(2 - \omega^2 LC)/R_H}.$$

Откуда для модуля частотной передаточной функции получаем

$$|W_U(j\omega)| = \frac{1}{\sqrt{(1 - \omega^2 LC)^2 + (\omega L(2 - \omega^2 LC)/R_H)^2}}.$$

График амплитудной частотной характеристики такого фильтра приведен на рис. 5.45. На этом же рисунке приведена в укрупненном масштабе, часть характеристики

вблизи точки резонансной частоты, поскольку именно она представляет практический интерес.

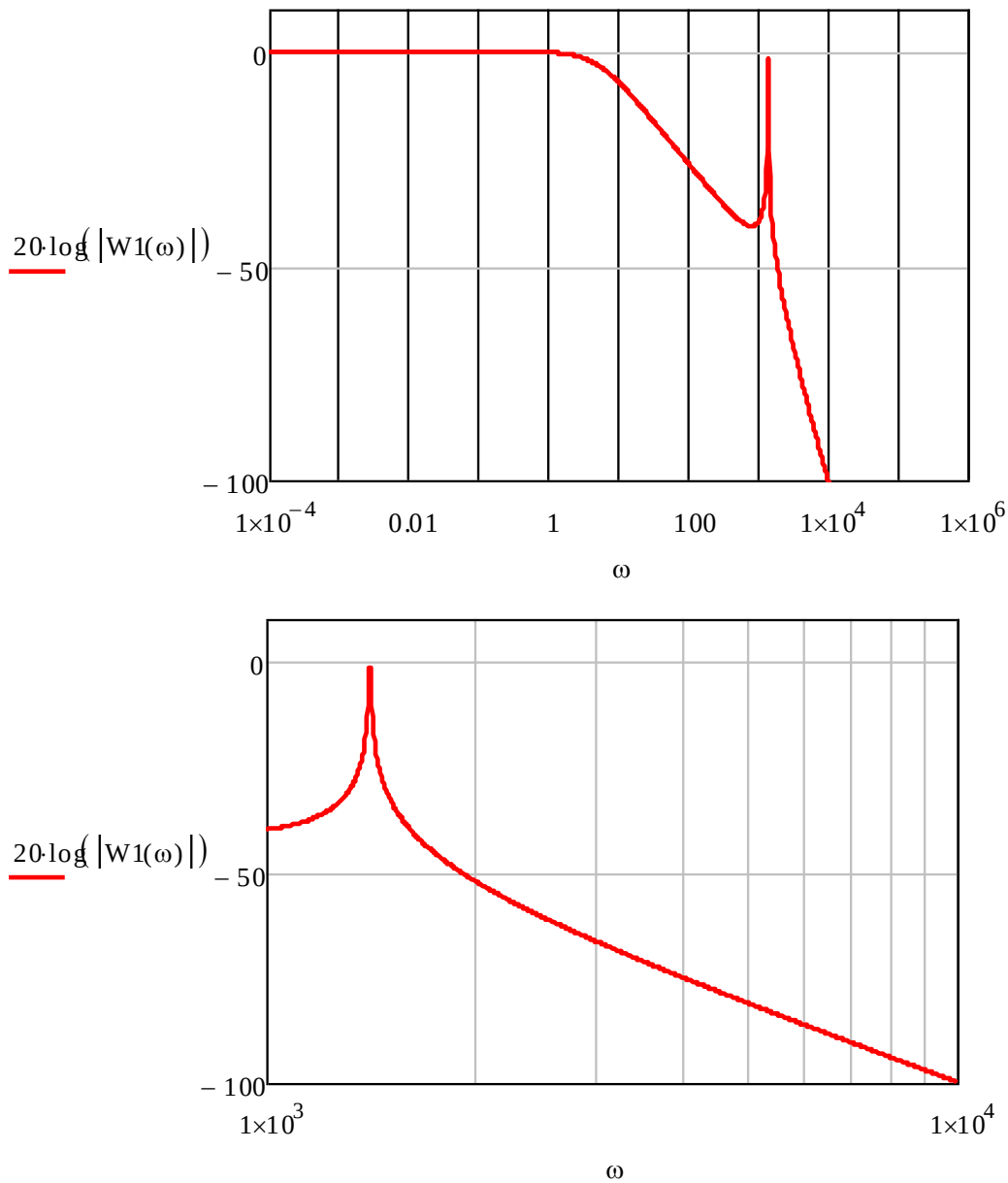


Рис.5.45. Амплитудно-частотная характеристика фильтра и ее укрупненный вариант вблизи резонансной частоты.

Пример 5.18.

Моделирование процесса исследования частотных характеристик реактивного фильтра нижних частот с помощью программы Multisim приведена на рис.5.46.

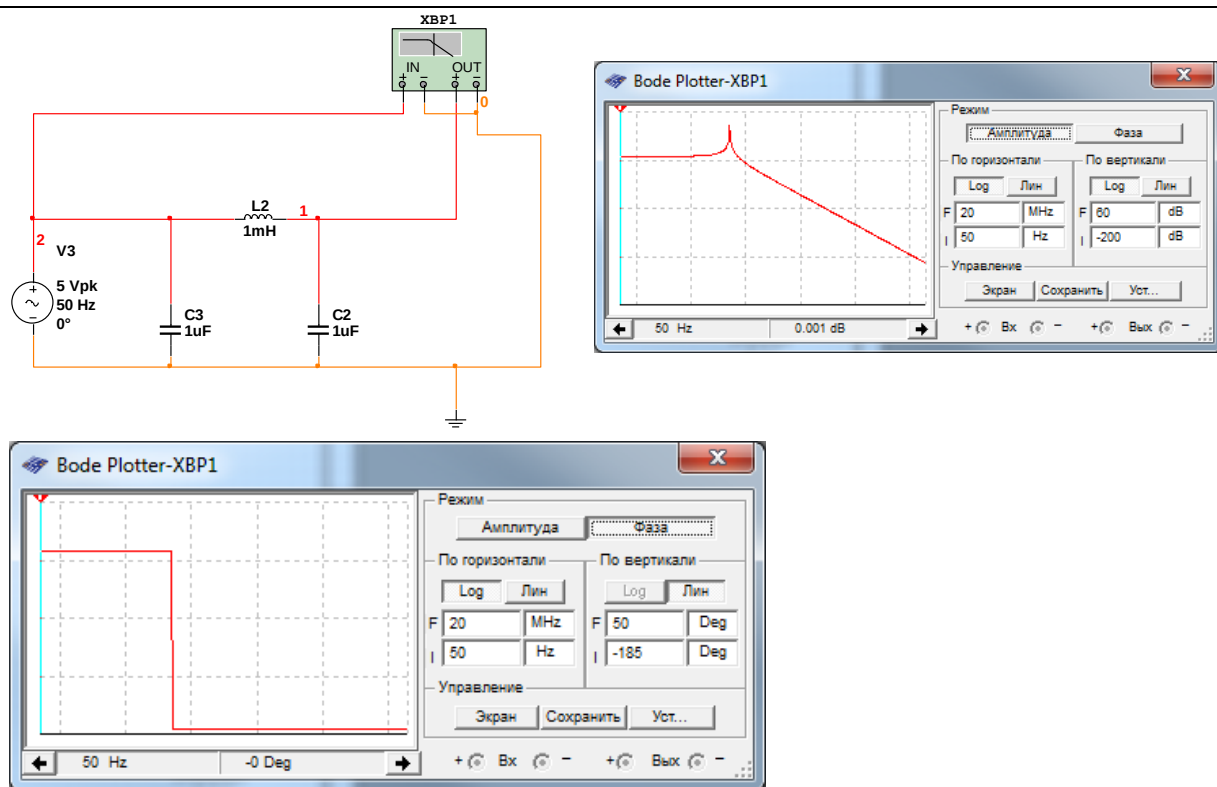
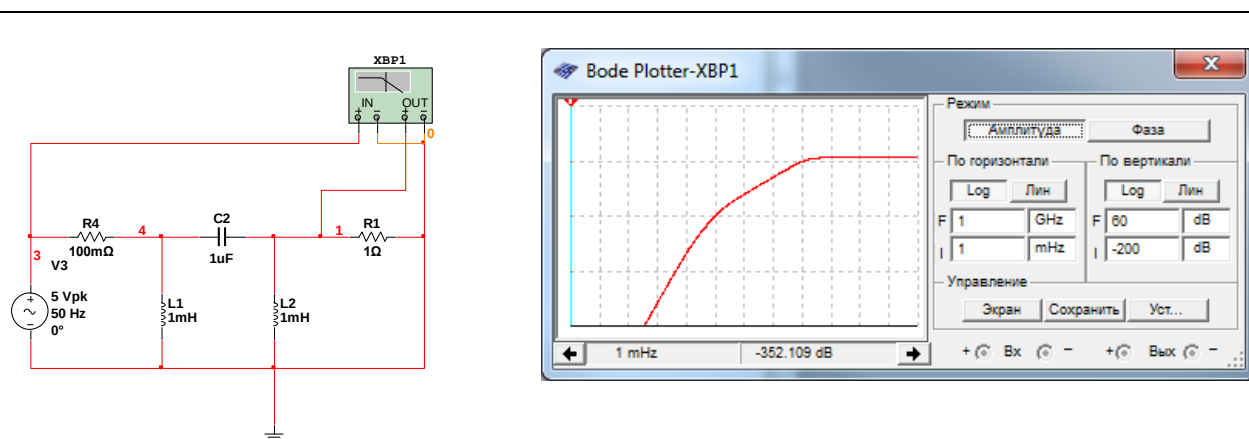
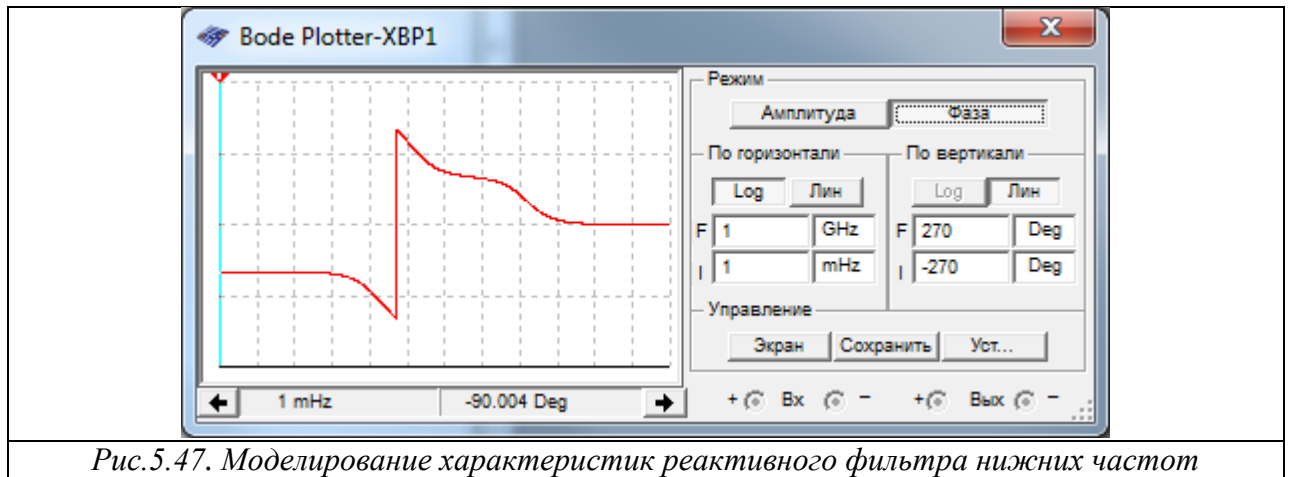


Рис.5.46. Моделирование характеристик реактивного фильтра нижних частот

Пример 5.19.

Моделирование процесса исследования частотных характеристик реактивного фильтра верхних частот с помощью программы Multisim приведена на рис.5.47.

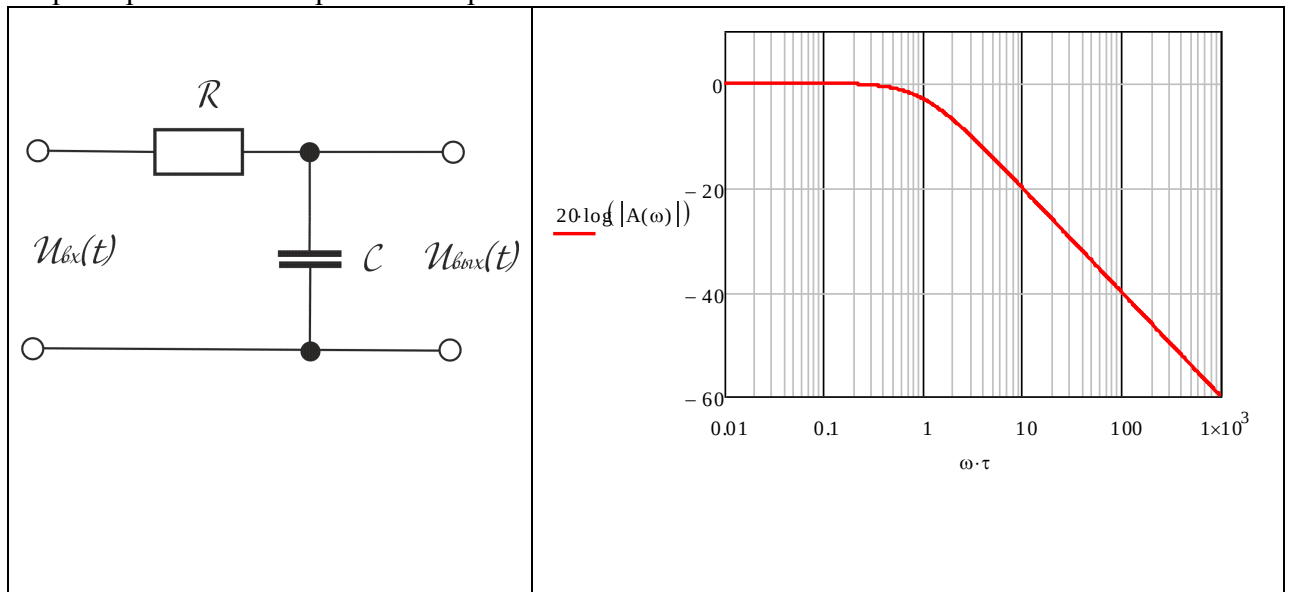


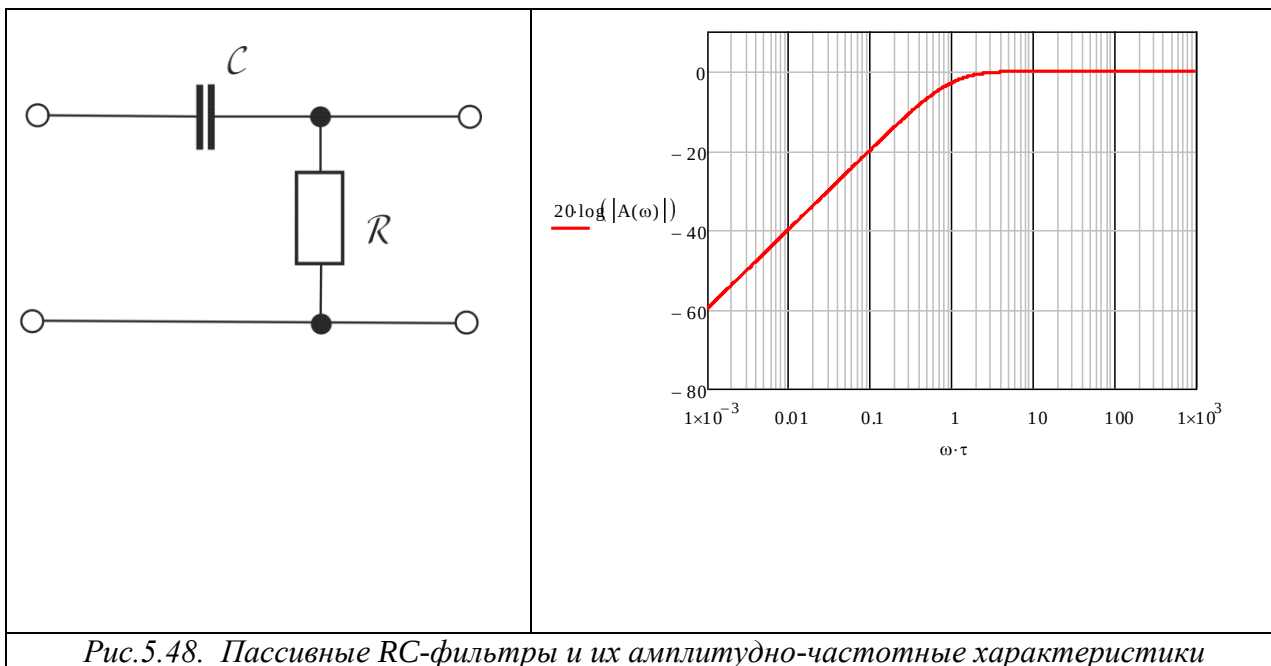


5.12.3. Пассивные RC- фильтры

При рассмотрении простейших RC и RL цепей мы подробно изучили временные и частотные характеристики четырехполюсников, получаемых на их основе. В рамках того рассмотрения мы не использовали понятие полосы пропускания и не анализировали их фильтрующие свойства.

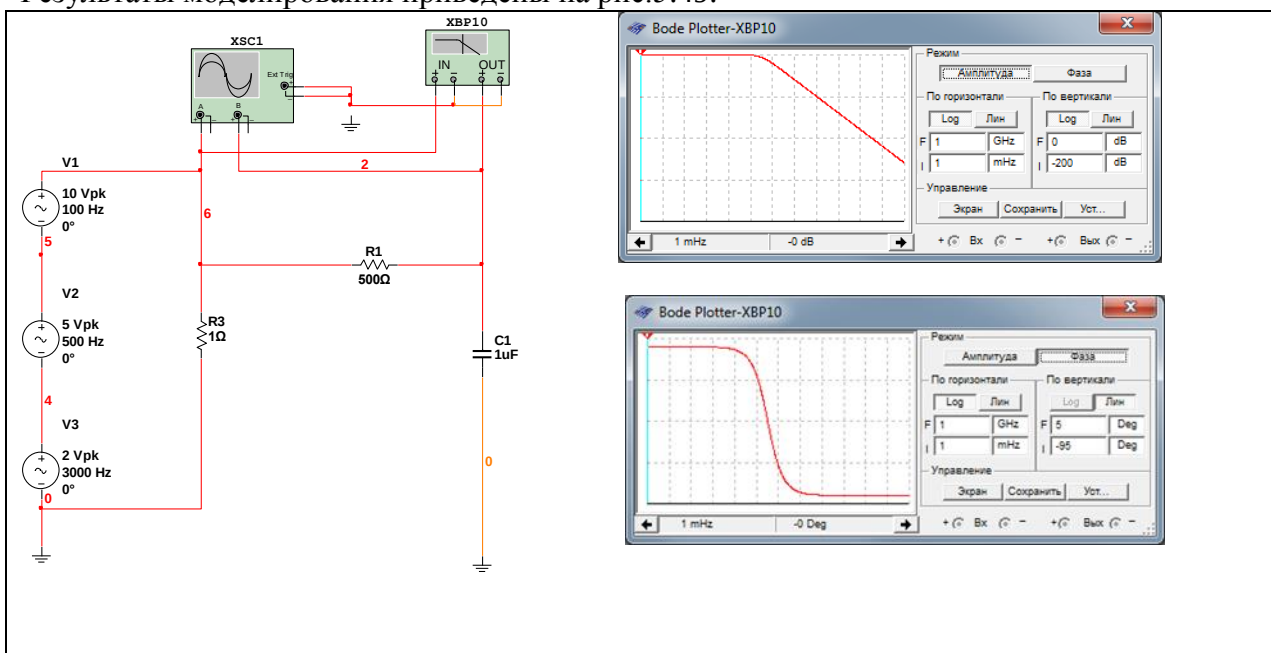
Поскольку все необходимые характеристики этих фильтров уже фактически нами получены, мы не будем повторно останавливаться на данном вопросе, а лишь ограничимся приведением соответствующих схем и амплитудных частотных характеристик, поясняющих фильтрующие возможности этих цепей. Схемы цепей и их характеристики изображены на рис. 5.48.





Пример 5.20.

Проведем моделирование прохождения полигармонического сигнала через простейшую интегрирующую RC-цепь, являющуюся фильтром нижних частот. Данный пример продемонстрирует ограничения эффекта фильтрации, связанные с низкой скоростью спада фронтов спектральной характеристики простейших фильтров. Результаты моделирования приведены на рис.5.49.



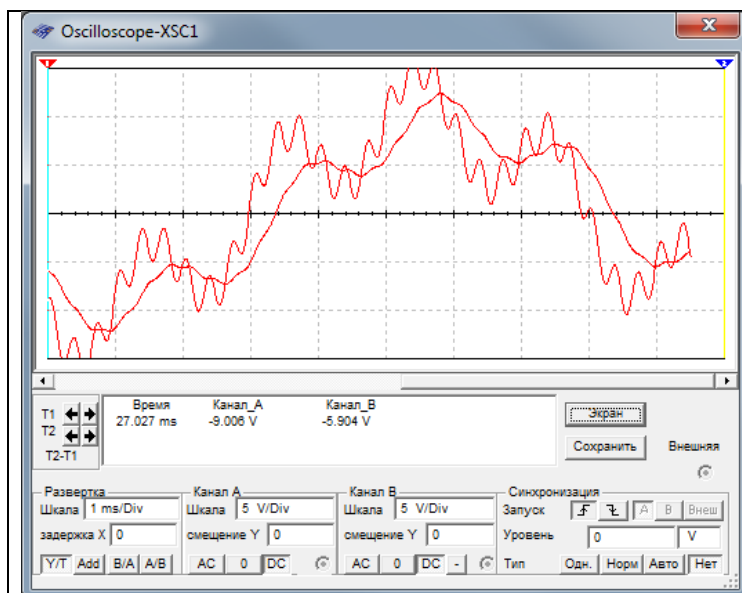


Рис.5.49. Моделирование прохождения полигармонического сигнала через RC-цепь.

5.12.4. Активные фильтры. Использование обратной связи для формирования желаемых частотных свойств фильтра

Активными называются электрические фильтры, в состав которых наряду с пассивными элементами входят усилительные или невзаимные элементы.

Наибольшее распространение получили активные RC-фильтры (ARC-фильтры). Они не содержат индуктивных элементов и могут быть легко реализованы в виде интегральной схемы.

В задачи нашего курса не входит подробное изучение схемных решений и характеристик электронных устройств на основе интегральных усилителей. Мы ограничимся кратким рассмотрением общих методических вопросов на примере использования идеализированного дифференциального усилителя с коэффициентом усиления стремящемся к бесконечности.

Упрощенная схема четырехполюсника на основе такого усилителя приведена на рис.5.50.

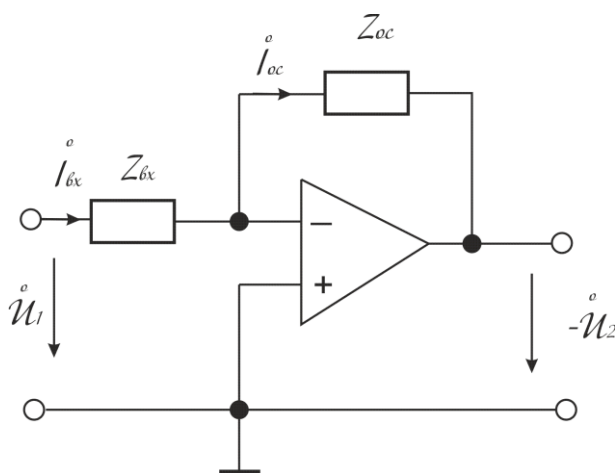


Рис.5.50. Активный четырехполюсник на основе идеального операционного усилителя

Бесконечное значение коэффициента усиления идеального операционного усилителя позволяет очень простым способом получить уравнения, связывающие его входное и выходное напряжение при использовании схем с обратной связью.

Поскольку коэффициент усиления (в разомкнутом состоянии) у идеального операционного усилителя стремится к бесконечности и его входное сопротивление также велико, то можно принять следующее допущение

$$I_{\text{вх}}^{\circ} = I_{\text{ос}}^{\circ}.$$

Соответственно,

$$\frac{U^{\circ}_1}{Z_{\text{BX}}} = -\frac{U^{\circ}_2}{Z_{\text{OC}}}$$

или

$$U^{\circ}_2 = -\frac{Z_{\text{OC}}}{Z_{\text{BX}}} U^{\circ}_1,$$

то есть частотная передаточная функция данного четырехполюсника будет определяться только отношением комплексных сопротивлений входной ветви и обратной связи

$$W(j\omega) = -\frac{Z_{\text{OC}}(j\omega)}{Z_{\text{BX}}(j\omega)}.$$

В частности, если выбрать $Z_{\text{BX}} = R$ и $Z_{\text{OC}} = 1/j\omega C$, как показано на рис.5.51.

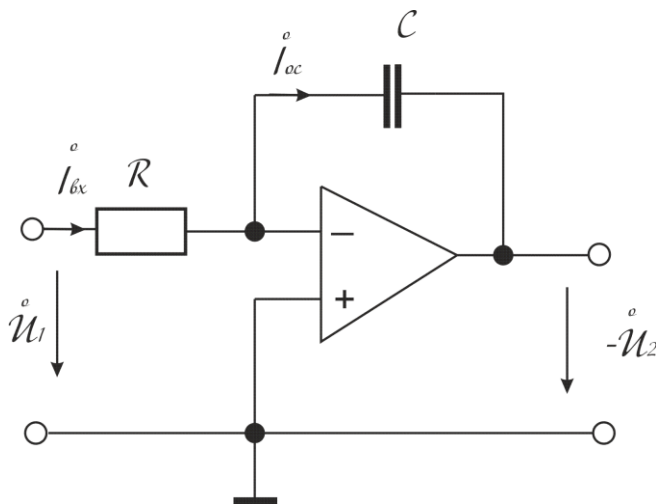


Рис.5.51. Активный RC- фильтр

Частотная передаточная функция такого фильтра будет равна

$$W(j\omega) = -\frac{1}{j\omega C R} = j \frac{1}{\omega C R},$$

что соответствует идеальному интегрирующему звену.

Если в цепь обратной связи параллельно конденсатору добавить резистор, как показано на рис.5.50, то характеристики четырехполюсника существенно изменятся:

$$W(j\omega) = -\frac{\frac{1}{j\omega C} R}{R \left(R + \frac{1}{j\omega C} \right)} = \frac{1}{1 + j\omega RC}.$$

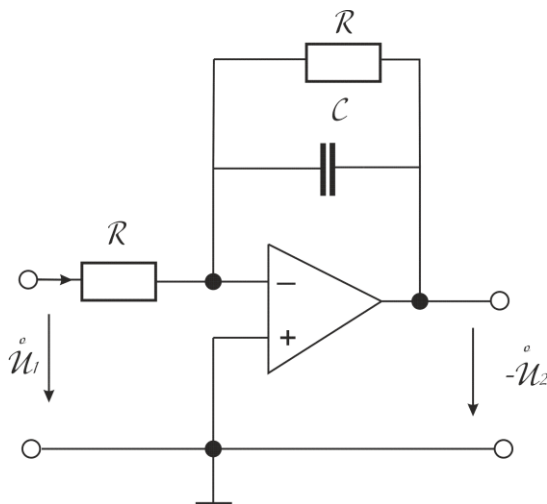


Рис.5.52 Активный RC- фильтр.

Пример 5.21.

Произведем моделирование характеристик активного RC- фильтра с помощью программы Multisim. Теоритические расчеты для моделируемой цепи рекомендуем провести самостоятельно в качестве дополнительного упражнения и сравнить результаты, полученные обоими методами. Процесс моделирования представлен на рис. 5.53.

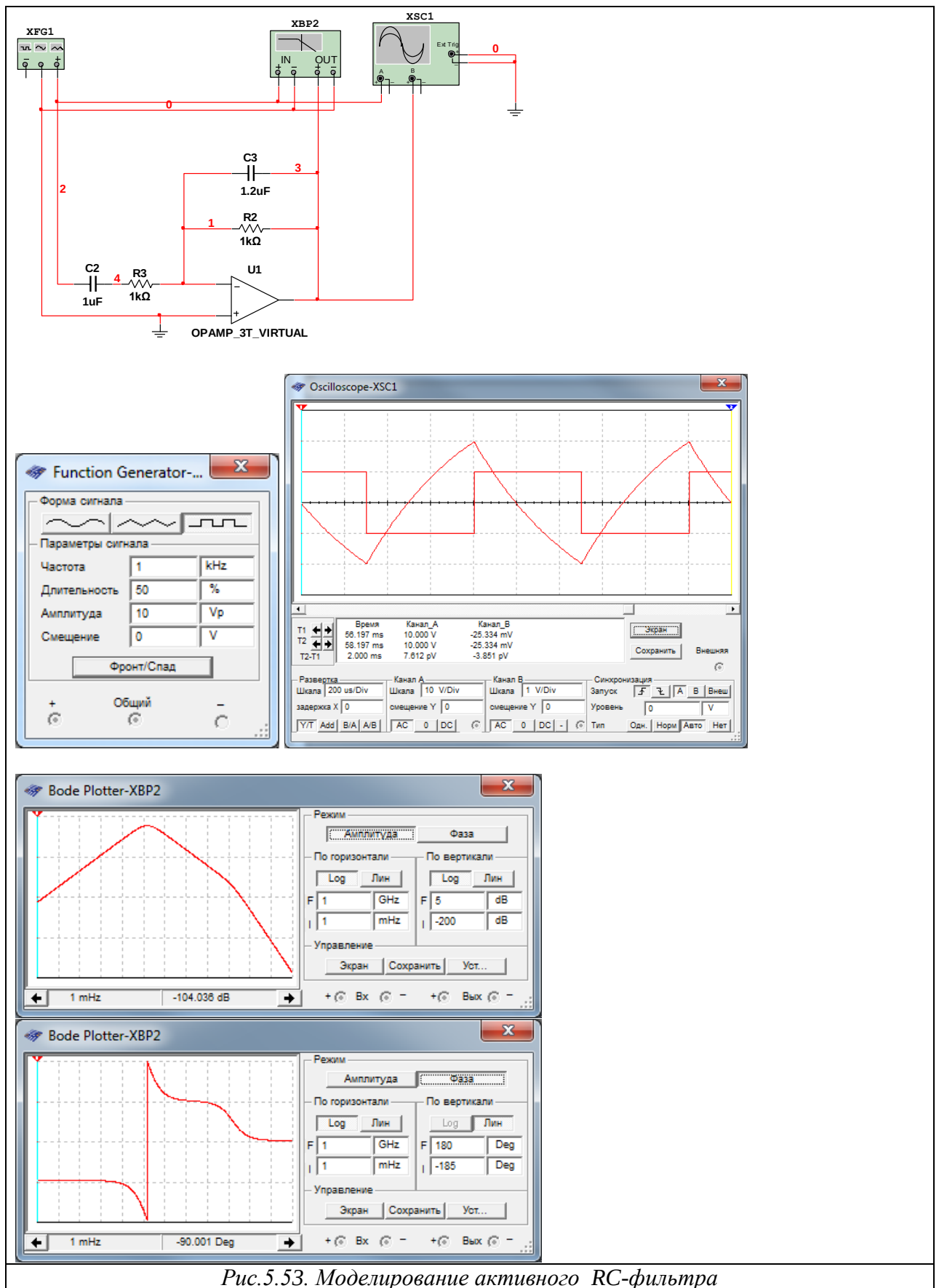


Рис.5.53. Моделирование активного RC-фильтра

5.13. Прохождение случайного сигнала через линейную цепь

Рассмотрим линейную цепь как линейную стационарную систему с постоянными коэффициентами, имеющую передаточную функцию $W(s)$ и импульсную функцию $g(t)$.

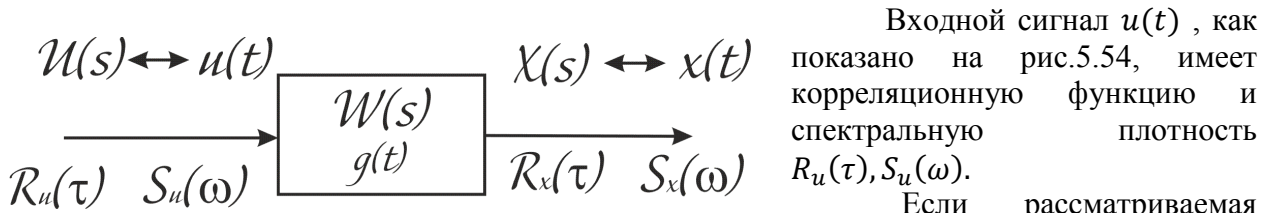


Рис.5.54. Прохождение случайного сигнала через линейную систему

Если рассматриваемая линейная система устойчива и стационарна, то на выходе системы установившийся сигнал $x(t)$ также будет стационарным, и, следовательно, может быть описан некой корреляционной функцией $R_x(\tau)$ и спектральной плотностью $S_x(\omega)$.

Установим связь между корреляционными функциями и спектральными плотностями на входе и выходе системы.

Для отдельной реализации процесса в соответствии с уравнением свертки для линейной системы можно записать

$$x(t) = \int_{-\infty}^{\infty} g(t - \gamma)u(\gamma)d\gamma = \int_{-\infty}^{\infty} g(\gamma)u(t - \gamma)d\gamma.$$

Если корреляционная функция входного стационарного случайного процесса равна $R_u(t_1, t_2)$, то для корреляционной функции выходного сигнала $x(t)$ можно записать

$$R_x(t_1, t_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(t_1 - \gamma_1)g(t_2 - \gamma_2)R_u(t_1 - t_2)d\gamma_1d\gamma_2,$$

то есть выходной сигнал также будет стационарным.

Вводя замену переменных $t_1 - \gamma_1 = s_1$, $t_2 - \gamma_2 = s_2$, получаем

$$R_x(\tau) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(s_1)g(s_2)R_u(\tau + s_1 - s_2)ds_1ds_2.$$

Воспользуемся уравнением Винера-Хинчина

$$S_{Px}(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} R_x(\tau)e^{-j\omega\tau}d\tau$$

и запишем

$$S_{Px}(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(s_1)g(s_2)R_u(\tau + s_1 - s_2)ds_1ds_2 \right) e^{-j\omega\tau}d\tau.$$

Переходя к новой переменной $z = \tau + s_1 - s_2$, и принимая во внимание, что частотная передаточная функция линейной системы связана с импульсной характеристикой соотношением

$$W(j\omega) = \int_{-\infty}^{\infty} g(s)e^{-j\omega s}ds,$$

получаем

$$S_{Px}(\omega) = \int_{-\infty}^{\infty} g(s_1) e^{-j\omega s_1} ds_1 \int_{-\infty}^{\infty} g(s_2) e^{j\omega s_2} ds_2 \frac{1}{2\pi} \int_{-\infty}^{\infty} R_u(z) e^{-j\omega z} dz.$$

Откуда следует

$$S_{Px}(\omega) = W(j\omega)W(-j\omega)S_{Pu}(\omega) = |W(j\omega)|^2 S_{Pu}(\omega). \quad (5.61)$$

Для корреляционной функции выходного сигнала теперь запишем

$$R_x(\tau) = \int_{-\infty}^{\infty} S_{Pu}(\omega) |W(j\omega)|^2 e^{j\omega\tau} d\omega. \quad (5.62)$$

Дисперсия выходного сигнала в этом случае может быть определена следующим образом

$$D_x = \int_{-\infty}^{\infty} S_{Pu}(\omega) |W(j\omega)|^2 d\omega. \quad (5.63)$$

Если для импульсной характеристики линейной системы $g(t)$ ввести корреляционную функцию в виде

$$B_g(\tau) = \int_{-\infty}^{\infty} g(t)g(t-\tau)dt,$$

то для корреляционной функции выходного сигнала можно записать

$$R_x(\tau) = \int_{-\infty}^{\infty} R_u(t) B_g(\tau-t)dt.$$

Отдельный интерес представляет собой случай, когда на вход линейной системы поступает белый шум с постоянной спектральной плотностью мощности

$$S_P(\omega) = S_N = const.$$

Тогда спектральная плотность мощности случайного сигнала на выходе данной линейной системы будет равна

$$S_x(\omega) = S_N |W(j\omega)|^2 \quad (5.64)$$

Приведенные уравнения (5.61) - (5.64) позволяют произвести необходимые вычисления для определения вероятностных характеристик выходного сигнала по заданным характеристикам входного сигнала и параметрам радиотехнической цепи.

5.14. Вопросы для самоконтроля и повторения

1. Дайте определение четырехполюсника.
2. Какие четырехполюсники называются пассивными?
3. Какие четырехполюсники называются активными?
4. Какие четырехполюсники называются автономными?
5. Какие четырехполюсники называются обратимыми?
6. Какие четырехполюсники называются симметричными?
7. Запишите уравнение четырехполюсника в Z, Y и A формах.

8. Каким образом можно произвести экспериментальное определение параметров четырехполюсника?
9. Какие цепи (системы) называются линейными?
10. Запишите уравнение прямого преобразования Лапласа.
11. Запишите уравнение обратного преобразования Лапласа.
12. Перечислите основные свойства преобразования Лапласа.
13. Дайте определение передаточной функции цепи.
14. Что такое частотная передаточная функция (комплексный коэффициент передачи)?
15. Каким образом находятся амплитудная и фазовая частотные функции?
16. В чем состоит физический смысл амплитудной и фазовой частотных функций.
17. Какие временные характеристики линейных цепей вам известны?
18. Каким образом связаны между собой передаточная и весовая функции цепи?
19. Запишите уравнение интеграла наложения (интеграла Дюамеля).
20. Какие формы представления передаточной функции вам известны?
21. Что такое фундаментальная матрица и какие способы ее вычисления вам известны?
22. В чем состоит разница между однонаправленными цепями и цепями с обратной связью?
23. Какие типы соединений четырехполюсников вам известны?
24. Какие цепи называют устойчивыми?
25. Какие критерии устойчивости вам известны?
26. Каковы основные характеристики интегрирующей RC цепи?
27. Каковы основные характеристики дифференцирующей RC цепи?
28. Дайте определение резонанса.
29. Каково условие возникновения резонанса в последовательном колебательном контуре?
30. Что такое добротность колебательного контура?
31. Как изменяются параметры амплитудно модулированного сигнала при прохождении через колебательный контур?
32. Что такое электрический фильтр?
33. Приведите классификацию электрических фильтров.
34. Что такое полоса пропускания электрического фильтра?
35. Приведите схему простейшего RC фильтра нижних частот и рассчитайте его характеристики.
36. Приведите схему простейшего RC фильтра верхних частот и рассчитайте его характеристики.
37. Приведите схему простейшего режекторного фильтра и рассчитайте его характеристики.
38. Приведите схему простейшего полосового фильтра и рассчитайте его характеристики.
39. Поясните разницу между активными и пассивными фильтрами.
40. Какие явления наблюдаются при прохождении через линейную цепь случайного сигнала?

ГЛАВА 6 . Преобразование сигналов в нелинейных и параметрических цепях

6.1. Линейные и нелинейные элементы цепей

Рассмотренные ранее методы анализа сигналов и цепей в качестве одной из основных операций использовали замену исходного сигнала суммой других сигналов, математическая обработка которых более проста или представление которых более полно раскрывает какой-либо информационный аспект исходного сигнала.

В линейных системах и цепях реакция на сумму воздействий равна сумме реакций на каждое из воздействий в отдельности, что составляет основу принципа линейной суперпозиции.

Если исходный сигнал имеет вид

$$x(t) = \alpha x_1(t) + \beta x_2(t), \quad (6.1)$$

то преобразование реакции линейной системы (выходной сигнал) может быть представлен суммой

$$S[x(t)] = S[\alpha x_1(t) + \beta x_2(t)] = \alpha S[x_1(t)] + \beta S[x_2(t)], \quad (6.2)$$

где α, β – коэффициенты,

S – обозначение результата прохождения сигнала через линейную цепь.

Для нелинейных элементов принцип линейной суперпозиции (6.1), (6.2) не выполняется. При этом, если хотя бы один из элементов цепи является нелинейным, то и вся цепь будет нелинейной.

Нарушение принципа суперпозиции существенно усложняет процесс анализа нелинейных цепей, так как в данном случае становится неприменим ни принцип наложения (интеграл Дюамеля), ни операторное представление в форме передаточной функции или комплексного коэффициента передачи в том виде, как они использовались для анализа линейных сигналов и цепей.

Нелинейность может проявляться в двух видах: нелинейность статических характеристик элементов или нелинейность их динамических характеристик.

Примером нелинейной статической характеристики может служить вольт-амперная характеристика (ВАХ) большинства полупроводниковых элементов, пример одной из которых изображен на рис.6.1.

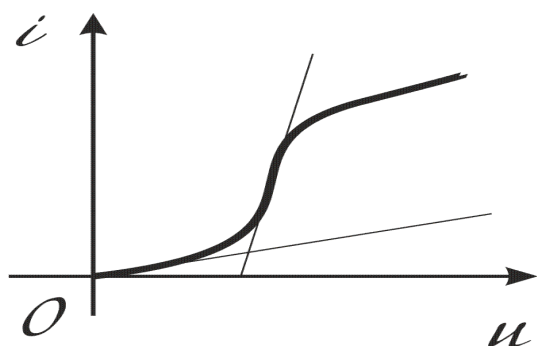


Рис.6.1. Нелинейная статическая характеристика

В общем виде ВАХ может быть описана уравнением

$$i = f(u), \quad (6.3)$$

где i – ток, протекающий по элементу,

u – напряжение, приложенное к элементу,

f – некоторая функция, связывающая между собой

приложенное напряжение и возникающий при этом ток.

Приведенная ВАХ (6.3) не содержит в явном виде переменную времени, поэтому ее и называют статической или безынерционной, подчеркивая тем самым, что изменения тока происходят мгновенно с изменением приложенного напряжения без каких-либо динамических задержек или искажений.

В радиотехнике различают резистивные (сопротивления) и реактивные (емкости и индуктивности) нелинейные элементы. Последние вносят в процесс определенные динамические искажения, величина и характер которых зависит от характеристик соответствующего нелинейного элемента.

В реальности строго резистивных нелинейных элементов не существует, так как всегда имеются (пусть и незначительные) паразитные емкости и индуктивности элементов. Поэтому, строго говоря, безынерционных элементов практически нет. Однако при проектировании радиотехнической аппаратуры она, как правило, рассчитывается на эксплуатацию в определенных диапазонах изменения параметров сигналов, что сопровождается соответствующим выбором расположения рабочих точек. В результате в большинстве случаев предположение о возможности описания характеристик нелинейных элементов только статическими (безынерционными) зависимостями оказывается допустимым.

6.2. Аппроксимация характеристик безынерционных нелинейных элементов

В большинстве практических ситуаций характеристики нелинейных элементов получают экспериментальным путем, в результате чего для анализа являются доступными только табличные данные соответствия значений входной и выходной величины. Для ВАХ такими значениями будут соответственно напряжение, приложенное к элементу и возникший при этом ток.

Для анализа цепей требуется от табличного вида ВАХ необходимо перейти к ее аналитическому (в виде определенной формулы) описанию. При решении задачи аналитической аппроксимации экспериментального набора данных необходимо, во-первых, выбрать вид аппроксимирующей функции, а во-вторых, найти значения параметров (коэффициентов) выбранной функции, обеспечивающих хорошее согласование с экспериментальным набором на всем диапазоне рабочего отрезка ВАХ. Выбранная для целей аппроксимации функция с одной стороны должна максимально точно соответствовать реальности, а с другой – быть удобной для дальнейшего аналитического применения.

Наиболее часто для аппроксимации нелинейной характеристики радиотехнических элементов используют представление в виде полиномиального ряда, то есть

$$i = a_0 + \sum_{k=1}^{\infty} a_k u^k, \quad (6.4)$$

в котором ограничиваются конечным числом слагаемых

$$i = a_0 + a_1 u + a_2 u^2 + a_3 u^3 + \dots + a_n u^n, \quad (6.5)$$

где $a_k, k = 0, 1, \dots, n$ – набор искомых коэффициентов разложения.

Вопросы разложения произвольной функции по заданной базисной системе функций с использованием регрессии общего вида мы уже подробно рассматривали в данном курсе, поэтому повторно останавливаться на нем не будем.

Отметим только, что в случае, если рабочей точкой элемента, является точка, отличная от начала координат, то соответствующий сдвиг надо внести и в аппроксимирующую формулу (6.4) или (6.5)

$$i = a_0 + a_1(u - u_0) + a_2(u - u_0)^2 + a_3(u - u_0)^3 + \dots + a_n(u - u_0)^n, \quad (6.6)$$

где u_0 – абсцисса рабочей точки.

В случае, когда нелинейная характеристика задана сложной, но аналитической функцией, искомую аппроксимацию можно получить разложением в ряд Тейлора (Маклорена) или применив аналитическое разложение по заданной системе базисных функций, что нами также уже было подробно рассмотрено.

Для нелинейных статических характеристик используется такой параметр, как дифференциальная крутизна, являющаяся первой производной от самой характеристики и представляющая собой тангенс угла наклона касательной, проведенной к каждой точке характеристики.

$$S_d = \frac{d}{du} f(u).$$

Как видно из рис.6.1. в различных точках кривой наклон касательной может быть существенно разным. Поэтому для определенности вводят показатель, называемый крутизной ВАХ в рабочей точке

$$S_d(u_0) = \left. \frac{d}{du} f(u) \right|_{u=u_0}. \quad (6.7)$$

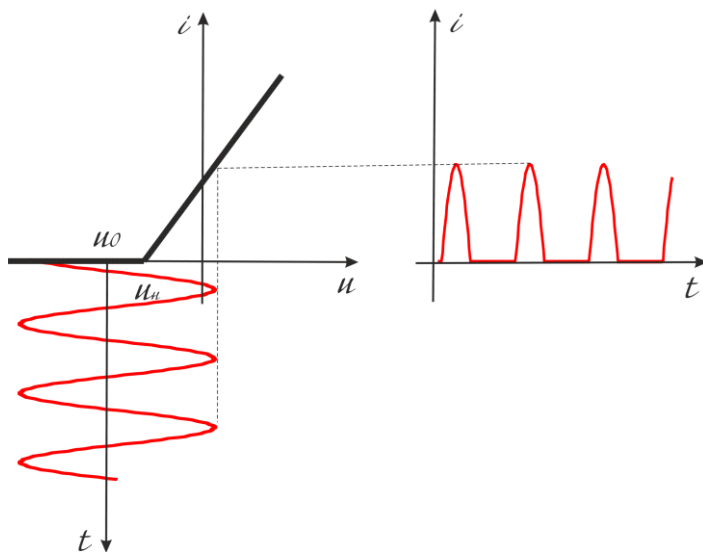


Рис.6.2. Кусочно-линейная аппроксимация ВАХ нелинейного элемента

приложенное синусоидальное напряжение.

Помимо полиномиальной и кусочно-линейной аппроксимации иногда для описания ВАХ полупроводниковых элементов используется экспоненциальная функция с двумя задаваемыми параметрами

$$i = a_1(e^{a_2 u} - 1), \quad (6.8)$$

где a_1, a_2 – подбираемые параметры аппроксимирующей функции.

График экспоненциальной аппроксимирующей функции приведен на рис.6.3.

В ряде случаев описать нелинейную характеристику единой аналитической функцией, как например, (6.65) или (6.6), крайне сложно, но зато можно получить эффективную аппроксимацию используя кусочный набор аппроксимирующих функций.

В простейшем случае это подход использует линейные функции, поэтому он называется кусочно-линейной аппроксимацией. На рис.6.2. приведен пример кусочно-линейной аппроксимации нелинейной характеристики двумя отрезками прямой и представлена реконструкция возникающего на элементе отклика тока на

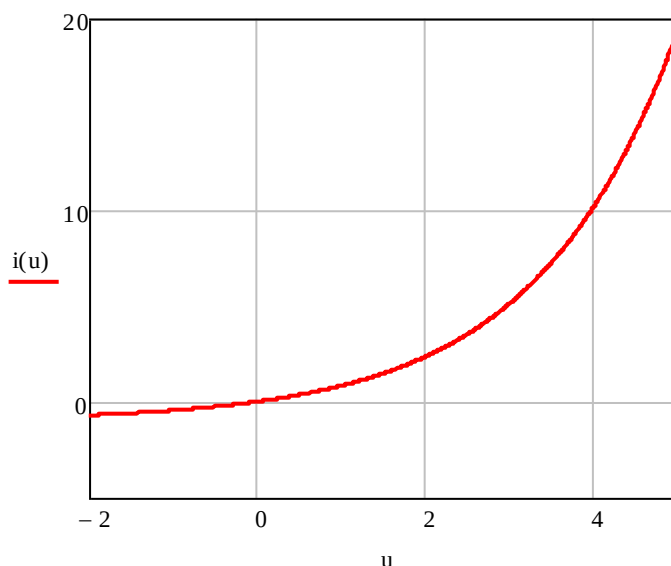


Рис.6.3. Аппроксимация нелинейной характеристики экспоненциальной функцией

Пример 6. 1.

Определение Воль-амперной характеристики нелинейного двухполюсника. В качестве нелинейного двухполюсника с кусочно-линейной ВАХ выберем диод. Схема и результат моделирования представлен на рис.6.4.

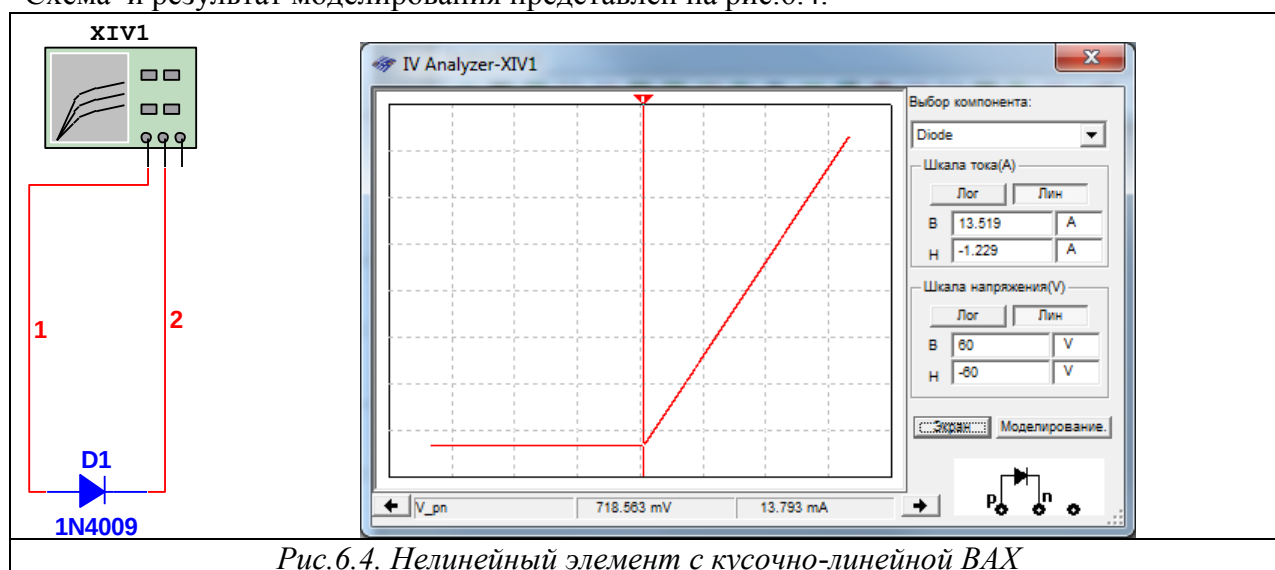


Рис.6.4. Нелинейный элемент с кусочно-линейной ВАХ

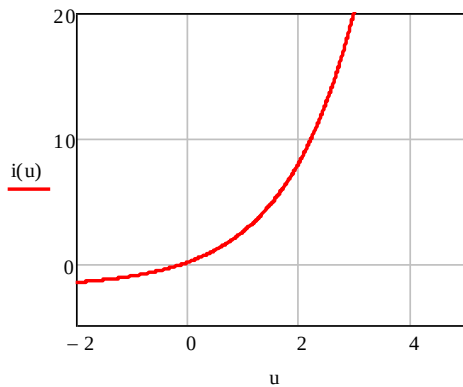
6.3. Нелинейные безынерционные элементы при гармонических воздействиях

Рассмотрим процесс воздействия гармонических колебаний на нелинейный элемент на примере простого сигнала вида

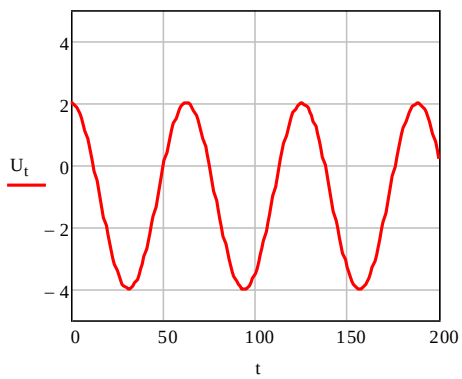
$$u(t) = u_0 + U_m \cos(\omega_1 t + \theta_1). \quad (6.9)$$

Предположим, что данное напряжение приложено к двухполюснику, имеющему нелинейную ВАХ, допускающую аппроксимацию степенным рядом, как было показано в предыдущем параграфе.

Нам требуется найти описание возникающего в двухполюснике тока при условии его подключения к заданному напряжению $u(t)$. Схематично процесс воздействия гармонического напряжения на нелинейный элемент представлен на рис.6.5
ВАХ нелинейного двухполюсника



Входное напряжение



Выходной ток

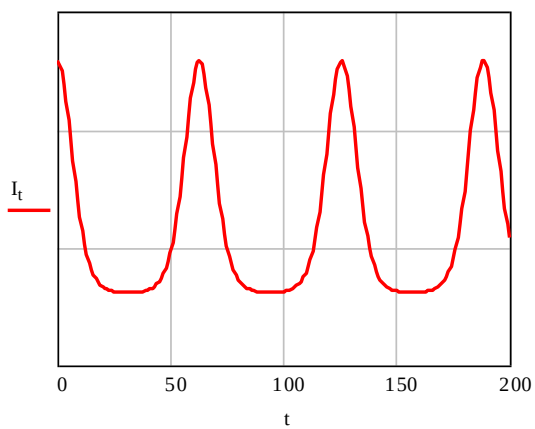


Рис.6.5. Воздействие гармонического напряжения на двухполюсник с нелинейной ВАХ

Пример 6.2.

На рис.6.6. представлены результаты моделирования в программе Simulink (Matlab) процесса прохождения гармонического сигнала через безынерционные цепи с различными нелинейными ВАХ.

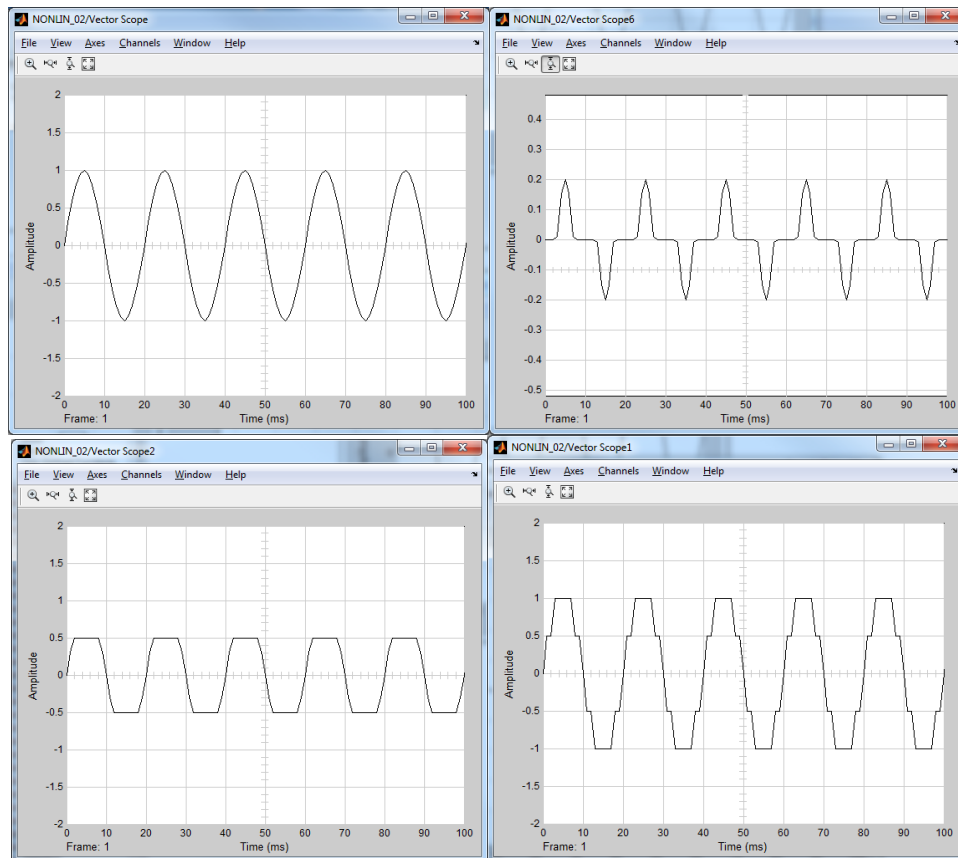
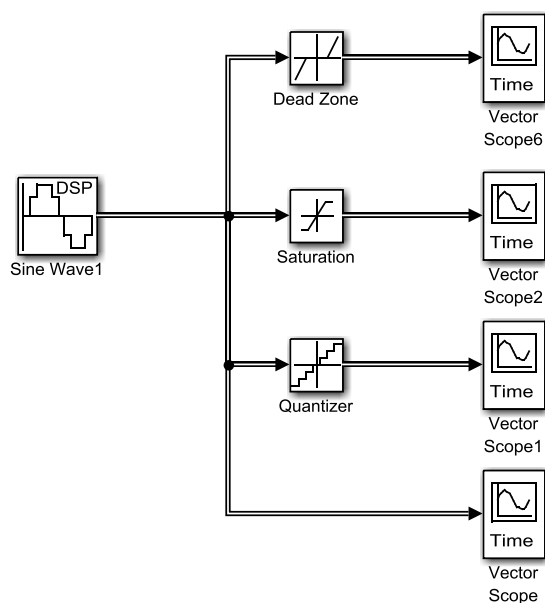


Рис.6.6. Прохождение гармонического сигнала через цепи с нелинейными ВАХ

6.3.1. Полиномиальная аппроксимация ВАХ

При использовании для аппроксимации ВАХ полиномиального ряда (6.5), (6.6)

$$\begin{aligned} i &= a_0 + a_1 u + a_2 u^2 + a_3 u^3 + \dots + a_n u^n + \dots \\ i &= a_0 + a_1(u - u_0) + a_2(u - u_0)^2 + a_3(u - u_0)^3 + \dots + a_n(u - u_0)^n + \dots \end{aligned}$$

постоянная составляющая в уравнении входного гармонического напряжения u_0 определит выбор рабочей точки, в результате уравнение для тока в цепи, исходя из (6.9) может быть записано следующим образом

$$i = a_0 + a_1 U_m \cos(\omega_1 t + \theta_1) + a_2 U_m^2 \cos^2(\omega_1 t + \theta_1) + \dots + a_n U_m^n \cos^n(\omega_1 t + \theta_1) + \dots$$

Для дальнейших преобразований можно воспользоваться тригонометрическими формулами вида

$$\begin{aligned} \cos^2 \beta &= \frac{1}{2}(1 + \cos 2\beta), \\ \cos^3 \beta &= \frac{1}{4}(3\cos \beta + \cos 3\beta), \\ \cos^4 \beta &= \frac{1}{8}(3 + 4\cos 2\beta + \cos 4\beta), \\ \cos^5 \beta &= \frac{1}{16}(10\cos \beta + 5\cos 3\beta + \cos 5\beta) \end{aligned}$$

и так далее.

Теперь путем обычной подстановки для тока в цепи легко записать

$$\begin{aligned} i(t) &= a_0 + \frac{1}{2}a_2 U_m^2 + \frac{3}{8}a_4 U_m^4 + \dots \\ &+ \left(a_1 U_m + \frac{3}{4}a_3 U_m^3 + \frac{5}{8}a_5 U_m^5 + \dots \right) \cos(\omega_1 t + \theta_1) + \\ &+ \left(\frac{1}{2}a_2 U_m^2 + \frac{1}{8}a_4 U_m^4 + \dots \right) \cos[2(\omega_1 t + \theta_1)] + \\ &+ \left(\frac{1}{4}a_3 U_m^3 + \frac{5}{16}a_5 U_m^5 + \dots \right) \cos[3(\omega_1 t + \theta_1)] + \dots = \\ &= I_0 + I_1 \cos(\omega_1 t + \theta_1) + I_2 \cos[2(\omega_1 t + \theta_1)] + I_3 \cos[3(\omega_1 t + \theta_1)] + \dots, \end{aligned} \quad (6.10)$$

где $I_0 = a_0 + \frac{1}{2}a_2 U_m^2 + \frac{3}{8}a_4 U_m^4 + \dots$

$$I_1 = a_1 U_m + \frac{3}{4}a_3 U_m^3 + \frac{5}{8}a_5 U_m^5 + \dots$$

$$I_2 = \frac{1}{2}a_2 U_m^2 + \frac{1}{8}a_4 U_m^4 + \dots$$

$$I_3 = \frac{1}{4}a_3 U_m^3 + \frac{5}{16}a_5 U_m^5 + \dots$$

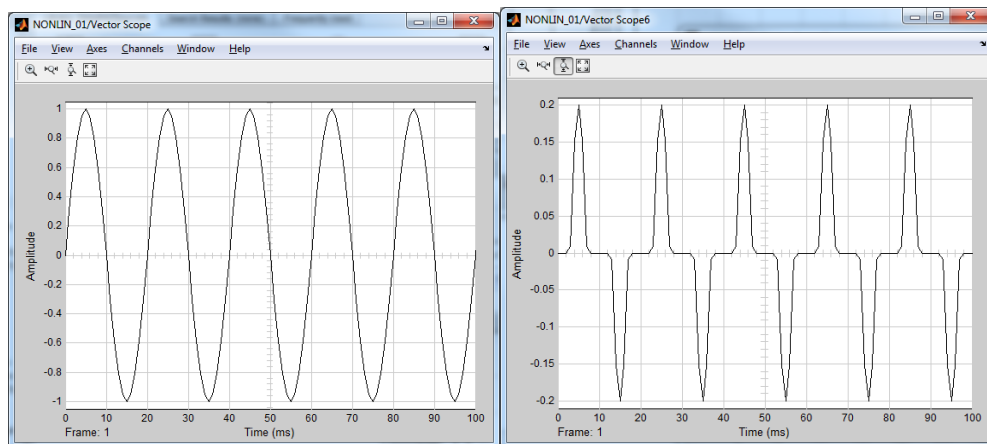
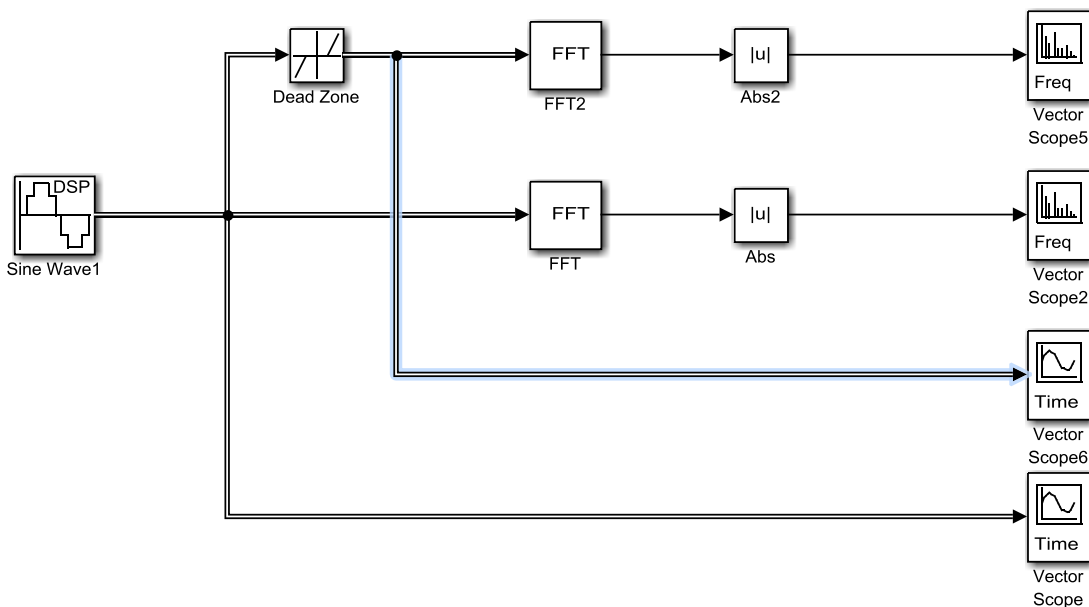
Анализ полученных уравнений (6.10) показывает наличие следующих закономерностей:

- в отличие от моногармонического характера приложенного напряжения ток в цепи имеет в своем составе высшие гармоники кратные частоте приложенного напряжения, причем гармоники с четными значениями кратности обусловлены четными степенями аппроксимирующего полинома, а гармоники с нечетными коэффициентами кратности – нечетными степенями полиномиальной аппроксимации;

- максимальное значение кратности высшей гармоники равно степени полинома, использованного для аппроксимации нелинейной характеристики цепи,
- амплитуда первой гармоники возникающего в цепи тока, то есть основной частоты приложенного напряжения, связана с амплитудой возбуждающего напряжения нелинейной зависимостью, обусловленной нечетными степенями аппроксимирующего полинома.

Пример 6.3.

На рис.6.7. представлен пример моделирования прохождения гармонического сигнала через цепь с нелинейной ВАХ с помощью программы Simulink (Matlab). На графиках полученных спектров входного и выходного сигналов отчетливо наблюдается эффект появления новых частотных составляющих.



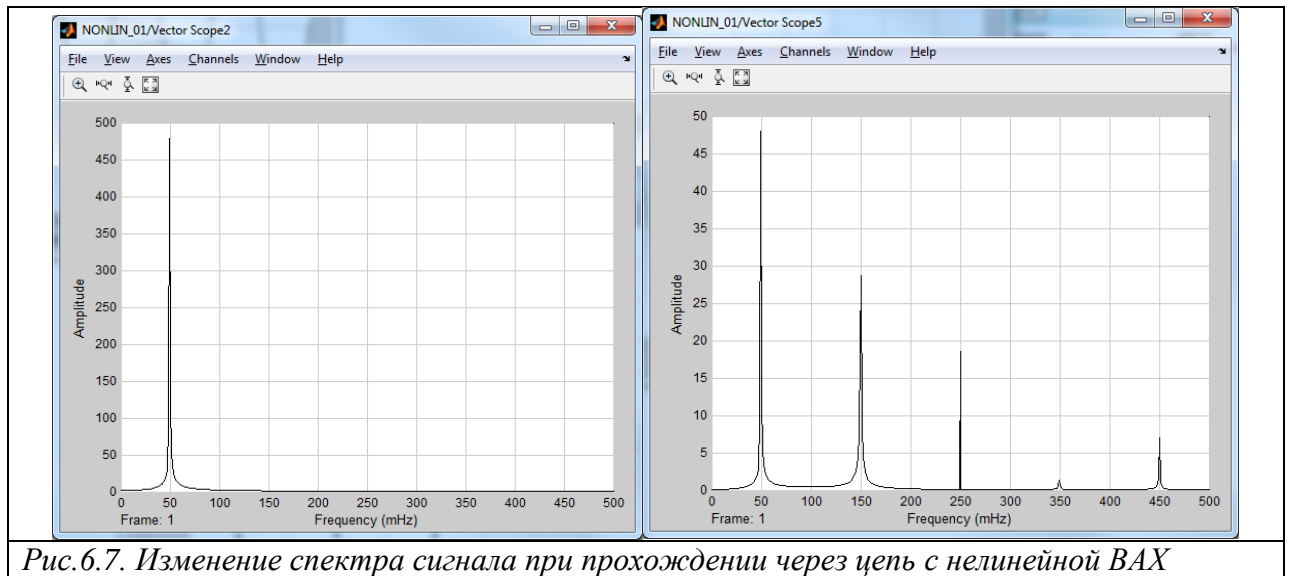


Рис.6.7. Изменение спектра сигнала при прохождении через цепь с нелинейной ВАХ

6.3.2. Кусочно-линейная аппроксимация ВАХ

Если для аппроксимации нелинейной ВАХ двухполюсника использовать кусочно-линейную функцию, то выражения, описывающие характер возникающего в цепи тока, получатся иными.

Предположим, что нелинейная ВАХ может быть аппроксимирована следующим выражением

$$i = \begin{cases} S(u - u_n), & \text{при } u \geq u_n, \\ 0, & \text{при } u < u_n \end{cases} \quad (6.11)$$

где S - крутизна характеристики,

u_n - абсцисса начала линейно-возрастающего участка аппроксимации ВАХ.

Подставляя в данное уравнение функцию, описывающую возбуждающее гармоническое колебание, получим

$$i = \begin{cases} S(u_0 + U_m \cos(\omega_1 t) - u_n), & \text{при } |\omega_1 t - 2\pi k| \leq \theta, \\ 0, & \text{при } |\omega_1 t - 2\pi k| \geq \theta \end{cases} \quad (6.12)$$

где θ - угол отсечки, величина которого находится из очевидного соотношения

$$\cos \theta = \frac{u_n - u_0}{U_m}. \quad (6.13)$$

Графики, поясняющие введенные обозначения, приведены на рис.6.8а, результаты моделирования изменений в выходном сигнале в зависимости от выбора рабочей точки продемонстрированы с помощью программы Multisim и приведены на рис. 6.8.б.

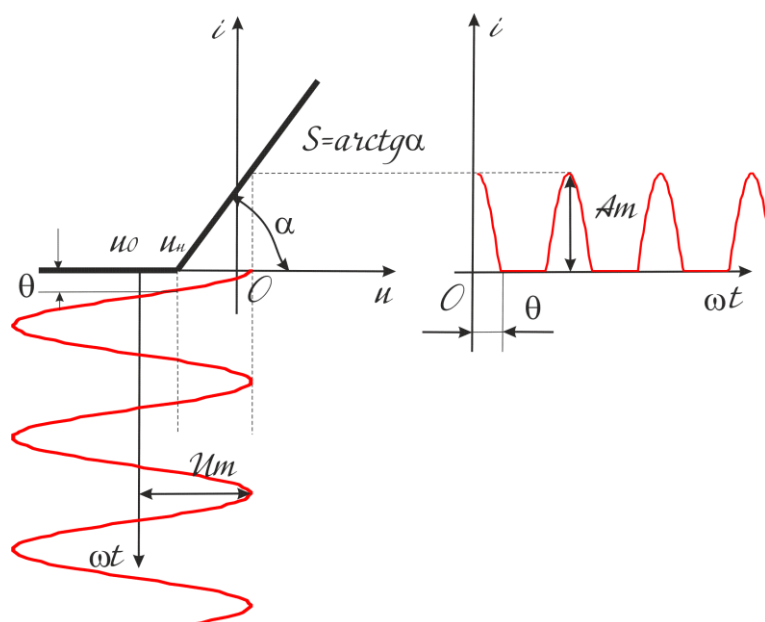


Рис.6.8а. Кусочно-линейная аппроксимация нелинейной ВАХ

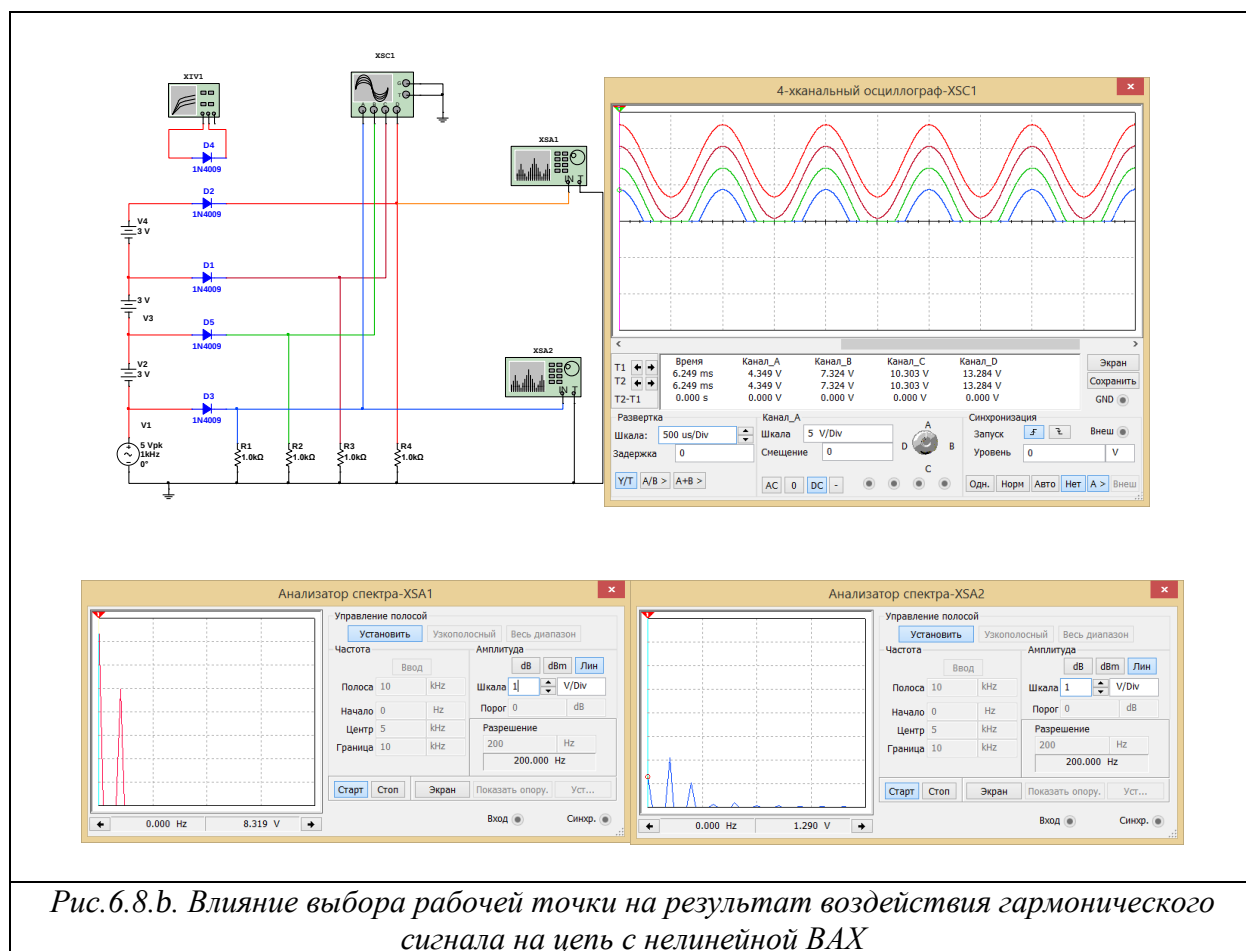


Рис.6.8.б. Влияние выбора рабочей точки на результат воздействия гармонического сигнала на цепь с нелинейной ВАХ

С учетом угла отсечки (6.13) уравнение для тока в цепи (6.12) может быть преобразовано к виду

$$i(t) = SU_m(\cos\omega_1 t - \cos\theta), \quad \text{при } 2k\pi - \theta \leq \omega_1 t \leq 2k\pi + \theta.$$

Максимальное значение тока достигается в моменты времени $t = \frac{2k\pi}{\omega_1}$, что позволяет записать

$$I_m = SU_m(\cos\omega_1 t - \cos\theta) \Big|_{t=\frac{2k\pi}{\omega_1}}$$

откуда

$$I_m = SU_m(1 - \cos\theta).$$

Полученные соотношения позволяют сделать вывод о том, что протекающий в исследуемой нелинейной цепи ток будет периодической функцией с периодом $T = 2\pi/\omega_1$. Поэтому соответствующую току временную функцию можно разложить в ряд Фурье

$$i(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos(k\omega_1 t) + b_k \sin(k\omega_1 t)).$$

Кроме того, ток в рассматриваемой цепи представляет собой четную функцию, поэтому все коэффициенты при синусных членах ряда будут нулевыми, то есть $b_k = 0$ для $\forall k$.

Тогда

$$a_k = \frac{2}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} i(t) \cos(k\omega_1 t) dt = \frac{2}{T\omega_1} \int_{-\theta}^{\theta} i(t) \cos(k\omega_1 t) d(\omega_1 t).$$

С учетом ранее полученных соотношений для тока и принимая во внимание, что $T\omega_1 = 2\pi$, получаем

$$a_k = \frac{I_m}{\pi(1 - \cos\theta)} \int_{-\theta}^{\theta} (\cos\omega_1 t - \cos\theta) \cos(k\omega_1 t) d(\omega_1 t).$$

Интегрирование данного уравнения позволяет получить формулы для амплитуды k -ой гармоники (при $k > 0$) тока в виде

$$I_k = I_m \frac{2(\sin k\theta \cos\theta - k \cos k\theta \sin\theta)}{k\pi(k^2 - 1)(1 - \cos\theta)}. \quad (6.14)$$

Постоянная составляющая тока $I_0 = a_0/2$.

Для соответствующих гармоник из полученного (6.14) следует

$$I_0 = I_m \frac{\sin\theta - \theta \cos\theta}{\pi(1 - \cos\theta)},$$

$$I_1 = I_m \frac{\theta - \sin\theta \cos\theta}{\pi(1 - \cos\theta)},$$

$$I_2 = I_m \frac{2\sin^3\theta}{3\pi(1 - \cos\theta)}$$

и так далее.

Коэффициенты, связывающие между собой амплитуды гармоник и максимальное значение тока, называют коэффициентами Берга

$$\alpha_k(\theta) = \frac{I_k}{I_m} \text{ или } I_m = \alpha_k(\theta) I_k. \quad (6.15)$$

Иногда вместо коэффициентов Берга (6.15) используется функция Берга, определяемая как

$$\gamma_k(\theta) = \frac{I_k}{SU_m}. \quad (6.16)$$

Связь между коэффициентами Берга (6.15),(6.16) и функцией Берга определяется из уравнения

$$\gamma_k(\theta) = (1 - \cos\theta)\alpha_k(\theta).$$

Графики коэффициентов Берга для нескольких значений k приведены на рис.6.9.

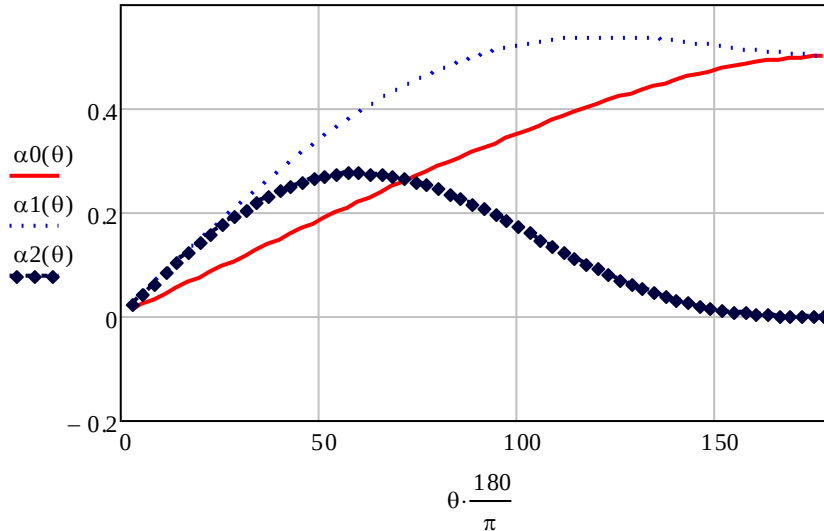


Рис.6.9. Графики коэффициентов Берга

6.3.3. Воздействие на нелинейную цепь бигармонического колебания

В предыдущих параграфах мы рассмотрели результаты воздействия на двухполусник с нелинейной ВАХ одного гармонического колебания.

Рассмотрим процессы, возникающие в нелинейной цепи при ее возбуждении суммой двух гармонических колебаний.

Возбуждающее напряжение, прикладываемое к цепи, в этом случае будем описывать уравнением

$$u(t) = U_{m1} \cos(\omega_1 t) + U_{m2} \cos(\omega_2 t). \quad (6.17)$$

Для упрощения анализа предположим, что ВАХ (6.5) анализируемой цепи может быть аппроксимирована полиномом второй степени без нулевого члена

$$i = a_1 u + a_2 u^2.$$

После подстановки выражений для возбуждающего напряжения в уравнение для тока двухполусника и проведения преобразований, аналогичных тем, что были выполнены для случая моногармонического сигнала, получаем следующее.

$$i(t) = a_1 U_{m1} \cos(\omega_1 t) + a_1 U_{m2} \cos(\omega_2 t) + a_2 (U_{m1}^2 \cos^2(\omega_1 t) + U_{m2}^2 \cos^2(\omega_2 t) + 2U_{m1} U_{m2} \cos(\omega_1 t) \cos(\omega_2 t)).$$

Квадратичный член этого выражения в соответствии с уже использованным тригонометрическим равенством $\cos^2 \beta = \frac{1}{2}(1 + \cos 2\beta)$ можно далее преобразовать к виду

$$a_2 u^2 = \frac{1}{2} a_2 (U_{m1}^2 + U_{m2}^2) + \frac{1}{2} a_2 U_{m1}^2 \cos 2(\omega_1 t) + \frac{1}{2} a_2 U_{m2}^2 \cos 2(\omega_2 t) + a_2 U_{m1} U_{m2} [\cos(\omega_1 + \omega_2) t + \cos(\omega_1 - \omega_2) t]. \quad (6.18)$$

Анализ полученных соотношений (6.17), (6.18) показывает, что в составе уравнения для тока нелинейной цепи имеется несколько принципиально различных составляющих:

- постоянная составляющая, которая не зависит от времени и определяет приращение постоянного тока,
- составляющие с удвоенными частотами $2\omega_1, 2\omega_2$, представляющие собой вторые гармоники от соответствующих возбуждающих напряжений,
- комбинационные колебания $(\omega_1 + \omega_2)$ и $(\omega_1 - \omega_2)$.

Если для аппроксимации ВАХ цепи использовать полином более высокого порядка N , то среди составляющих возникающего тока могут присутствовать следующие частоты:

$\omega = 0$ – постоянная составляющая,

$\omega = n\omega_1$ и $\omega = n\omega_2$, где $n = 1, 2, \dots, N$ – гармоники частот входных напряжений,

$\omega = |n\omega_1 \pm m\omega_2|$, где $n = 1, 2, \dots, N, m = 1, 2, \dots, N, n + m \leq N$ – комбинационные частоты.

Свойство квадратичного нелинейного элемента, приводящее к появлению комбинационных частот, широко применяется в радиотехнике для сдвига частоты сигнала.

Аналогично проведенному анализу нетрудно получить, что при любом периодическом (с основной частотой ω) воздействии на нелинейную цепь на выходе также будет наблюдаться периодический процесс с той же основной частотой ω , однако спектр этого выходного сигнала по сравнению со спектром возбуждающего напряжения может быть существенно обогащен за счет более высоких гармоник $n\omega$.

Рассмотренные выше воздействия гармонического сигнала на нелинейный двухполусник были связаны с преобразованием приложенного возбуждающего напряжения в ток. Для того чтобы получить дальнейшее преобразование тока в выходное напряжение требуется перейти к схеме четырехполусника, простейший вариант которого приведен на рис.6.10.

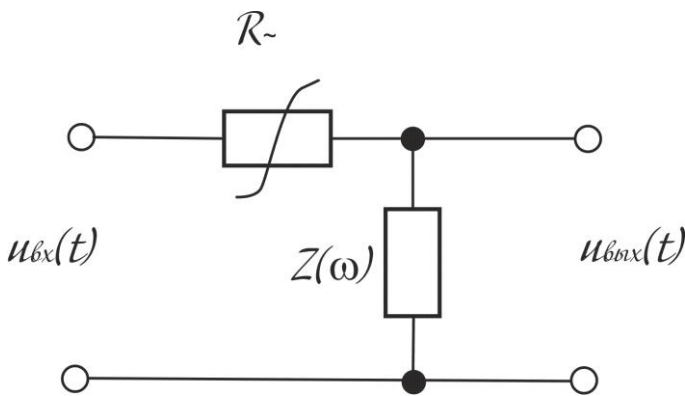


Рис.6.10. Схема четырехполусника с нелинейным элементом

В данном случае включенное в состав четырехполусника комплексное сопротивление нагрузки $Z_n^\circ(\omega)$ выполняет роль фильтра, на котором выделяется выходное напряжение цепи. Внесение таким способом в цепь дополнительного сопротивления $Z_n^\circ(\omega)$ приводит к тому, что выходное напряжение начинает оказывать влияние на нелинейный элемент. Поэтому спектральный состав сигнала может отличаться от первоначального состава тока в цепи. Это явление, получившее название

реакцией нагрузки, следует учитывать при разработке конкретных радиотехнических цепей.

В ряде случаев, когда справедливо предположение $|u_{vbx}(t)| \ll |u_{bx}(t)|$, реакцией нагрузки можно пренебречь.

Проведенный анализ показывает, что использование нелинейного резистивного элемента в сочетании частотно избирательными цепями позволяет осуществить целый ряд

востребованных в радиотехнике преобразований сигнала, включая резонансное усиление, умножение частоты, амплитудную модуляцию и многие другие. Часть из этих преобразований мы подробно рассмотрим в дальнейших разделах курса.

6.3.4. Преобразование частоты сигнала (гетеродирование)

В радиотехнике часто возникает задача сдвига сигнала по оси частот на заданную постоянную величину при сохранении общей структуры сигнала. Данная операция называется преобразованием частоты.

Рассмотрим случай, когда на нелинейный элемент подается два сигнала – один гармонический сигнал

$$u_r(t) = U_r \cos(\omega_r t + \theta_r), \quad (6.19)$$

создаваемый специальным вспомогательным автогенератором, называемым гетеродином, а второй – АМ-сигнал.

Для простоты рассмотрения в качестве АМ-сигнала мы будем использовать вариант тональной модуляции, то есть

$$\begin{aligned} u(t) &= U_m(1 + M \cos(\Omega t)) \cos[\omega_0 t + \theta_0] = \\ &= U_m \cos[\omega_0 t + \theta_0] + \frac{MU_m}{2} \cos\{(\omega_0 + \Omega)t + \theta_0\} \\ &\quad + \frac{MU_m}{2} \cos\{(\omega_0 - \Omega)t + \theta_0\}. \end{aligned} \quad (6.20)$$

Если аппроксимировать ВАХ нелинейного элемента параболой, то есть

$$i = a_2 u^2,$$

после проведения соответствующего разложения по частотам и выделения необходимых гармоник суммарных и разностных частот с учетом (6.19) и (6.20) получим

$$\begin{aligned} i_\phi(t) &= a_2 U_m U_r (1 + M \cos(\Omega t)) \cos[(\omega_0 + \omega_r) t + \theta_0 + \theta_r] + \\ &\quad + a_2 U_m U_r (1 + M \cos(\Omega t)) \cos[(\omega_0 - \omega_r) t + \theta_0 - \theta_r]. \end{aligned} \quad (6.21)$$

В результате в составе тока нелинейного элемента (6.21) присутствуют компоненты исходного АМ-сигнала, но со смещенным относительно ω_0 несущим колебанием. Величина этого сдвига равна частоте колебаний гетеродина ω_r .

6.4. Нелинейное резонансное усиление и умножение частоты

Принципиальное отличие нелинейного резонансного усилителя от его линейного аналога состоит в использовании режима работы с отсечкой.

Упрощенная принципиальная схема такого усилителя приведена на рис.6.11.

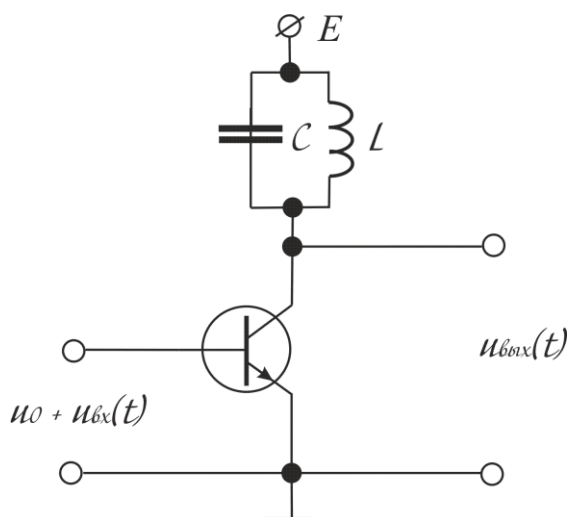


Рис.6.11. Схема резонансного усилителя

рассмотрено в предыдущем параграфе в уравнениях (6.11)-(6.13) , то без учета нагрузки в выходной цепи токи гармоник могут быть определены по уравнениям

$$I_k = \gamma_k(\theta) S U_m,$$

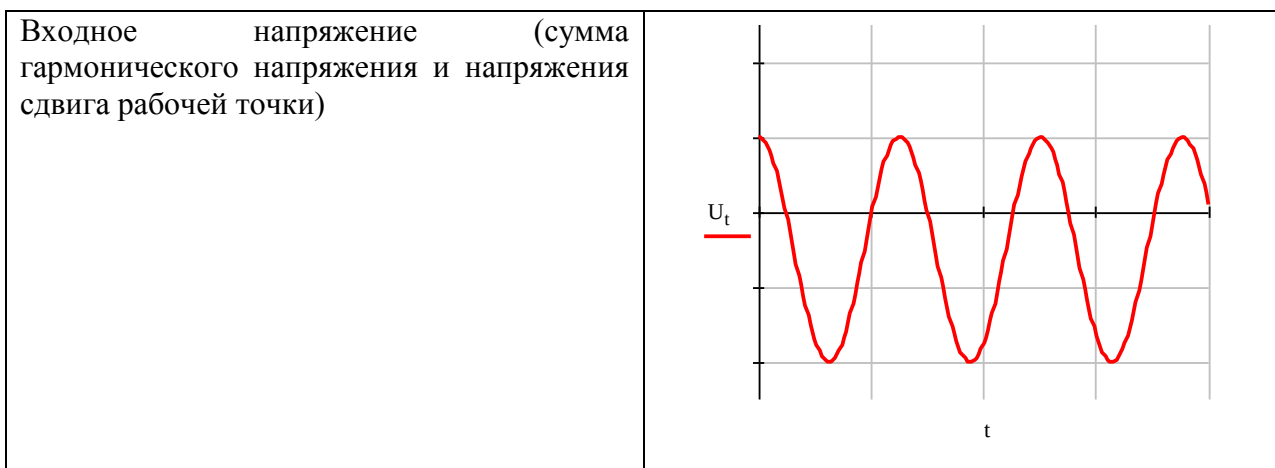
где θ – угол отсечки, определяемый выбором рабочей точки.

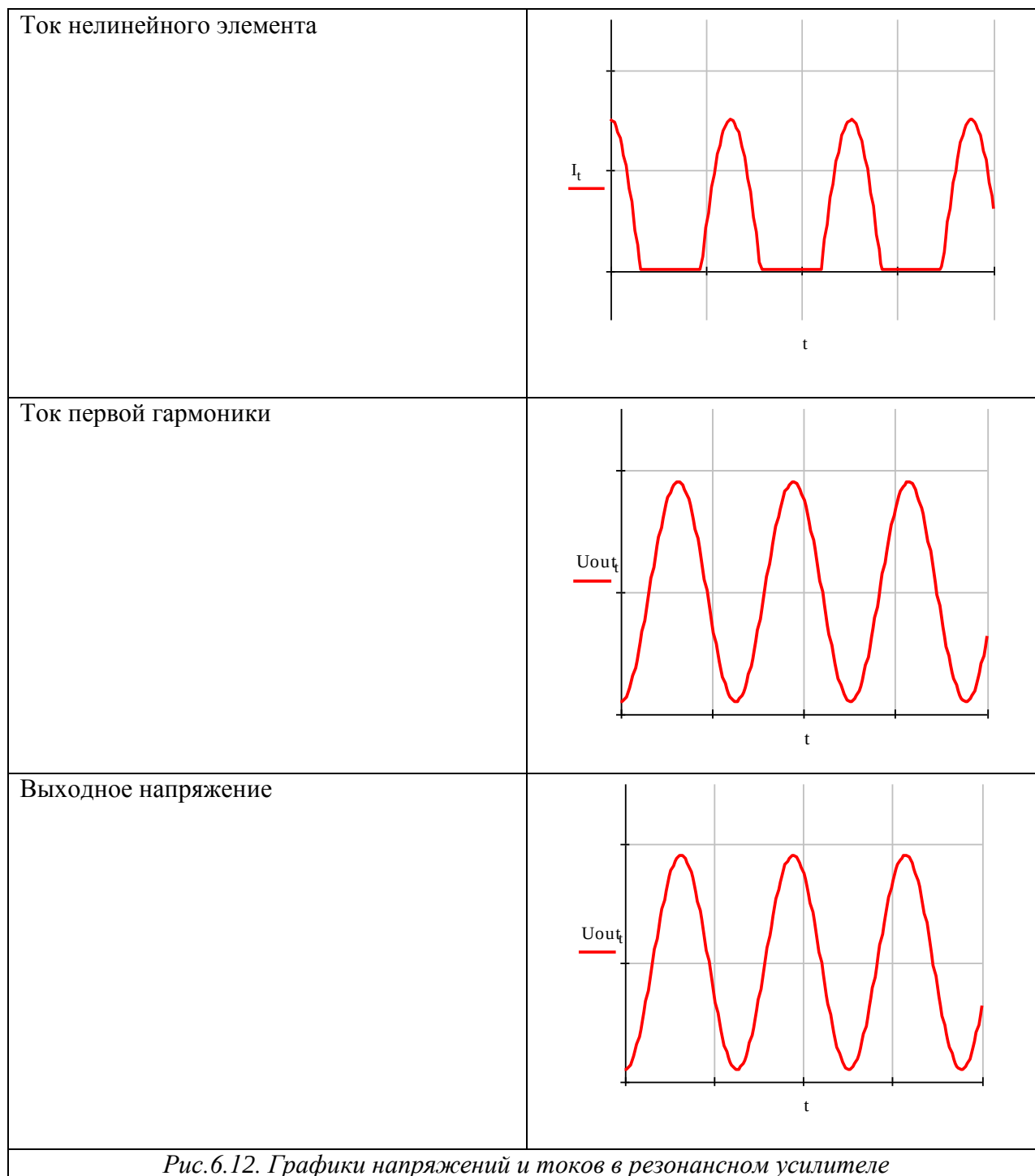
Если параметры LC контура выбраны таким образом, чтобы его резонансная частота соответствовала частоте первой гармоники ω_1 , то

$$U_{m_{\text{вых}}} = \gamma_k(\theta) S U_m r_{\Sigma},$$

где r_{Σ} – эквивалентное сопротивление контура.

Графики временных зависимостей напряжения и токов в анализируемом усилительном каскаде представлены на рис.6.12.





В этом случае выходное напряжение усилителя можно представить в виде

$$u_{\text{вых}}(t) = E - U_m \cos(\omega_1 t),$$

где E – напряжение питания усилительного каскада.

Из приведенных графиков видно, что выходной сигнал нелинейного резонансного усилителя, во-первых, находится в противофазе с входным напряжением (то есть при рассмотренной схеме включения происходит инвертирование входного сигнала), а во-вторых, представляет собой колебания относительно уровня, определяемого напряжением питания усилительного каскада. Следовательно, в некоторые моменты времени выходное

напряжение может превышать напряжение питания, что и обуславливает высокие значения коэффициентов усиления в нелинейных резонансных усилителях.

Качество работы нелинейного резонансного усилителя оценивают формой зависимости

$$U_{m\text{вых}} = f(U_{m\text{вх}}),$$

приведенной на рис.6.13. и называемой колебательной характеристикой усилителя.

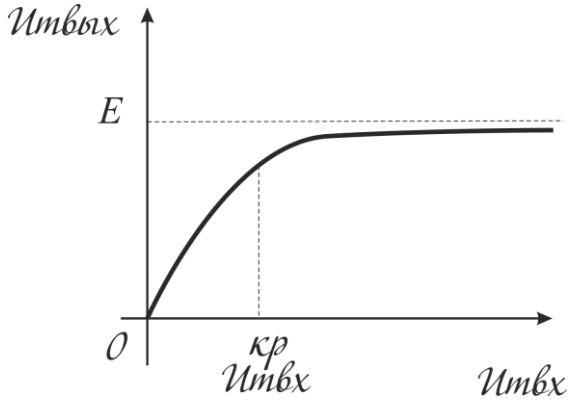


Рис.6.13. Колебательная характеристика нелинейного резонансного усилителя

Для усиления АМ-сигналов используют линейный участок данной характеристики, выбирая режим работы таким образом, чтобы

$$U_{m\text{вх}} < U_{m\text{вх}}^{\text{кр}}.$$

В противном случае режим работы усилителя называют перенапряженным и его используют для ЧМ и ФМ сигналов, поскольку в таком режиме работы дополнительно ослабляется паразитная амплитудная модуляция этих сигналов.

Если в рассмотренной схеме нелинейного резонансного усилителя LC контур настроить на частоту $n\omega_1$, то данное устройство может использоваться в качестве умножителя частоты.

Умножители частоты обычно применяются при создании источников гармонических колебаний с высокой стабильностью частоты, если непосредственное генерирование в заданном частотном диапазоне такой стабильности обеспечить не может. Тогда задача решается путём каскадного соединения низкочастотного стабильного генератора и одного или нескольких умножителей частоты.

Амплитуда выходного сигнала умножителя

$$U_{m\text{вых}} = \gamma_k(\theta) S U_{m\text{вх}} R_{\text{рез}},$$

где $R_{\text{рез}}$ – резонансное сопротивление контура, настроенного на $n\omega_1$.

Трудность реализации резонансных умножителей частоты заключается в том, что при большой кратности умножителя коэффициенты $\gamma_k(\theta)$ имеют низкие значения. Для максимизации значения коэффициента Берга $\gamma_k(\theta)$ углы отсечки рекомендуется выбирать следующим образом

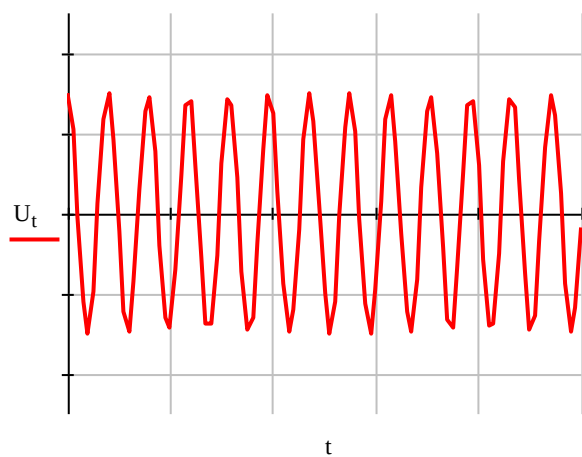
$$\theta_{\text{опт}} \approx \frac{180^\circ}{n}.$$

6.5. Амплитудная модуляция

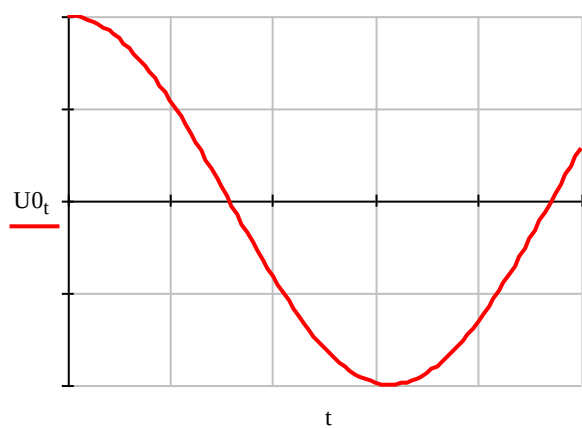
Рассмотренная в предыдущем параграфе схема нелинейного резонансного легко может быть преобразована в амплитудный модулятор (АМ-модулятор).

Для этого на вход нелинейного усилителя кроме постоянного напряжения смещения u_0 необходимо подать переменное напряжение, соответствующее передаваемому сигналу, а колебательный контур настроить на частоту несущего колебания. Форма сигналов на входе и выходе такого модулятора для случая однотоновой модуляции представлена на рис. 6.14.

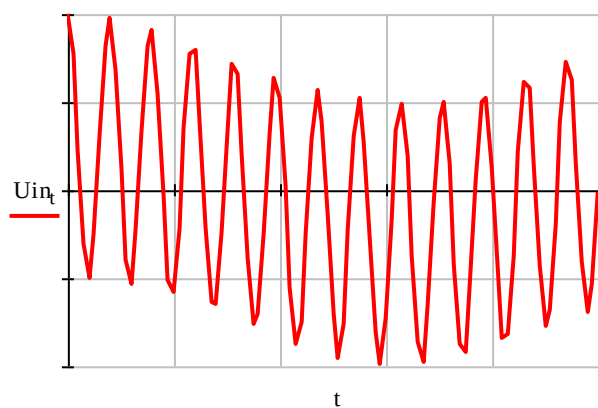
Входное несущее колебание



Модулирующее напряжение



Суммарное входное напряжение нелинейного усилителя



Выходной ток нелинейного элемента

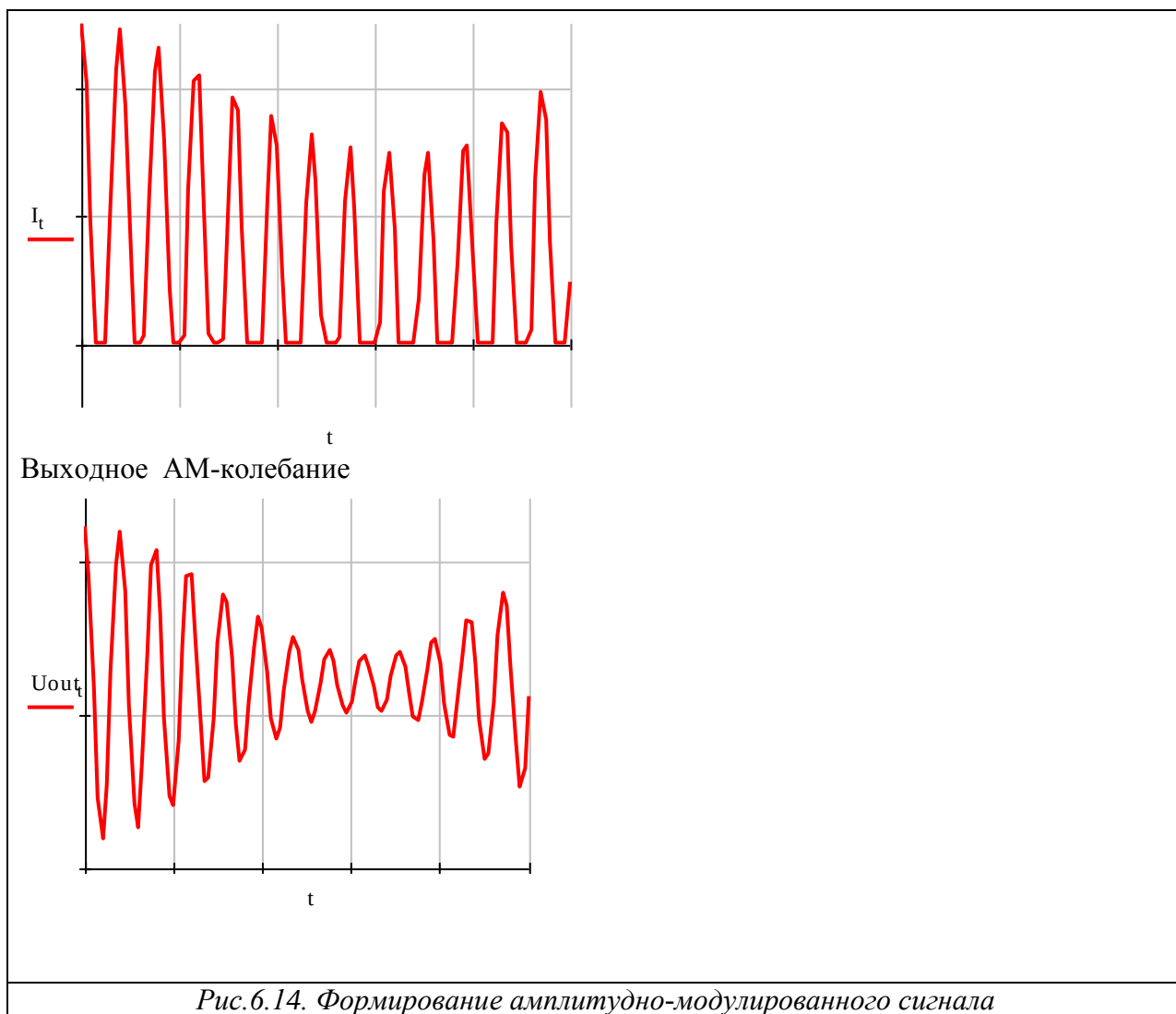


Рис.6.14. Формирование амплитудно-модулированного сигнала

Полученный описанным способом простейший амплитудный модулятор имеет входной сигнал вида

$$u(t) = u_0 + U_{\text{мнec}} \cos(\omega_1 t) + U_{\text{мmod}} \cos(\Omega t).$$

Пусть нелинейная характеристика транзистора аппроксимируется двумя прямолинейными отрезками (как показано на рис.6.15.).

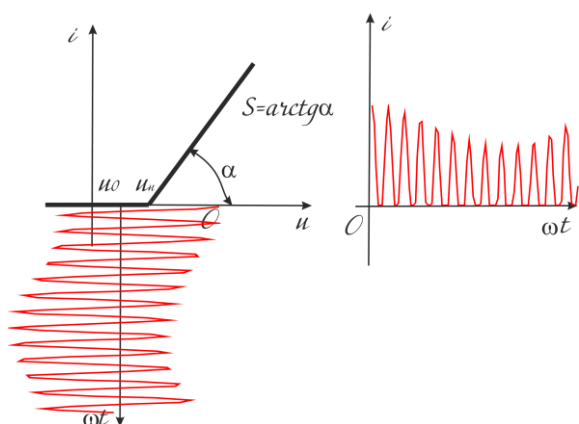


Рис.6.15. Процесс формирования АМ-сигнала

Тогда из-за наличия низкочастотного модулирующего колебания рабочая точка не будет оставаться на месте, а будет перемещаться синхронно с модулирующим колебанием, что приводит к непрерывному изменению угла отсечки. В результате амплитуда первой гармоники также не является постоянной, а изменяется в такт с перемещением рабочей точки.

Колебательный контур произведет фильтрацию возникающего коллекторного тока, выделяя на выходе АМ-сигнал.

Ранее рассматривая характеристики нелинейного двухполюсника в (6.13), мы получали

$$\cos\theta = \frac{u_n - u_0}{U_m}.$$

В настоящем случае постоянное напряжение сдвига u_0 было дополнено гармонической составляющей, в результате чего угол отсечки перестал быть постоянным.

Для максимального и минимального значений угла отсечки в данном случае можно записать

$$\begin{aligned}\theta_{\min} &= \arccos \frac{u_n - u_0 - U_{\text{мод}}}{U_{\text{нес}}}, \\ \theta_{\max} &= \arccos \frac{u_n - u_0 + U_{\text{мод}}}{U_{\text{нес}}},\end{aligned}$$

что в свою очередь позволяет рассчитать минимальную и максимальную амплитуды первой гармоники тока с учетом того, что $I_1 = \gamma_1(\theta)SU_{\text{нес}}$. Если параметры модулятора подобраны таким образом, что модуляционная зависимость $I_1 = f(U_{\text{мод}})$ является линейной, то коэффициент модуляции можно определить по формуле

$$M = \frac{I_{1\max} - I_{1\min}}{I_{1\max} + I_{1\min}}. \quad (6.22)$$

В последнее время в модуляторах АМ все шире начинают применяться специализированные аналоговые интегральные микросхемы, получившие название аналоговых перемножителей напряжений и реализующие функцию

$$u_{\text{вых}}(t) = Ku_1(t)u_2(t),$$

где K – постоянный коэффициент.

Если на входы такого устройства подать сигнал $U_{\text{мод}} \cos(\Omega t)$ и несущее колебание $U_{\text{нес}} \cos(\omega_1 t)$, то на выходе получим сигнал

$$\begin{aligned}u_{\text{вых}}(t) &= KU_{\text{мод}} \cos(\Omega t) U_{\text{нес}} \cos(\omega_1 t) = \\ &= \frac{KU_{\text{мод}}U_{\text{нес}}}{2} \cos((\omega_1 + \Omega)t) + \frac{KU_{\text{мод}}U_{\text{нес}}}{2} \cos((\omega_1 - \Omega)t),\end{aligned}$$

который состоит из нижней и верхней боковых частот АМ-сигнала. Такой сигнал мы изучали ранее и называли его АМ-сигналом с подавленной несущей. Устройство, используемое для его формирования, носит название балансного модулятора.

Если в АМ сигнале требуется сохранить несущее колебание, то это легко сделать путем простого суммирования (с масштабным коэффициентом) выходного сигнала балансного модулятора и самого несущего колебания.

Схематично процесс получения АМ-сигнала с восстановленной несущей представлен на рис.6.16.

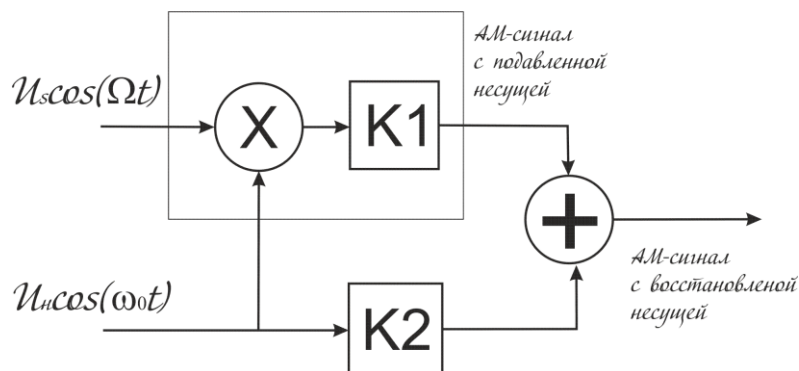


Рис.6.16. Формирование АМ-сигнала с восстановленной несущей

6.6. Амплитудное детектирование

6.6.1. Линейный диодный детектор

Амплитудное детектирование представляет собой операцию обратную амплитудной модуляции. Входным сигналом амплитудного детектора является сигнал

$$u_{AM}(t) = U_{mвх}(1 + M\cos(\Omega t))\cos\omega_0 t,$$

а на выходе требуется получить сигнал вида

$$u_{вых}(t) = U_{mвых}\cos(\Omega t).$$

Значения начальных фаз в рассматриваемой задаче не являются принципиальными и поэтому для простоты записи приняты равными нулю.

Эффективность работы амплитудных детекторов принято характеризовать коэффициентом детектирования, который равен отношению амплитуды выходного низкочастотного сигнала к амплитуде огибающей входного АМ-сигнала:

$$K_d = \frac{U_{mвых}}{MU_{mвх}}.$$

В зависимости от вида характеристики используемого для детектирования нелинейного элемента и амплитуды АМ-сигнала различают два основных режима детектирования:

- линейный режим (большие амплитуды и кусочно-линейная аппроксимация характеристики нелинейного элемента);
- квадратичный режим (малые значения амплитуд и аппроксимация рабочего участка характеристики полиномом 2-го порядка).

Схема простейшего диодного детектора приведена на рис.6.17.

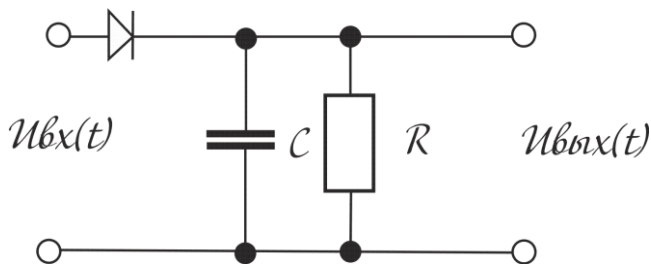


Рис.6.17. Схема диодного детектора

Для использованного в приведенной схеме диода воспользуемся упрощенной кусочно-линейной аппроксимацией ВАХ

$$i(u) = \begin{cases} Su, & u \geq 0 \\ 0, & u < 0 \end{cases}$$

где S – крутизна характеристики.

Сопротивление нагрузки выбирается таким образом, чтобы оно значительно превышало внутренне сопротивление открытого диода, то есть $R \gg r = 1/S$.

С другой стороны из схемы видно, что представленный детектор по своей сути является однополупериодным выпрямителем с емкостным фильтром. Поэтому требуется подавление имеющих место пульсаций на несущей частоте ω_0 . Соответственно,

$$RC \gg \frac{1}{\omega_0}.$$

При этом задача детектора состоит в отслеживании изменений амплитуды полезного сигнала, который в нашем случае представлен моногармоническим колебанием с частотой Ω . Следовательно, необходимо обеспечить выполнение условия

$$RC \ll \frac{1}{\Omega}.$$

Таким образом, диапазон допустимых значений для параметров детектора определяется двойным неравенством

$$\frac{1}{\omega_0} \ll RC \ll \frac{1}{\Omega}. \quad (6.23)$$

Детальный анализ работы диодного детектора приводит к следующей формуле для коэффициента детектирования

$$K_d = \cos \left(\sqrt[3]{\frac{3\pi}{SR}} \right),$$

то есть всегда остается меньше единицы.

Важным параметром диодного детектора является его входное сопротивление. Понятно, что входное сопротивление схемы, работающей с отсечкой, не является постоянным. Поэтому в данном случае под входным сопротивлением понимается некоторая усредненная величина, определяемая как отношение амплитуды входного напряжения к амплитуде первой гармоники тока.

На частоте несущего колебания входное сопротивление диодного детектора оказывается равным

$$R_{\text{вх}} = \frac{R}{2}.$$

Пример 6.4.

Рассмотрим работу простейшего диодного амплитудного детектора.

Исследуемая схема и результат моделирования приведен на рис. 6.18.

В качестве дополнительного упражнения рекомендуем проверить оптимальность выбора параметров R и C исходя из частотных параметров детектируемого сигнала, заменить их значения в случае необходимости, сравнить полученный результат с приведенным на рисунке и объяснить обнаруженную разницу в форме выходного сигнала.

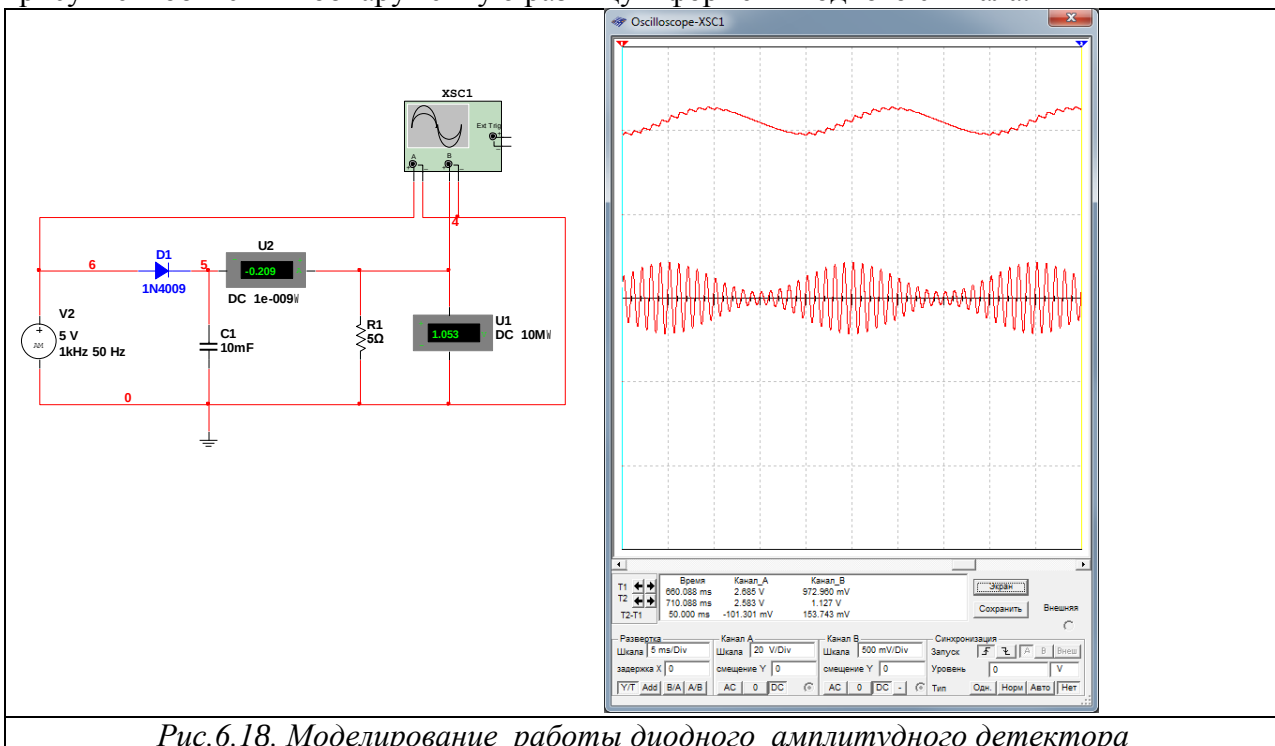


Рис.6.18. Моделирование работы диодного амплитудного детектора

6.6.2. Коллекторный детектор

Рассмотрим схему коллекторного детектора, который представляет собой усилительный каскад с нагрузкой в виде параллельной RC-цепи, приведенную на рис.6.19.

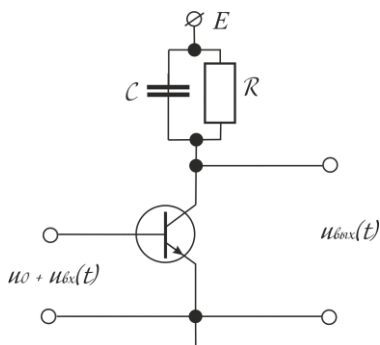


Рис.6.19. Схема коллекторного детектора

На вход данного каскада поступает АМ-сигнал совместно с напряжением смещения u_0 .

Для упрощения процесса анализа примем кусочно-линейную аппроксимацию ВАХ транзистора и будем полагать, что $u_0 = u_n$, что соответствует углу отсечки $\theta = 90^\circ$ и не зависит от входного напряжения. Схема преобразования сигналов для такого случая приведена на рис.6.20.

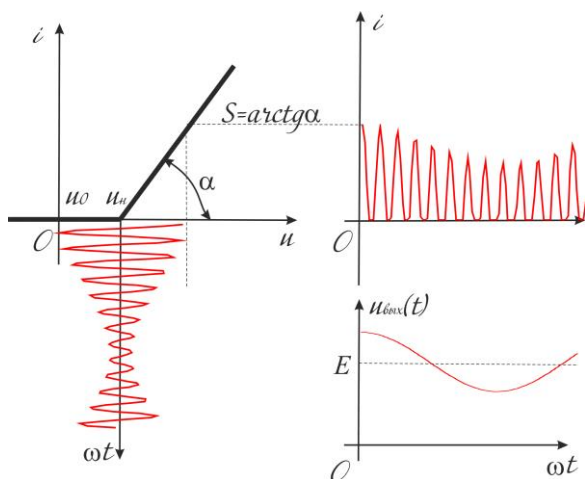


Рис.6.20. Амплитудное детектирование с применением коллекторного детектора

Коллекторный ток может быть представлен как сумма возникающих в процессе детектирования гармоник

$$i(t) = \sum_{k=0}^{\infty} S U_{\text{вх}}(t) \gamma_k(90^\circ) \cos(k \omega_0 t).$$

Если величину емкости выбрать таким образом, чтобы выполнялось условие (6.23), которое мы приводили при рассмотрении диодного детектора

$$\frac{1}{\omega_0} \ll RC \ll \frac{1}{\Omega},$$

то на всех высших гармониках резистор будет фактически шунтирован емкостью, и реальное падение напряжения на нем будет вызывать только составляющая тока I_0 .

Тогда

$$\begin{aligned} u_{\text{вых}}(t) &= E - I_0 R = E - S R U_{\text{вх}}(t) \gamma_0(90^\circ) = \\ &= E - 0,318 S R U_{\text{мвх}} (1 + M \cos(\Omega t)). \end{aligned}$$

Отсюда коэффициент детектирования может быть определен по выражению

$$K_d = 0,318 S R.$$

Полученное выражение наглядно показывает преимущества коллекторного детектора, в котором коэффициент детектирования может быть доведен в реальных условиях до тысячи, в то время как диодный детектор имеет коэффициент детектирования меньше единицы.

6.6.3. Квадратичный детектор

При малых значениях амплитуды подлежащего детектированию сигнала реальные характеристики нелинейных элементов (6.4) хорошо аппроксимируются полиномами второй степени

$$i(t) = a_1 u + a_2 u^2.$$

Подставляя в это уравнение выражение для входного АМ-сигнала

$$u_{AM}(t) = U_{mBX}(1 + M \cos(\Omega t)) \cos \omega_0 t = U_{BX}(t) \cos \omega_0 t,$$

получим

$$i(t) = a_1 U_{BX}(t) \cos \omega_0 t + \frac{1}{2} a_2 U_{BX}^2(t) + \frac{1}{2} a_2 U_{BX}^2(t) \cos 2\omega_0 t \quad (6.24)$$

Высокочастотные составляющие тока в (6.24) с частотами ω_0 и $2\omega_0$ отфильтровываются соответствующим фильтром, в результате чего остается только низкочастотная составляющая $\frac{1}{2} a_2 U_{BX}^2(t)$, которая может быть представлена как

$$i_{нч}(t) = a_2 U_{mBX}^2 M \cos(\Omega t) + \frac{1}{4} a_2 U_{mBX}^2 M^2 \cos(2\Omega t).$$

На нагрузке детектора R этот ток вызовет напряжение

$$u_{вых}(t) = R a_2 U_{mBX}^2 M \cos(\Omega t) + \frac{R}{4} a_2 U_{mBX}^2 M^2 \cos(2\Omega t). \quad (6.25)$$

Первое слагаемое этого напряжения в (6.25) представляет собой полезный результат детектирования. Поскольку амплитуда полезного выходного напряжения пропорциональна квадрату амплитуды информационного сигнала, то такой вид детектирования является квадратичным. Коэффициент детектирования при этом определяется выражением

$$K_d = R a_2 U_{mBX}.$$

Второе слагаемое полезной информации не несет, а лишь искажает получаемый результат. Эффект его воздействия оценивают так называемым коэффициентом нелинейных искажений, равным отношению амплитуд выходных колебаний на частотах Ω и 2Ω . В данном случае

$$K_{ни} = \frac{M}{4}.$$

Как видно из рассмотренного, нелинейные искажения в квадратичном детекторе оказываются весьма существенными, особенно при глубокой модуляции. По этой причине в радиоприемниках АМ-сигнал обязательно предварительно усиливается, чтобы на входе детектора он составлял единицы Вольт, что переводит детектор в режим линейного детектирования с малым количеством нелинейных искажений.

6.6.4. Синхронное детектирование

Синхронное детектирование является частным случаем преобразования частоты, при котором частота колебаний гетеродина совпадает с частотой несущих колебаний АМ-сигнала.

Ранее для операции преобразования частоты мы получали уравнение

$$i_\phi(t) = a_2 U_m U_r (1 + M \cos(\Omega t)) \cos[(\omega_0 + \omega_r) t + \theta_0 + \theta_r] + a_2 U_m U_r (1 + M \cos(\Omega t)) \cos[(\omega_0 - \omega_r) t + \theta_0 - \theta_r],$$

что при подстановке $\omega_r = \omega_0$ приводит к выражению для медленно меняющейся составляющей тока

$$i_\phi(t) = a_2 U_m U_r (1 + M \cos(\Omega t)) \{ \cos(\theta_0 - \theta_r) + \cos[2\omega_0 t + \theta_0 + \theta_r] \},$$

которая после НЧ-фильтрации дает составляющую

$$i_{нч}(t) = i_\phi(t) = a_2 U_r u(t) \cos(\theta_0 - \theta_r),$$

совпадающую по форме с огибающей АМ-сигнала.

Данный метод детектирования называется синхронным и его основное достоинство состоит в возможности повышения качества приема слабых сигналов. К сожалению, практическая реализация данного метода связана с необходимостью обеспечения синхронизма частоты гетеродина с частотой принимаемого сигнала, что является достаточно сложной технической задачей особенно в условиях присутствия значительных шумов.

Пример 6.5.

Моделирование амплитудного детектора на операционном усилителе. Схема детектора и результаты моделирования приведены на рис. 6.21.

В качестве дополнительного упражнения рекомендуем объяснить принцип работы приведенной схемы и предложить значения параметров, наиболее подходящие детектируемому сигналу.

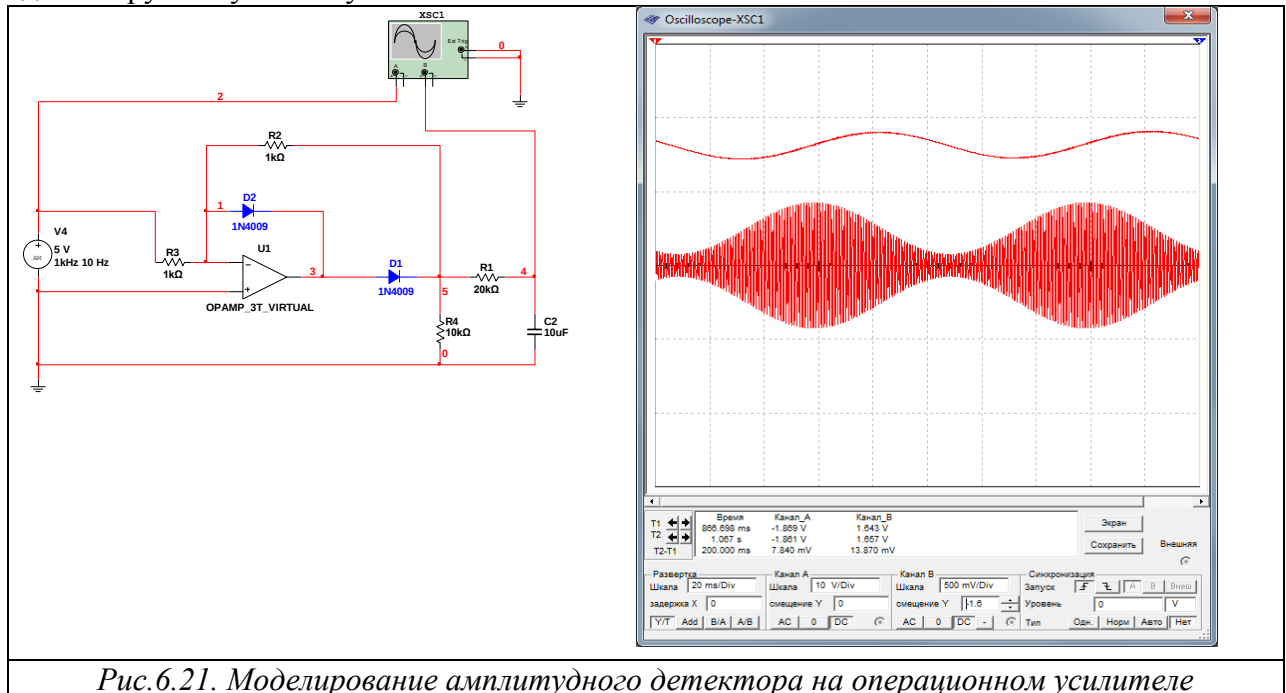


Рис.6.21. Моделирование амплитудного детектора на операционном усилителе

6.7. Частотное и фазовое детектирование сигналов

При использовании угловой модуляции информационный радиосигнал, как мы рассматривали ранее, можно представить в виде модулированного колебания

$$x(t) = A(t) \cos[\omega_0 t + \theta(t)] = A(t) \cos \psi(t),$$

где $A(t)$ и $\theta(t)$ являются информационными параметрами передаваемого сигнала.

Если $A(t) = \text{const}$ и $\theta(t) \neq \text{const}$, то получается модель радиосигнала с угловой модуляцией.

Поэтому задачей частотного детектора является выделение полезной информации из высокочастотного сигнала в виде

$$u_{\text{чд}}(t) = S_{\text{чд}} \frac{d}{dt} \theta(t),$$

где $S_{\text{чд}}$ – крутизна характеристики частотного детектора.

Соответственно для фазового детектора можно записать

$$u_{\text{фд}}(t) = S_{\text{фд}} \theta(t),$$

где $S_{\text{фд}}$ – крутизна характеристики фазового детектора.

При этом предполагается, что $u_{\text{чд}}(t)$ и $u_{\text{фд}}(t)$ являются "медленными" функциями времени.

Особенность частотного детектирования в отличие от амплитудного состоит в том, что использование для детектирования только одного нелинейного элемента будет недостаточно, так как при постоянстве амплитуды входного напряжения элемент не реагирует на изменения частоты входного напряжения. Поэтому для частотного детектирования нужны два устройства:

- избирательная линейная цепь, служащая для преобразования частотной модуляции в амплитудную;
- амплитудный детектор.

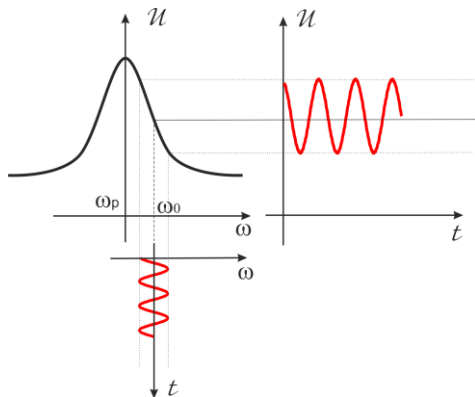


Рис.6.22. Принцип частотного детектирования

выполненная в программе Multisim.

В качестве линейной цепи может быть использована любая цепь с неравномерной частотной характеристикой, хотя в высокочастотных цепях наиболее часто используются колебательные контура и цепи на их основе.

Если колебательный контур настраивается таким образом, что его резонансная частота ω_p отличается от средней частоты модулированного сигнала ω_0 , как показано на рис. 6. 22 напряжение на контуре в определенных пределах повторяет изменение частоты входного напряжения.

Схема простейшего частотного детектора, состоящего из двух описанных частей, представлена на рис. 6.23а. На рис.6.23.б представлена модель схемы преобразования частотно-модулированного сигнала в амплитудно-модулированный,

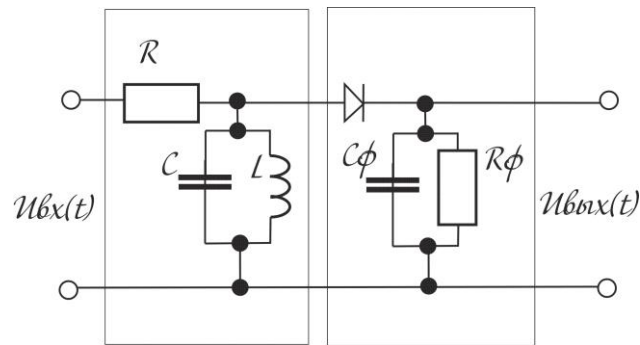


Рис.6.23а. Схема простейшего частотного детектора

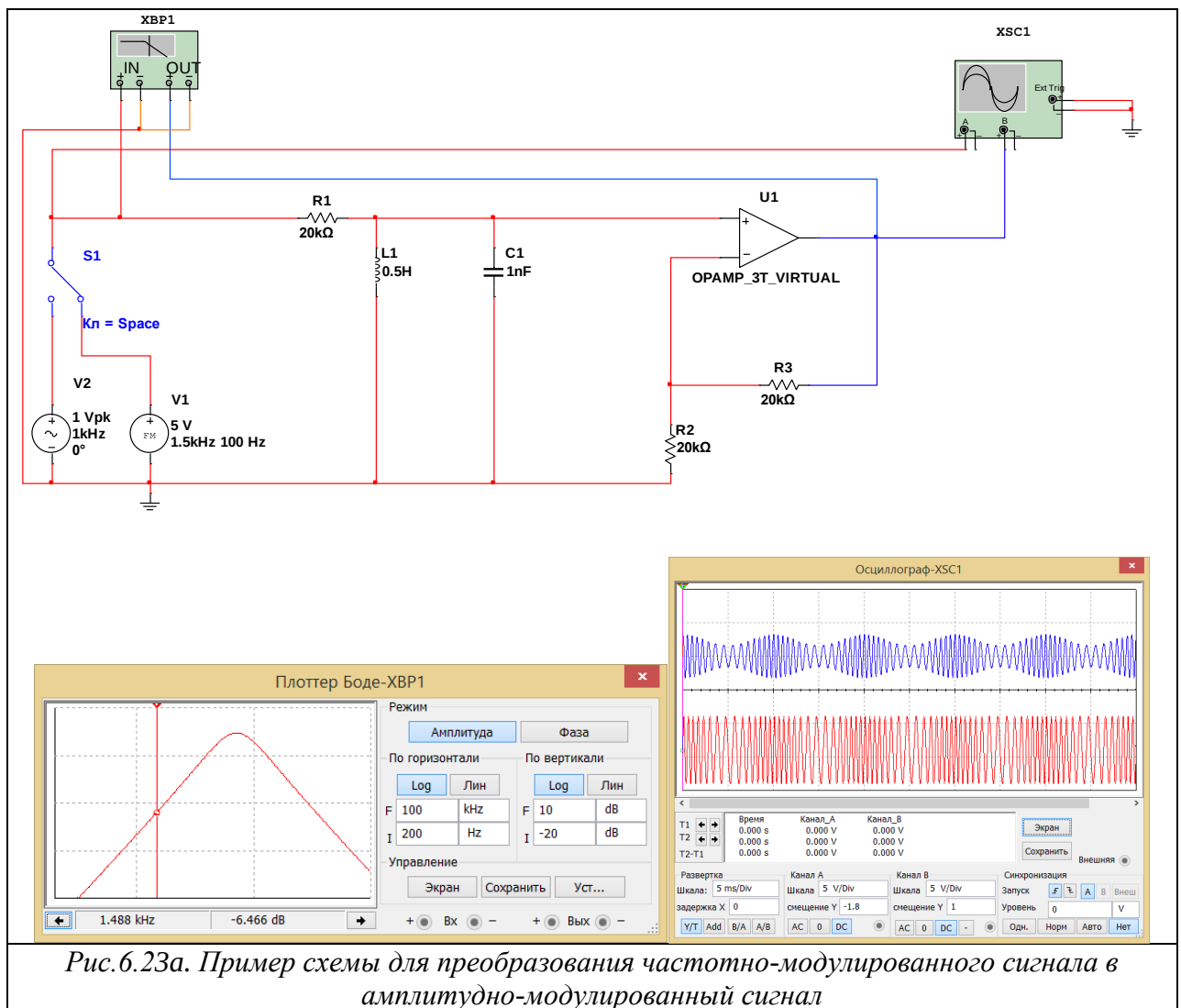


Рис.6.23а. Пример схемы для преобразования частотно-модулированного сигнала в амплитудно-модулированный сигнал

По аналогии с рассмотренным выше амплитудным синхронным детектированием рассмотрим нелинейное преобразование суммы ФМ-сигнала

$$u(t) = U \cos[\omega_0 t + \theta(t)]$$

и гармонического сигнала гетеродина

$$u_r(t) = U_r \cos(\omega_r t + \theta_r)$$

с помощью нелинейного элемента с ВАХ, описываемой уравнением

$$i = a_2 u^2.$$

Для тока на суммарной и разностной частотах в данном случае можно записать

$$i_{\phi}(t) = a_2 U U_{\Gamma} \cos[(\omega_0 + \omega_{\Gamma}) t + \theta(t) + \theta_{\Gamma}] + \\ + a_2 U U_{\Gamma} \cos[(\omega_0 - \omega_{\Gamma}) t + \theta(t) - \theta_{\Gamma}].$$

При выборе частоты колебаний гетеродина равной средней частоте колебаний исходного ФМ-сигнала $\omega_{\Gamma} = \omega_0$ получаем

$$i_{\phi}(t) = a_2 U U_{\Gamma} \cos(2\omega_0 t + \theta(t) + \theta_{\Gamma}) + a_2 U U_{\Gamma} \cos(\theta(t) - \theta_{\Gamma}), \quad (6.26)$$

что после низкочастотной фильтрации дает

$$i_{\text{нч}}(t) = a_2 U U_{\Gamma} \{ \cos(\theta(t)) \cos \theta_{\Gamma} + \sin(\theta(t)) \sin \theta_{\Gamma} \}. \quad (6.27)$$

Если принять $\theta_{\Gamma} = \pi/2$ и диапазон изменения $\theta(t)$ полагать таким, что справедливо представление $\sin \alpha \approx \alpha$, то

$$i_{\text{нч}}(t) = a_2 U U_{\Gamma} \theta(t). \quad (6.28)$$

В результате нам удалось получить весьма простое уравнение для преобразования сигнала, однако, следует помнить, что все проведенные преобразования (6.26)-(6.28) сохраняют свою справедливость только при малых значениях индекса модуляции.

Сложность реализации синхронного фазового детектирования аналогично синхронному амплитудному детектированию состоит в необходимости строгой синхронизации частоты гетеродина детектора и принимаемого сигнала

6.8. Воздействие случайных сигналов на нелинейный безынерционный элемент

Как и для линейных цепей общая задача изучения прохождения сигналов через нелинейные цепи состоит в нахождении статистических характеристик выходного сигнала по известным статистическим характеристикам входного сигнала и заданным параметрам цепи.

В общем случае входной сигнал может представлять собой стационарный или нестационарный случайный процесс, заданный либо своим законом распределения либо корреляционной функцией (или энергетическим спектром).

Во многих практических задачах приходится рассматривать сигналы, являющиеся суммой полезного сигнала и шумовой помехи

$$y(t) = x(t) + n(t), \quad (6.29)$$

где $y(t)$ - наблюдаемый сигнал,

$x(t)$ - полезный сигнал,

$n(t)$ - помеха.

При таком виде сочетания полезного сигнала и помехи ее называют аддитивной помехой.

Помимо аддитивного воздействия (6.29) на полезный сигнал помехи могут возникать из-за изменения параметров цепей или каналов связи, что приводит к более сложному виду формирования наблюдаемого сигнала

$$y(t) = x(t) + \mu(t)x(t) + n(t), \quad (6.30)$$

где $\mu(t)$ - мультипликативная помеха.

Уравнение для мультипликативной помехи (6.30) ясно показывает принципиальную возможность нелинейной связи исходного и наблюдаемого сигнала,

которая в свою очередь может быть усложнена использованием в процессе преобразования сигналов различных нелинейных элементов.

В задачи нашего курса не входит подробное изучение вопросов статистического описания и анализа нелинейных преобразований случайных сигналов. Мы кратко остановимся лишь на концептуальных вопросах воздействия стационарных случайных сигналов на нелинейные безынерционные цепи.

6.8.1. Преобразования плотности вероятности входного сигнала

Рассматривая ранее преобразование статистических характеристик случайного сигнала при прохождении через линейную цепь, нам удалось получить достаточно простые уравнения для связи корреляционных функций (или спектральных плотностей мощности) на входе и выходе цепи. В случаях нормального характера распределения плотности вероятности входного процесса, достаточно просто решается и задача нахождения соответствующих распределений для выходного сигнала.

В нелинейных безынерционных цепях получение плотности вероятности выходного сигнала представляет собой сравнительно простую задачу, в то время как вычисление корреляционной функции выходного процесса может оказаться весьма затруднительным.

Как и во всех разделах касающихся анализа нелинейных элементов мы будем рассматривать их как некоторые двухполюсники с ВАХ, заданной известной функцией

$$y = f(x),$$

где y – выходная величина, чаще всего представляющая ток в цепи,

x – входная величина, чаще всего являющаяся приложенным к зажимам двухполюсника напряжением.

Если плотность вероятности входного сигнала задана в виде $p(x)$, то исходя из заданной безынерционной формы связи входного и выходного сигнала, для плотности вероятности выходной величины можно записать

$$p(y)dy = p(x)dx.$$

Поскольку плотность вероятности по определению является неотрицательной величиной, то справедливо следующее соотношение

$$p(y) = p(x) \left| \frac{dx}{dy} \right|.$$

Если обратная к f функция f^{-1} однозначна, то легко получаем

$$p(y) = p(x) \frac{1}{|dy/dx|}. \quad (6.31)$$

Для неоднозначной функции f^{-1} каждую из ее отдельных ветвей требуется рассмотреть самостоятельно, а результаты сложить, то есть

$$p(y) = [p(x) 1/|dy/dx|]_{x1=f^{-1}(y)} + [p(x) 1/|dy/dx|]_{x2=f^{-1}(y)} + \dots, \quad (6.32)$$

где $x1 = f^{-1}(y), x2 = f^{-1}(y), \dots$ – отдельные однозначные ветви функциональной зависимости $f^{-1}(\cdot)$.

Если функция $y = f(x)$ имеет участок, на котором величина y является константой, то выражение для плотности вероятности дополняется дельта-функцией.

Процесс определения плотности вероятности выходного сигнала по уравнениям (6.31) и (6.32) продемонстрируем на нескольких примерах.

Пример.6.6.

Допустим, что входной сигнал является нормальным центрированным шумом с дисперсией σ^2 .

Характеристика нелинейного элемента представляет собой кусочно-линейную функцию

$$y = \begin{cases} Sx, & x > 0 \\ 0, & x < 0 \end{cases}.$$

Вид данной зависимости представлен на рис. 6.24.

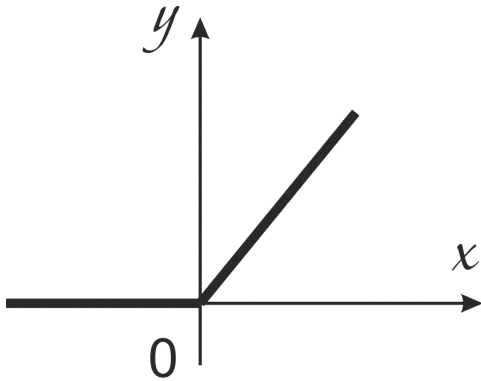


Рис.6.24. Выходная характеристика нелинейного элемента

Из анализа характеристики нелинейного элемента ясно, что на интервале $x > 0$

имеет место линейное преобразование исходного сигнала с коэффициентом усиления равным крутизне нелинейной характеристики S . Поэтому для данного интервала искомая плотность вероятности находится как

$$p(y) = \frac{1}{S\sigma\sqrt{2\pi}} e^{-\frac{y^2}{2S^2\sigma^2}}, y > 0.$$

Поскольку входной процесс является нормальным и центрированным, то его математическое ожидание равно нулю, а появление отрицательных значений x равновероятно появлению положительных значений, то должны быть равны и вероятности событий $y > 0$ и $y = 0$.

Вследствие этого

$$\int_0^{\infty} p(y) dy = 1/2 \quad \text{и} \quad \int_{-\infty}^{\infty} p(y) dy = 1.$$

Полученный выше закон плотности вероятности для $y > 0$ требуется дополнить дельта-функцией с коэффициентом $1/2$.

Окончательно для искомого распределения выходного сигнала получаем

$$p(y) = \frac{1}{2} \delta(y) + \frac{1}{S\sigma\sqrt{2\pi}} e^{-\frac{y^2}{2S^2\sigma^2}}, y \geq 0.$$

График полученного распределения приведен на рис.6.25.

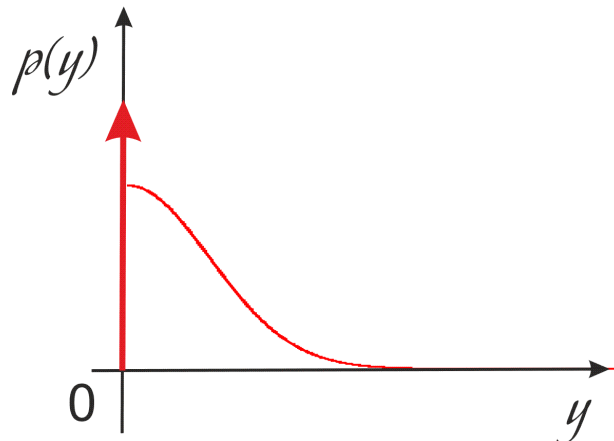


Рис.6.25. Плотность вероятности выходного сигнала

Пример.6.7.

Воспользуемся условиями примера 6.4., но в качестве нелинейной характеристики будем применять функцию вида

$$y = ax^2,$$

где $a = \text{const} > 0$.

График заданной нелинейной характеристики и обратной к ней функции приведен на рис.6.26.

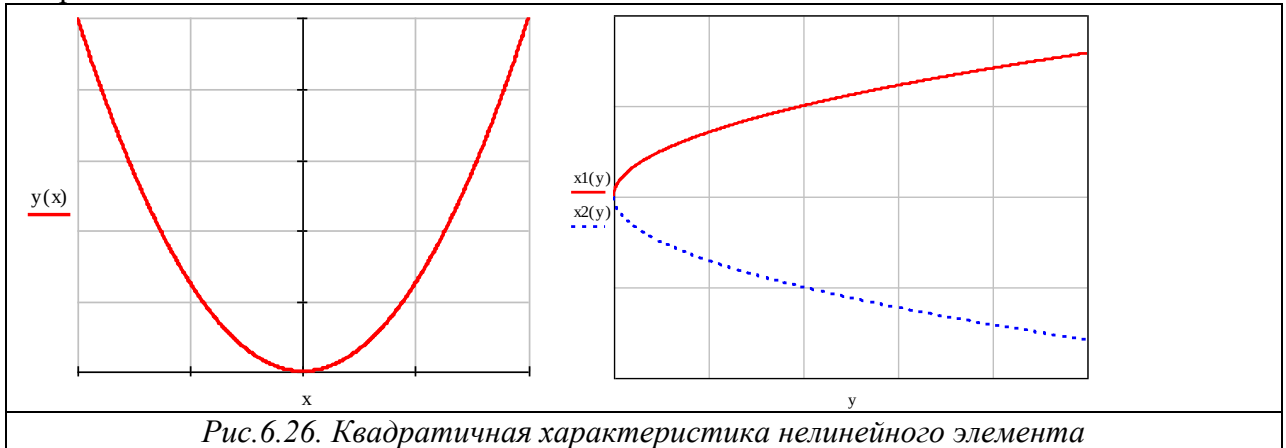


Рис.6.26. Квадратичная характеристика нелинейного элемента

На рис.6.26. наглядно показано, что функция, обратная к заданной характеристике нелинейного элемента является неоднозначной и содержит в себе две однозначные ветви.

Поэтому в соответствии с правилом указанным выше определяем две функции

$$x_{1,2} = \pm \sqrt{\frac{y}{a}},$$

что для модуля производной дает выражением в виде

$$\left| \frac{dx_{1,2}}{dy} \right| = \frac{1}{2\sqrt{ay}}.$$

Тогда для искомой плотности вероятности запишем

$$p_y(y) = p_x\left(\sqrt{\frac{y}{a}}\right) \frac{1}{2\sqrt{ay}} + p_x\left(-\sqrt{\frac{y}{a}}\right) \frac{1}{2\sqrt{ay}},$$

что в окончательном виде дает

$$p_y(y) = \begin{cases} \frac{1}{\sigma\sqrt{2\pi ay}} e^{-\frac{y}{2a\sigma^2}}, & y \geq 0 \\ 0, & y < 0 \end{cases}$$

Графики плотностей вероятности входного и выходного сигналов представлены на рис.6.27.

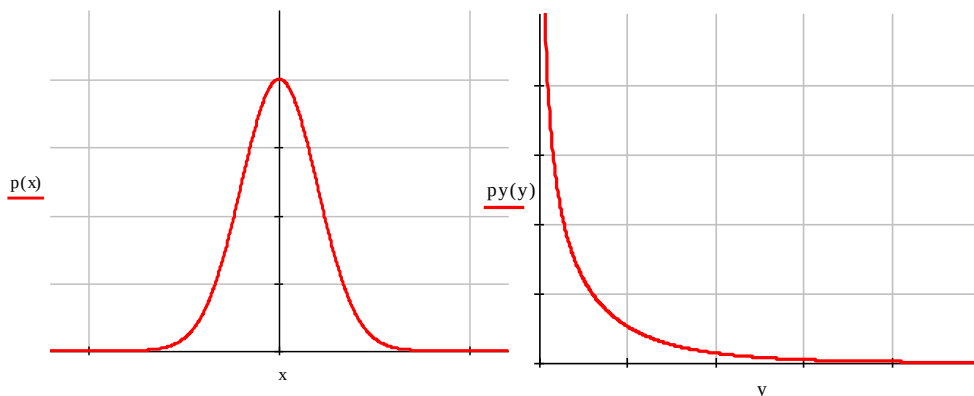


Рис.6.27. Плотности вероятности входного и выходного сигналов

6.8.2. Преобразование корреляционной функции случайного процесса при прохождении нелинейной безынерционной цепи

В силу безынерционности рассматриваемых нами характеристик нелинейных элементов, для ковариационной и корреляционной функций на выходе нелинейного элемента можно использовать те же определения, которые использовались при анализе линейных систем.

$$\begin{aligned} K_y(t_1, t_2) &= M\{y(t_1)y(t_2)\}, \\ R_y(t_1, t_2) &= M\{[y(t_1) - m_y(t_1)][y(t_2) - m_y(t_2)]\} = \\ &= K_y(t_1, t_2) - m_y(t_1)m_y(t_2). \end{aligned}$$

В общем случае корреляционная функция выходного процесса может быть найдена из уравнения

$$R_y(t_1, t_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y_1 y_2 p_y(y_1, y_2, t_1, t_2) dy_1 dy_2.$$

Поскольку предполагается, что выходная переменная нелинейного элемента связана с входной однозначной безынерционной зависимостью $y = f(x)$, то как было показано выше

$$p(y)dy = p(x)dx.$$

Таким образом,

$$R_y(t_1, t_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(y_1)f(y_2) p_y(y_1, y_2, t_1, t_2) dx_1 dx_2, \quad (6.33)$$

что для стационарного входного сигнала уравнение (6.33) можно записать в виде

$$R_y(\tau) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(y_1)f(y_2) p_y(y_1, y_2, \tau) dx_1 dx_2. \quad (6.34)$$

Для эргодических процессов усреднение по ансамблю может быть заменено усреднением по времени, что дает вторую формулу для нахождения корреляционной функции

$$R_y(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f(x(t))f(x(t + \tau))d\tau. \quad (6.35)$$

С учетом стационарности выражения (6.34), (6.35) могут быть представлены в более простом виде и с сокращением записи времени t

$$\begin{aligned} K_y(\tau) &= M\{y(t)y(t + \tau)\} = M\{yy_\tau\} = \overline{yy_\tau}, \\ R_y(\tau) &= \overline{yy_\tau} - \overline{y}^2. \end{aligned}$$

К сожалению, вычисление корреляционных функций по приведенным формулам представляет собой довольно трудную задачу.

Дело в том, что при анализе нелинейных систем не действует принцип суперпозиции, поэтому методы вычисления корреляционной функции в данном случае требуют использование двумерной плотности вероятности даже для стационарных

эргодических процессов и никаких упрощений в отличие от линейного случая произвести невозможно.

В качестве иллюстрации возникающих при этом осложнений приведем формулу для двумерной плотности вероятности нормального стационарного процесса

$$p(x, x_\tau, \tau) = \frac{1}{2\pi\sigma^2\sqrt{1-r_x^2}} e^{-\frac{x^2+x_\tau^2-2r_x x x_\tau}{2\sigma^2(1-r_x^2)}},$$

где $r_x = r_x(\tau) = \frac{R_x(\tau)}{\sigma^2}$ – коэффициент корреляции.

Поскольку аналитические выводы искомым корреляционных функций достаточно громоздки даже для простейших случаев, мы ограничимся приведением без доказательства уравнения корреляционной функции выходного сигнала для случая, когда нелинейная характеристика задана уравнением

$$y = ax^2.$$

При условии, что входным сигналом является нормальный центрированный стационарный процесс с дисперсией σ^2 и коэффициентом корреляции $r_x(\tau)$, корреляционная функция выходного сигнала имеет вид

$$R_y(\tau) = a^2\sigma^4 + 2a^2\sigma^4 r_x^2(\tau) = a^2\sigma^4 + 2a^2 R_x^2(\tau). \quad (6.36)$$

6.8.3. Спектральная функция случайного процесса на выходе нелинейного безынерционного звена

Прямое получение спектральной плотности мощности случайного процесса на выходе нелинейной цепи по заданному спектру входного сигнала, к сожалению, не представляется возможным. Поэтому для решения данной задачи производится определение корреляционной функции выходного сигнала с последующим применением к ней закона Винера-Хинчина. При таком подходе спектральная плотность мощности выходного сигнала определяется как преобразование Фурье от корреляционной функции выходного процесса

$$S_{P_x}(\omega) = \int_{-\infty}^{\infty} R_x(\tau) e^{-j\omega\tau} d\tau = 2 \int_0^{\infty} R_x(\tau) \cos(\omega\tau) d\tau. \quad (6.37)$$

Рассмотрим процесс определения спектральной плотности мощности (6.37) выходного сигнала нелинейной цепи на примере нелинейности, заданной выражением

$$y = ax^2$$

и входного нормального узкополосного случайного процесса с корреляционной функцией

$$R_x(\tau) = \sigma^2 \rho_x(\tau) \cos(\omega_0 \tau),$$

где ω_0 – центральная частота входного процесса,

$\rho_x(\tau)$ – нормированная огибающая корреляционной функции входного процесса.

С учетом сделанных допущений, используя полученную выше формулу (6.36) для вычисления корреляционной функции выходного процесса, можно записать

$$R_y(\tau) = a^2\sigma^4 + 2a^2 R_x^2(\tau) = a^2\sigma^4 + 2a^2 [\sigma^2 \rho_x(\tau) \cos(\omega_0 \tau)]^2 = \quad (6.38)$$

$$= a^2\sigma^4 + 2a^2\sigma^4 \rho_x^2(\tau) + a^2\sigma^4 \rho_x^2(\tau) \cos(2\omega_0 \tau).$$

Рассмотрим полученный результат на качественном уровне без окончательного количественного вычисления соответствующих выражений. Применение уравнению (6.38) и, состоящему из трех слагаемых, преобразования Фурье показывает, что спектральная плотность мощности выходного сигнала будет иметь три составляющие.

$$S_{Py}(\omega) = S_{Py0}(\omega) + S_{PyHЧ}(\omega) + S_{PyBЧ}(\omega)$$

Первая компонента соответствует нулевой составляющей и является дельта функцией, расположенной в точке $\omega = 0$. Вторая компонента – низкочастотная составляющая, третья компонента – высокочастотная составляющая, сконцентрированная в окрестностях частоты $\omega = \omega_0$. Схематично данный результат представлен на рис.6.28.

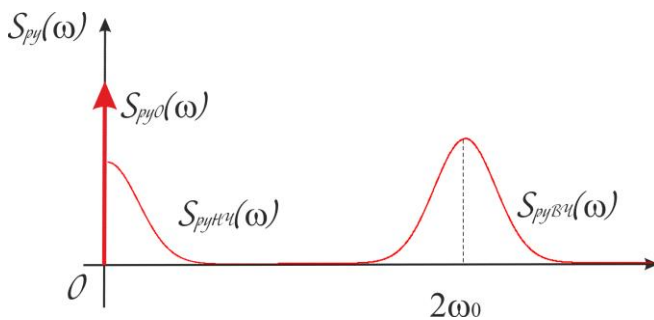


Рис.6.28. Спектральная плотность мощности на выходе нелинейного элемента с квадратичной характеристикой

В рассмотренном случае прохождения через нелинейную цепь случайного сигнала просматривается прямая аналогия с эффектом расширения спектра, который был рассмотрен при изучении процессов воздействия детерминированного гармонического сигнала на нелинейную цепь.

6.9. Общие сведения о параметрических цепях

До сих пор мы рассматривали радиотехнические цепи, параметры элементов которых оставались постоянными. Если в цепи присутствует хотя бы один элемент, параметры которого меняются в соответствии с некоторым заданным правилом, то такие цепи называются *параметрическими*.

Параметрические цепи могут быть линейными и нелинейными. В линейно-параметрической цепи параметры элементов не зависят от уровня сигнала, но могут изменяться во времени. Реально параметрический элемент получают из нелинейного элемента (резистивного, емкостного или индуктивного), на вход которого подается сумма двух сигналов, один сигнал – информационный, несущий информацию и имеющий, как правило, малую амплитуду. Поэтому в области изменения информационного сигнала параметры цепи практически не изменяются. Вторым сигналом является управляющий сигнал большой амплитуды, который изменяет положение рабочей точки нелинейного элемента, а, следовательно, параметр цепи.

Для резистивного элемента таким управляемым параметром является дифференциальная крутизна. График, поясняющий процесс формирования изменяющегося параметра приведен на рис.6.29.

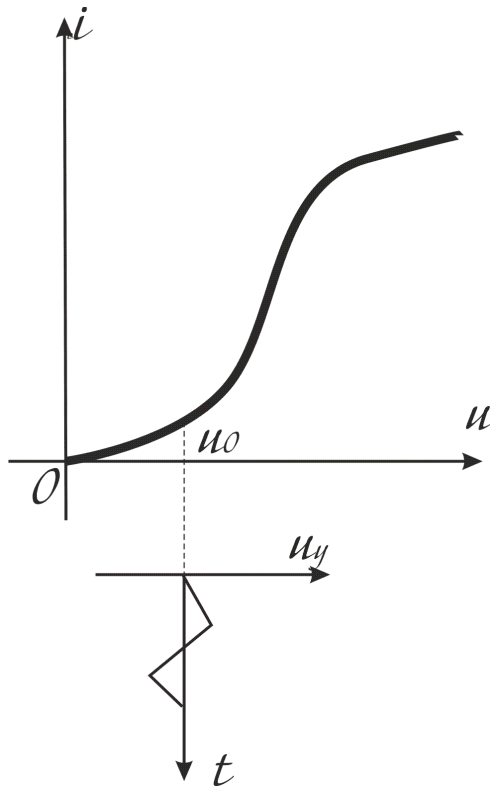


Рис.6.29. Управление крутизной характеристики

Дифференциальная крутизна характеристики в этом случае будет определяться как

$$S(t) = S(u_y(t)) = \left(\frac{di}{du}\right)_{u=u_0+u(t)}. \quad (6.39)$$

Если управляющий сигнал $u(t)$ в (6.39) достаточно мал, то в пределах сдвига рабочей точки сохраняется линейность цепи и можно говорить о формировании параметрического резистора $R(t)$, связывающего мгновенные значения тока и напряжения

$$u(t) = R(t)i(t).$$

На практике значительно чаще входным сигналом является не ток, а напряжение, поэтому вместо параметрического сопротивления можно ввести параметрическую проводимость

$$i(t) = g(t)u(t).$$

Существенная особенность дифференциальной крутизны $S(t)$, а, следовательно, и дифференциального

сопротивления, состоит в том, что они могут принимать отрицательные значения.

Подобным образом могут быть определены параметрические емкость и индуктивность с той лишь разницей, что дифференциальная емкость и индуктивность всегда остаются положительными. (Чтобы сделать эти величинами отрицательными требуется применение более сложных составных схем с использованием обратной связи).

Говоря о параметрических изменениях в цепи, следует различать два варианта таких изменений:

- управляемое изменение с целью придания цепи желаемых свойств,
- неуправляемые изменения, происходящие из-за различного рода физических процессов, воздействующих на цепь или канал связи.

Для описания динамических свойств линейных цепей с постоянными параметрами мы использовали передаточную функцию $W(s)$, частотную передаточную функцию (комплексный коэффициент передачи) $W(j\omega)$ (или $K(j\omega)$) и импульсную характеристику $g(t)$, которая связана с передаточной функцией преобразованием Лапласа.

В параметрических цепях характер прохождения сигнала существенно меняется, поскольку все указанные характеристики теперь дополнительно изменяются во времени из-за непостоянства параметров цепи.

Поэтому для цепей с переменными параметрами импульсную характеристику вводят в виде $g(t, \tau)$ или $g(t, x)$ и определяют как величину отклика цепи в момент времени t на воздействия импульса в момент времени $x = t - \tau$, что соответствует входному воздействию $\delta(t - x)$.

Тогда по аналогии с линейными системами для связи произвольного входного сигнала $u(t)$ с соответствующим ему выходным сигналом $u_{\text{вых}}(t)$ можно записать

$$u_{\text{вых}}(t) = \int_{-\infty}^{\infty} u(t)g(t, x)dx = \int_{-\infty}^{\infty} u(t - \tau)g_1(t, \tau)d\tau, \quad (6.40)$$

где $g_1(t, \tau) = g(t, t - \tau)$. При этом условие физической реализуемости цепи требует, чтобы $g_1(t, \tau) = 0$ при $\tau > t$.

Таким же образом можно для параметрической цепи ввести передаточную функцию в виде

$$W(s, t) = \int_{-\infty}^{\infty} g_1(t, \tau)e^{-j\omega\tau}d\tau \quad (6.41)$$

и соответственно получить частотную передаточную функцию (комплексный коэффициент передачи)

$$W(j\omega, t) = K(j\omega, t) = W(s, t)_{s=j\omega}. \quad (6.42)$$

Зависимость передаточной функции параметрической цепи от времени приводит к тому, что при прохождении гармонического сигнала через такую цепь в выходном сигнале обнаруживаются новые дополнительные частоты.

Для демонстрации этого эффекта предположим, что передаточная функция, соответствующая уравнениям (6.40) - (6.42) некоторой параметрической цепи изменяется по закону

$$W(j\omega, t) = W_0(j\omega) + W_1(j\omega) \cos(\Omega t + \theta_1),$$

а на вход данной цепи подается гармонический сигнал вида

$$u(t) = U \cos(\omega_0 t + \theta_0).$$

Переходя к комплексному сигналу и используя принцип линейности запишем

$$u_{\text{вых}}^{\circ}(t) = W(j\omega, t)e^{j(\omega_0 t + \theta_0)} = W_0(j\omega)e^{j(\omega_0 t + \theta_0)} + W_1(j\omega)e^{j(\omega_0 t + \theta_0)} \cos(\Omega t + \theta_1).$$

Далее

$$u_{\text{вых}}^{\circ}(t) = W_0(\omega)e^{j(\omega_0 t + \varphi_0 + \theta_0)} + W_1(\omega)e^{j(\omega_0 t + \varphi_1 + \theta_0)} \cos(\Omega t + \theta_1),$$

где φ_0, φ_1 – фазовые сдвиги, внесенные передаточными функциями $W_0(j\omega)$ и $W_1(j\omega)$ соответственно.

Тогда выходной сигнал можно представить как действительную часть комплексного сигнала

$$u_{\text{вых}}(t) = \text{Re}(u_{\text{вых}}^{\circ}(t)) = W_0(\omega_0)U \cos(\omega_0 t + \theta_0 + \varphi_0) + \\ + W_1(\omega_0)U \cos(\omega_0 t + \theta_0 + \varphi_1) \cos(\Omega t + \theta_1).$$

Откуда окончательно получаем

$$u_{\text{вых}}(t) = W_0(\omega_0)U \cos(\omega_0 t + \theta_0 + \varphi_0) + \\ + \frac{1}{2}W_1(\omega_0)U \cos((\omega_0 + \Omega)t + \theta_0 + \theta_1 + \varphi_1) + \\ + \frac{1}{2}W_1(\omega_0)U \cos((\omega_0 - \Omega)t + \theta_0 - \theta_1 + \varphi_1). \quad (6.43)$$

Таким образом, из уравнения (6.43) видно, что при прохождении сигнала через параметрические цепи, так же как и при прохождении через нелинейные цепи в спектре выходного сигнала появляются дополнительные составляющие на комбинационных частотах. В рассмотренном примере такими частотами были $\omega_0 + \Omega$ и $\omega_0 - \Omega$. При усложнении вида функции, определяющей временные изменения параметров цепи, изменения спектра выходного сигнала могут быть более значительными.

Обнаруженные свойства параметрических цепей позволяют по-иному взглянуть на рассмотренный ранее процесс получения АМ-сигнала. Если процесс модуляции

рассмотреть как прохождение несущего колебания через параметрическую цепь с законом изменения параметров, определяемым информационным сигналом, то становится ясно, использованное резонансного усилителя и нелинейного элемента в данном процессе является не его сутью, а только способом технической реализации требуемых параметрических изменений модулирующей цепи. Поэтому АМ модуляцию следует рассматривать как линейный параметрический процесс.

6.10. Вопросы для самоконтроля и повторения

1. Сформулируйте принцип линейной суперпозиции.
2. Продемонстрируйте нарушение принципа линейной суперпозиции в нелинейной системе.
3. Приведите пример нелинейной вольт-амперной характеристики (ВАХ) цепи.
4. Какие методы аппроксимации нелинейной ВАХ вам известны.
5. Продемонстрируйте результат прохождения гармонического сигнала через цепь с нелинейной ВАХ.
6. Как проводится кусочно-линейная аппроксимация ВАХ.
7. Как проводится полиномиальная аппроксимация ВАХ.
8. Что такое коэффициенты Берга?
9. Какие изменения происходят в спектре моногармонического сигнала при прохождении через цепь с нелинейной ВАХ?
10. Какие изменения происходят в спектре полигармонического сигнала при прохождении через цепь с нелинейной ВАХ?
11. Поясните принцип гетеродирования.
12. Поясните принцип работы нелинейного резонансного усилителя.
13. Каким образом может быть проведено умножение частоты сигнала?
14. Приведите схему простейшего амплитудного модулятора.
15. Какие схемы реализации амплитудного детектирования вам известны?
16. Каким образом реализуется синхронное детектирование?
17. Поясните принцип проведения частотного и фазового детектирования сигналов.
18. Какие явления наблюдаются при воздействии случайных сигналов на безынерционный нелинейный элемент?
19. Какие цепи называют параметрическими?
20. Для каких целей могут использоваться параметрические цепи?

ГЛАВА 7. Генерирование гармонических колебаний

7.1. Понятие автоколебательной системы

В главе 5. при рассмотрении вариантов формирования составных четырехполюсников был показан способ соединения цепей, получивший название обратной связи.

Обобщенная схема такого типа соединения представлена на рис. 7.1.

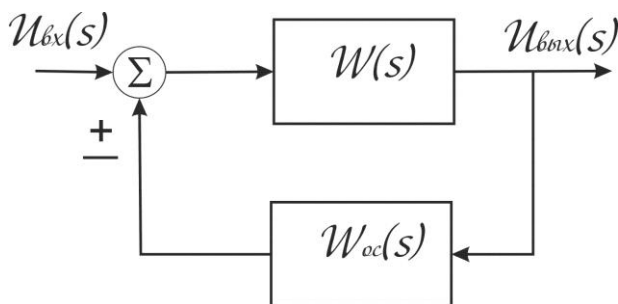


Рис.7.1. Общая схема соединения цепей по принципу обратной связи

В зависимости от знака, с которым сигнал обратной связи подается на сумматор, передаточная функция составной цепи принимает вид

$$W_{\Sigma}(s) = \frac{W(s)}{1 \mp W(s)W_{oc}(s)} \quad (7.1)$$

или для частотной передаточной функции

$$W_{\Sigma}(j\omega) = \frac{W(j\omega)}{1 \mp W(j\omega)W_{oc}(j\omega)} \quad (7.2)$$

Знак "+" в уравнениях (7.1), (7.2) соответствует случаю отрицательной обратной связи, а знак "-" - положительной.

Представив частотные передаточные функции (7.2) в виде

$$W(j\omega) = |W(j\omega)|e^{j\varphi(\omega)}$$

и

$$W_{oc}(j\omega) = |W_{oc}(j\omega)|e^{j\varphi_{oc}(\omega)}$$

для случая положительной обратной связи получаем

$$W_{\Sigma}(j\omega) = \frac{|W(j\omega)|e^{j\varphi(\omega)}}{1 - |W(j\omega)||W_{oc}(j\omega)|e^{j(\varphi(\omega)+\varphi_{oc}(\omega))}} \quad (7.3)$$

Внимательный анализ полученного выражения (7.3) показывает, что в зависимости от соотношения сдвигов фаз, обусловленных прямой и обратной ветвью прохождения сигнала, в знаменателе может происходить как вычитание, так и суммирование слагаемых.

При обеспечении синфазности сигналов прямой и обратной связи, например, при $(\varphi(\omega) + \varphi_{oc}(\omega)) = 2\pi$ в знаменателе частотной передаточной функции имеем

$$1 - |W(j\omega)||W_{oc}(j\omega)| \rightarrow 0 \text{ при } |W(j\omega)||W_{oc}(j\omega)| \rightarrow 1,$$

что в свою очередь ведет к бесконечному возрастанию модуля передаточной функции составной цепи и выведению ее на границу устойчивости.

В случае, когда в цепи, охваченной обратной связью, присутствуют активные элементы, способные передавать часть энергии от внешних источников в цепь, существует потенциальная возможность возникновения незатухающих собственных

колебаний. Эти колебания могут происходить в цепи без участия входного сигнала, то есть при условии $u_{\text{вх}}(t) = 0$. Таким образом, приходим к обобщенной схеме автоколебательной системы, представленной на рис. 7.2.

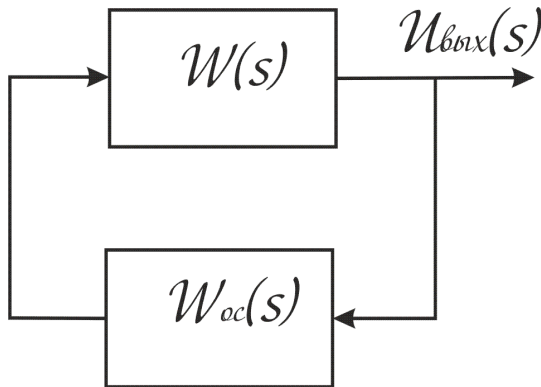


Рис. 7.2. Схема автоколебательной системы

Устройства, генерирующие автоколебания, часто называют автогенераторами. Принцип работы каждого автогенератора основан на передаче энергии от источника питания через управляемый элемент в колебательную систему. После возникновения колебаний их амплитуда начинает возрастать, поэтому, чтобы обеспечить установившийся колебательный режим этот рост амплитуды необходимо каким-либо образом ограничить. Роль такого ограничителя чаще всего выполняет нелинейная характеристика используемого активного элемента, благодаря которой после достижения определенного уровня рост амплитуды колебаний практически прекращается и остается постоянным. При достижении такого уровня режим работы автогенератора называют **стационарным**.

Из выше изложенного понятно, что при анализе автоколебательных систем возникают два основных практических вопроса:

- при каких условиях использование обратной связи приводит к самовозбуждению системы;
- каковыми будут частота и амплитуда автоколебаний при достижении стационарного режима.

Рассмотрим эти вопросы более подробно.

7.2. Баланс фаз и баланс амплитуд. Возникновение колебаний в автогенераторе

Предположим, что автоколебательная система прошла режим возбуждения и находится в стационарном режиме. Тогда из приведенных ранее уравнений для передаточной функции автоколебательной системы следует

$$U(s) = W_{\text{oc}}(s)U_{\text{вых}}(s),$$

где $U(s)$ – изображение Лапласа для выходного сигнала цепи обратной связи, который одновременно является входным сигналом для цепи прямой связи автоколебательной системы.

Следовательно,

$$W(s)U(s) = U_{\text{вых}}(s).$$

Объединяя два последних уравнения, получаем искомое условие для стационарного режима работы

$$W(s)W_{\text{oc}}(s) = 1 \quad (7.4)$$

или для частотного варианта записи

$$W(j\omega)W_{\text{oc}}(j\omega) = 1. \quad (7.5)$$

Переходя от полученного комплексного результата в (7.4), (7.5) к вещественной форме условий существования стационарного режима, получаем два соответствующих уравнения

$$|W(j\omega)||W_{oc}(j\omega)| = 1 \quad (7.6)$$

и

$$\varphi(\omega) + \varphi_{oc}(\omega) = 2\pi k, \quad k = 0, \pm 1, \pm 2, \dots \quad (7.7)$$

Первое из этих уравнений (7.6) называют уравнением баланса амплитуд или просто **балансом амплитуд**, а второе (7.7) – **балансом фаз**.

Если условия баланса амплитуд и баланса фаз выполняются для единственной частоты ω_0 , то колебания будут иметь гармонический характер, если же они справедливы для полосы частот, то колебания будут иметь релаксационный характер.

Рассмотрим процесс возникновения автоколебаний более подробно на примере схемы с LC-автогенератором, приведенной на рис.7.3.

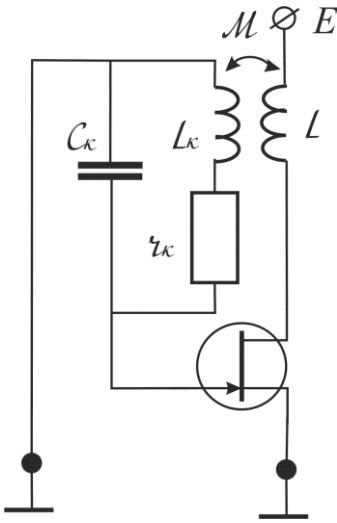


Рис.7.3. Упрощенная схема LC-автогенератора

Предположим, что ВАХ в окрестности рабочей точки может быть аппроксимирована уравнением

$$i = Su.$$

Тогда для колебательного контура с учетом того, что напряжение на конденсаторе фактически равно напряжению на затворе транзистора, можно записать следующее

$$\left(L_k C_k \frac{d^2}{dt^2} + r_k C_k \frac{d}{dt} + 1 \right) u \pm M \frac{di}{dt} = 0,$$

где u – напряжение на затворе,

i – ток стока.

Подставляя в данное уравнение формулу ВАХ транзистора, и производя подстановку $\omega_0 = \frac{1}{\sqrt{L_k C_k}}$ где ω_0 – частота собственных колебаний LC без потерь,

получим

$$\left[\frac{d^2}{dt^2} + \left(\frac{r_k}{L_k} \pm \frac{MS}{L_k C_k} \right) \frac{d}{dt} + \omega_0^2 \right] u = 0. \quad (7.8)$$

Введем в уравнении (7.8) обозначение

$$2\alpha = \left(\frac{r_k}{L_k} \pm \frac{MS}{L_k C_k} \right)$$

и упростим запись полученного дифференциального уравнения

$$\left[\frac{d^2}{dt^2} + 2\alpha \frac{d}{dt} + \omega_0^2 \right] u = 0.$$

Этому дифференциальному уравнению соответствует характеристическое уравнение вида

$$p^2 + 2\alpha p + \omega_0^2 = 0.$$

Корни полученного характеристического уравнения суть

$$p_{1,2} = -\alpha \pm j\sqrt{\omega_0^2 - \alpha^2}.$$

Что приводит к решению вида

$$u(t) = Ae^{-\alpha t} \cos(\sqrt{\omega_0^2 - \alpha^2}t) + Be^{-\alpha t} \sin(\sqrt{\omega_0^2 - \alpha^2}t),$$

где $A, B = \text{const}$, зависящие от начальных условий.

В схемах генераторов применяются реактивные элементы с параметрами, способными обеспечить высокую добротность

$$Q \gg 1,$$

где $Q = \frac{\omega_0 L}{r} = \frac{1}{r} \sqrt{\frac{L}{C}}$ – добротность параллельного колебательного контура.

Поэтому

$$\omega_0 \gg \alpha$$

и для корней характеристического уравнения справедливо

$$p_{1,2} \approx -\alpha \pm j\omega_0.$$

Следовательно, частота генерируемых колебаний в стационарном режиме работы автогенератора будет близка к собственной частоте колебаний контура.

Для пассивной цепи (то есть при $S = 0$)

$$2\alpha = \left(\frac{r_k}{L_k}\right),$$

что определяет затухающий характер колебаний в контуре.

Наличие в рассмотренной схеме активного усилительного элемента с крутизной характеристики $S \neq 0$ позволяет обеспечить условие самовозбуждения автогенератора $\alpha < 0$, или иначе

$$\frac{r_k}{L_k} \pm \frac{MS}{L_k C_k} < 0, \quad (7.9)$$

что возможно получить только при отрицательном знаке перед вторым слагаемым левой части записанного неравенства. Требуемый знак обеспечивается выбором способа включения катушки контура и катушки связи схемы автогенератора, которые для этого должны быть включены встречно. В этом случае

$$r_k < \frac{MS}{C_k}. \quad (7.10)$$

Последние два неравенства (7.9), (7.10) помогают предложить новую трактовку роли активного усилительного элемента в автогенераторной цепи, а именно, он по своей сути привносит в цепь отрицательное активное сопротивление, компенсирующее собственные потери контура на внутреннем сопротивлении r_k . Если компенсация превышает внутренние потери, то возможно возникновение самовозбуждения генератора.

Пример 7.1.

Продemonстрируем работу автогенератора с трансформаторной связью, схема которого приведена на рис.7.4. На этом же рисунке приведена осциллограмма, показывающая процесс возбуждения автогенератора и нарастания амплитуды колебаний до установившегося стационарного значения.

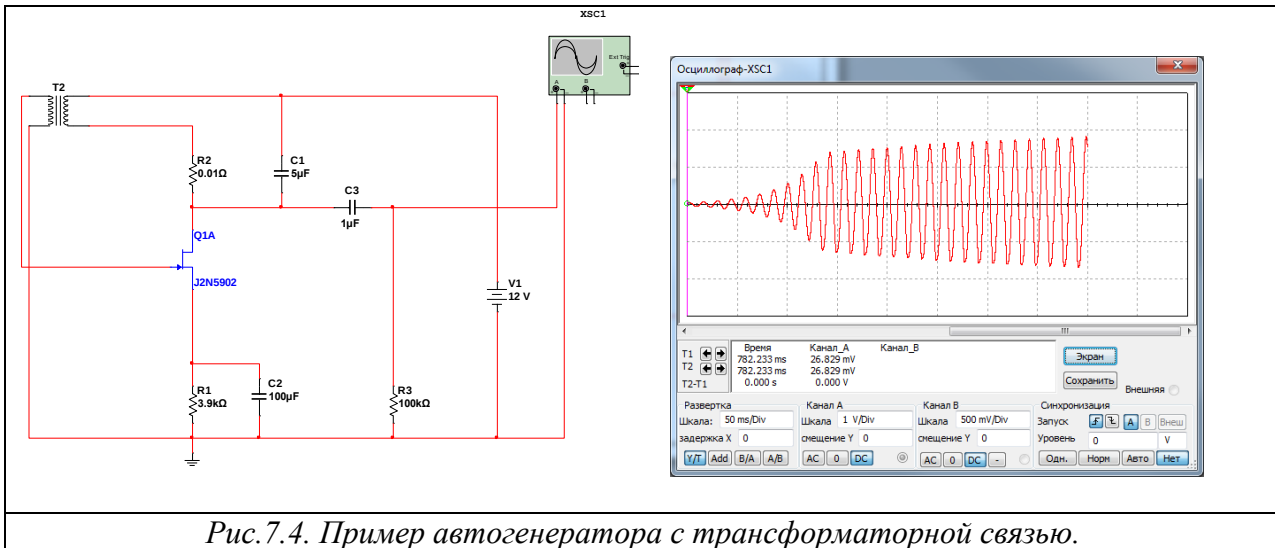


Рис.7.4. Пример автогенератора с трансформаторной связью.

7.3. Стационарный режим работы автогенератора

Анализ условий самовозбуждения автогенератора в предыдущем параграфе был произведен исходя из предположения линейности ВАХ усилительного элемента при малых токах и напряжениях, что позволяло использовать уравнения

$$i = Su \quad \text{и} \quad \frac{d}{dt}i = S \frac{d}{dt}u.$$

Теперь для случая больших отклонений от начала координат будем предполагать нелинейную зависимость тока от прилагаемого управляющего напряжения, то есть

$$i = f(u).$$

Дифференциальная крутизна, соответствующая этой зависимости, будет определяться как

$$S_d = \frac{df}{du},$$

а уравнение напряжения в контуре примет соответствующий нелинейный вид

$$\frac{d^2u}{dt^2} + \left(\frac{r_k}{L_k} - \frac{M}{L_k C_k} \frac{df}{du} \right) \frac{du}{dt} + \omega_0^2 u = 0. \quad (7.11)$$

В результате получили нелинейное однородное дифференциальное уравнение второго порядка. К сожалению, общие методы решения такого типа уравнений неизвестны. В математической физике решения таких уравнений ищутся исходя из физических предпосылок, базирующихся на особенностях функционирования исследуемого объекта.

Один из подходов приближенного анализа полученного уравнения (7.11) базируется на предположении о том, что напряжение $u(t)$, выделяемое на высокодобротном контуре автогенератора, может быть описано функцией времени

$$u(t) = U(t) \cos \omega_0 t,$$

где $U(t)$ – медленно изменяющаяся амплитуда колебаний автогенератора,

$\omega_0 \approx \omega_r$ – собственная частота колебаний контура, близкая к частоте колебаний генератора ω_r в силу предположения о том, что $\omega_0 \gg \alpha$.

Для первой и второй производных по времени от напряжения $u(t)$ нетрудно записать

$$\frac{du}{dt} = \frac{d}{dt}(U(t) \cos \omega_0 t) = \frac{dU(t)}{dt} \cos \omega_0 t - \omega_0 U(t) \sin \omega_0 t,$$

$$\begin{aligned}\frac{d^2u}{dt^2} &= \frac{d}{dt} \left[\frac{dU(t)}{dt} \cos \omega_0 t - \omega_0 U(t) \sin \omega_0 t \right] = \\ &= \frac{d^2U(t)}{dt^2} \cos \omega_0 t - 2 \frac{dU(t)}{dt} \omega_0 \sin \omega_0 t - \omega_0^2 U(t) \cos \omega_0 t.\end{aligned}$$

Сделанное допущение о медленном (по сравнению с $\cos \omega_0 t$) изменении функции $U(t)$, можно записать следующие приближенные уравнения

$$\begin{aligned}\frac{du}{dt} &\approx -\omega_0 U(t) \sin \omega_0 t, \\ \frac{d^2u}{dt^2} &\approx -2 \frac{dU(t)}{dt} \omega_0 \sin \omega_0 t - \omega_0^2 U(t) \cos \omega_0 t.\end{aligned}$$

В результате получаем упрощенное уравнение вида

$$\frac{dU(t)}{dt} + \frac{1}{2} \left(\frac{r_k}{L_k} - \frac{M}{L_k C_k} \frac{df}{du} \right) U(t) = 0. \quad (7.12)$$

Полученное уравнение (7.12) называют укороченным дифференциальным уравнением LC-автогенератора с высокочастотным контуром.

Дальнейшее упрощение полученного уравнения можно достигнуть, если для стационарного режима колебаний для функции

$$i(t) = f(u(t))$$

применить разложение в ряд с ограниченным числом членов. Например, ограничившись анализом постоянной составляющей и первой гармоники, данная аппроксимация примет вид

$$i(t) \approx I_0 + I_1 \cos \omega_0 t.$$

Тогда производная $\frac{df}{du}$, входящая в состав укороченного уравнения автогенератора может быть вычислена как

$$\frac{df}{du} = \frac{\frac{df}{dt}}{\frac{du}{dt}} = \frac{-\omega_0 I_1 \sin \omega_0 t}{-\omega_0 U(t) \sin \omega_0 t} = \frac{I_1}{U(t)}$$

При фиксированном значении амплитуды $U(t) = U = \text{const}$ можно ввести дифференциальную крутизну характеристики, связывающей амплитуду первой гармоники тока стока и амплитуду напряжения на затворе. Эта величина получила название средней крутизны ВАХ (или крутизны по первой гармонике)

$$S_1(U) = \frac{I_1(U)}{U}. \quad (7.13)$$

С учетом введения средней крутизны ВАХ (7.13) укороченное уравнение (7.12) принимает вид

$$\frac{dU(t)}{dt} + \frac{1}{2} \left(\frac{r_k}{L_k} - \frac{M}{L_k C_k} S_1(U) \right) U(t) = 0.$$

В стационарном режиме работы автогенератора амплитуда вырабатываемых колебаний должна оставаться постоянной, поэтому

$$\frac{dU(t)}{dt} = 0.$$

Откуда

$$\frac{r_k}{L_k} - \frac{M}{L_k C_k} S_1(U) = 0 \quad (7.14)$$

и средняя крутизна характеристики в рабочей точке (7.14) определится как

$$S_1(U_{\text{ст}}) = \frac{r_k C_k}{M}. \quad (7.15)$$

Наличие известной функции, описывающей ВАХ $i = f(u)$, позволяет легко определить положение рабочей точки (7.15) стационарного режима $U_{\text{ст}}$.

7.4. Мягкий и жесткий режимы самовозбуждения автогенератора

В зависимости от вида ВАХ и расположения на ней рабочей точки автогенератора средняя крутизна $S_1(U)$, как функция амплитуды напряжения вырабатываемых колебаний, может иметь одну из двух принципиально различающихся форм. Графики соответствующих вариантов характеристики приведены на рис.7.5

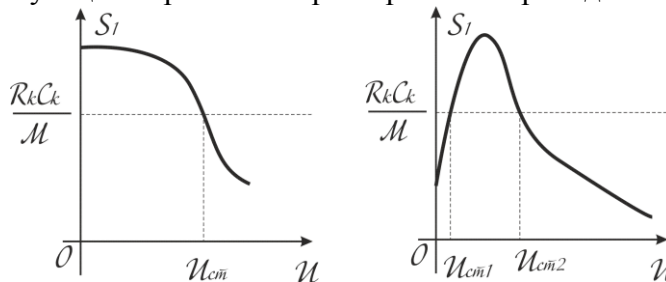


Рис.7.5. Графики средней крутизны $S_1(U)$ для мягкого и жесткого режимов самовозбуждения

Если средняя крутизна имеет монотонно убывающий характер при увеличении управляющего напряжения, то такой режим самовозбуждения называют мягким. Амплитуда колебаний стационарного режима в данном случае легко находится из графика, как точка его пресечения

горизонтальной линией, проведенной на уровне $\frac{r_k C_k}{M}$, определенном в (7.15) (Данная прямая часто называется прямой обратной связи.) При монотонном характере изменения крутизны точка пересечения может быть только единственной и никаких трудностей в данном случае не возникает.

Если средняя крутизна не является монотонной функцией, то точек ее пересечения с прямой обратной связи может быть две и соответственно возможны два стационарных режима работы $U_{\text{ст1}}$ и $U_{\text{ст2}}$.

С практической точки зрения весьма важным является вопрос о том, насколько устойчивым будет достигнутый в автогенераторе стационарный режим колебаний. Под устойчивостью режима в данном случае понимают стремление автоколебательной системы вернуться к колебаниям со стационарной амплитудой, если малые внешние воздействия отклонили систему от состояния стационарных колебаний. В отличие от устойчивой точки стационарных колебаний, в которой наблюдается тенденция возвращения в ту же самую точку, попадание системы в неустойчивую точку стационарных колебаний будет сопровождаться не возвращением, а переходом в иную устойчивую стационарную точку.

Для анализа данных процессов при малых отклонениях от рабочей точки можно воспользоваться укороченным уравнением автогенератора.

Допустим, что возникло некоторое отклонение от стационарной рабочей точки

$$U = U_{\text{ст}} + \Delta U.$$

Понятно, что при этом произойдет изменение средней крутизны, которое в рамках линейного приближения может быть оценено как

$$S_1(U) \approx S_1(U_{\text{ст}}) + \frac{dS_1(U)}{dU} \Delta U = S_1(U_{\text{ст}}) + A \Delta U,$$

где $A = \frac{dS_1(U)}{dU}$ – коэффициент наклона средней крутизны в стационарной точке.

Тогда с учетом ранее полученных уравнений (7.12)- (7.15) нетрудно записать

$$\frac{d\Delta U}{dt} + \frac{1}{2} \left(\frac{r_k}{L_k} - \frac{M}{L_k C_k} (S_1(U_{ст}) + A\Delta U) \right) (U_{ст} + \Delta U) = 0$$

Откуда для производной от отклонения можно получить

$$\frac{d\Delta U}{dt} \approx \frac{AMU_{ст}}{2L_k C_k} \Delta U. \quad (7.16)$$

Из уравнения (7.16) понятно, что, так как все параметры автогенератора, кроме A , являются положительными величинами, то для обеспечения устойчивости стационарного режима величина A должна быть отрицательной. Только в этом случае производная будет иметь знак противоположный знаку возникшего отклонения ΔU , что с течением времени вернет автоколебательную систему в исходное стационарное состояние.

Таким образом, стационарный режим автоколебаний является устойчивым, если

$$A = \frac{dS_1(U)}{dU} < 0. \quad (7.17)$$

В противном случае стационарный режим будет неустойчивым.

Возвращаясь к графикам, соответствующим мягкому и жесткому режимам самовозбуждения и рассмотрим некоторые особенности их работы.

Из рис. 7.4 видно, что при форме кривой, соответствующей мягкому режиму самовозбуждения взаимное положение прямой обратной связи $\frac{r_k C_k}{M}$ может меняться в зависимости от величины коэффициента взаимной индукции M . При уменьшении M прямая поднимается вверх, а при его увеличении опускается вниз. Поэтому при малых значениях M (то есть при слабой обратной связи) возможна ситуация, при которой характеристика $S_1(U)$ вообще не имеет с прямой обратной связи точек пересечения. В этом случае возможно устойчивое состояние системы, но оно соответствует нулевой амплитуде колебаний, что не представляет интереса с практической точки зрения.

По мере увеличения M обратная связь между контуром и управляющим напряжением усиливается. Начиная с некоторого значения коэффициента взаимной индукции, называемого критическим, автогенератор будет самовозбуждаться при любой амплитуде стационарных колебаний, отличной от нуля. Дальнейшее увеличение M ведет к повышению амплитуды стационарных колебаний.

В случае жесткого самовозбуждения вариантов поведения автогенератора оказывается больше. Для демонстрации их различий обратимся к графикам, изображенным на рис.7.6.

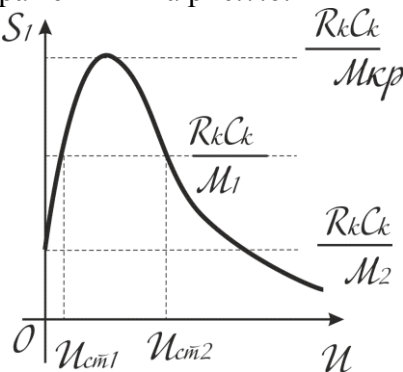


Рис.7.6. Особенности жесткого режима самовозбуждения автогенератора

В области малых значений M в данном случае, так же как и при мягком возбуждении устойчивый стационарный режим соответствует нулевому значению амплитуды колебаний. При увеличении M сверх критического значения начинается область характеристики, которая имеет две точки пересечения с прямой обратной связи. Для всех значений M из этого интервала картина развития автоколебаний является схожей, а именно, несмотря на наличие точек пересечения, самовозбуждение в генераторе не происходит.

Если же какая-либо внешняя сила выведет автогенератор в режим соответствующий первой точке пересечения характеристики и прямой обратной связи

(точки на восходящей ветви характеристики), то возникшие колебания будут неустойчивыми. Амплитуда данных колебаний будет нарастать пока система не перейдет в состояние, соответствующее второй точке пересечения (на нисходящей ветви). В этом состоянии будут наблюдаться устойчивые колебания со стационарной амплитудой.

Увеличение коэффициента взаимной индукции до величины M_2 переводит автогенератор в состояние, в котором возможно его самовозбуждение. Дальнейшее увеличение M будет первоначально приводить к постепенному снижению амплитуды стационарных колебаний, а затем и срыву колебаний с падением амплитуды до нуля.

Рассмотренные особенности возникновения колебаний в автогенераторе при жестком режиме самовозбуждения показывают, что при различных значениях коэффициента обратной связи принципиально меняется картина процессов возбуждения. Это явление часто называют **колебательным гистерезисом**.

Аналитический анализ функции средней крутизны

$$S_1(U) = \frac{I_1(U)}{U}$$

можно провести, например, применив ее разложение в степенной ряд с учетом выбора рабочей точки

$$S_1(U) = a_0 + a_1(U - U_0) + a_2(U - U_0)^2 + a_3(U - U_0)^3 + \dots + a_n(U - U_0)^n,$$

где $a_k, k = 0, 1, 2, \dots, n$

Для глубоко анализа часто требуется учет степеней аппроксимирующего полинома до 4-5 степени.

7.5. Нелинейное уравнение автогенератора

Для получения прямого аналитического решения укороченного нелинейного дифференциального уравнения автогенератора (7.12)

$$\frac{dU(t)}{dt} + \frac{1}{2} \left(\frac{r_k}{L_k} - \frac{M}{L_k C_k} S_1(U) \right) U(t) = 0$$

воспользуемся аппроксимацией средней крутизны ВАХ активного элемента. При рассмотрении вопросов воздействия гармонических сигналов на нелинейные элементы для случая описания характеристики элемента степенным рядом было получено уравнение

$$I_1 = a_1 U_m + \frac{3}{4} a_3 U_m^3 + \frac{5}{8} a_5 U_m^5 + \dots$$

Откуда с учетом того, что $S_1(U) = \frac{I_1(U)}{U}$, и ограничившись первыми двумя членами разложения запишем

$$S_1(U) \approx a_1 + \frac{3}{4} a_3 U^2.$$

Из данного уравнения понятно, что при $a_3 > 0$ реализуется жесткий режим самовозбуждения, а при $a_3 < 0$ – мягкий.

Подстановка этого аппроксимирующего выражения в укороченное уравнение дает

$$\frac{dU}{dt} + \frac{1}{2} \left(\frac{r_k}{L_k} - a_1 \frac{M}{L_k C_k} - \frac{3}{4} a_3 \frac{M}{L_k C_k} U^2 \right) U = 0$$

Допустим, что в начальный момент времени в автогенераторе существуют колебания с некой малой амплитудой

$$U(0) = U_0.$$

Для сокращения громоздкости записи введем обозначения

$$\alpha = \frac{1}{2} \left(a_1 \frac{M}{L_k C_k} - \frac{r_k}{L_k} \right),$$

$$\beta = \frac{3}{8} a_3 \frac{M}{L_k C_k}.$$

Из выше сказанного ясно, что $\alpha > 0, \beta < 0$.

Таким образом, запись укороченного уравнения упростилась до следующего вида

$$\frac{dU}{dt} = (\alpha + \beta U^2)U. \quad (7.18)$$

Умножив обе части уравнения (7.18) на U и произведя процедуру разделения переменных ($U \frac{dU}{dt} = \frac{1}{2} \frac{dU^2}{dt}$), получаем

$$\frac{dU^2}{2(\alpha + \beta U^2)U^2} = dt.$$

Представив левую часть данного уравнения как сумму элементарных дробей имеем

$$\frac{B_1}{\alpha + \beta U^2} + \frac{B_2}{U^2} = dt,$$

где $B_1 = -\frac{1}{2\alpha} \beta dU^2,$
 $B_2 = \frac{1}{2\alpha} dU^2.$

Тогда можно записать

$$\frac{-\frac{1}{2\alpha} \beta dU^2}{\alpha + \beta U^2} + \frac{\frac{1}{2\alpha} dU^2}{U^2} = \frac{1}{2\alpha} \left[\frac{dU^2}{U^2} - \frac{d(\alpha + \beta U^2)}{\alpha + \beta U^2} \right] = dt.$$

Интегрируя полученное уравнение получаем

$$\frac{1}{2\alpha} [\ln U^2 - \ln(\alpha + \beta U^2) + \ln A] = \frac{1}{2\alpha} \ln \frac{AU^2}{\alpha + \beta U^2} = t,$$

где A – постоянная интегрирования, которую необходимо найти.

Для нахождения постоянной интегрирования A воспользуемся предположением относительно начальной амплитуды колебаний, которое мы сделали в начале анализа

$$U(0) = U_0.$$

Тогда

$$\frac{1}{2\alpha} \ln \frac{AU_0^2}{\alpha + \beta U_0^2} = 0,$$

и для искомой постоянной интегрирования получаем

$$A = \frac{\alpha + \beta U_0^2}{U_0^2}.$$

Окончательный результат для амплитуды колебаний теперь получается в виде

$$U(t) = \frac{U_0 \sqrt{\alpha} e^{\alpha t}}{\sqrt{\alpha + \beta U_0^2 (1 - e^{2\alpha t})}}.$$

Стационарное значение амплитуды колебаний можно получить из данного уравнения при предельном переходе $t \rightarrow \infty$, что дает следующий результат

$$\lim_{t \rightarrow \infty} U(t) = \lim_{t \rightarrow \infty} \frac{1}{\sqrt{\frac{\alpha}{\alpha e^{2\alpha t} U_0^2} + \frac{\beta U_0^2}{\alpha e^{2\alpha t} U_0^2} - \frac{e^{2\alpha t} \beta U_0^2}{\alpha e^{2\alpha t} U_0^2}}} = \sqrt{-\frac{\alpha}{\beta}}.$$

подстановка в полученное уравнение значений элементов цепи автогенератора дает

$$U_{ст} = \sqrt{-\frac{\alpha}{\beta}} = \sqrt{\frac{\frac{r_k C_k}{M} - a_1}{\frac{3}{4} a_3}}. \quad (7.19)$$

Справедливость полученного выражение (7.19) легко проверить, так как стационарный режим должен соответствовать точке пересечения прямой обратной связи (7.15)

$$S_1(U) = \frac{r_k C_k}{M}$$

и аппроксимации характеристики средней крутизны, принятой при анализе,

$$S_1(U) \approx a_1 + \frac{3}{4} a_3 U^2.$$

Подстановка в эти два уравнения $U_{ст}$ и их решение дает результат идентичный полученному путем решения укороченного дифференциального уравнения автогенератора.

7.6. Основные схемы реализации LC-генераторов

Среди всего многообразия схем реализации LC-автогенераторов различают трансформаторную схему (вариант которой подробно рассматривался в предыдущих разделах) и так называемые трехточечные схемы.

Особенностью трехточечных схем является то, что сигнал обратной связи снимается не целиком с контура, а с отдельной его части. Упрощенные принципиальные схемы автогенераторов представлены на рис.7.7.

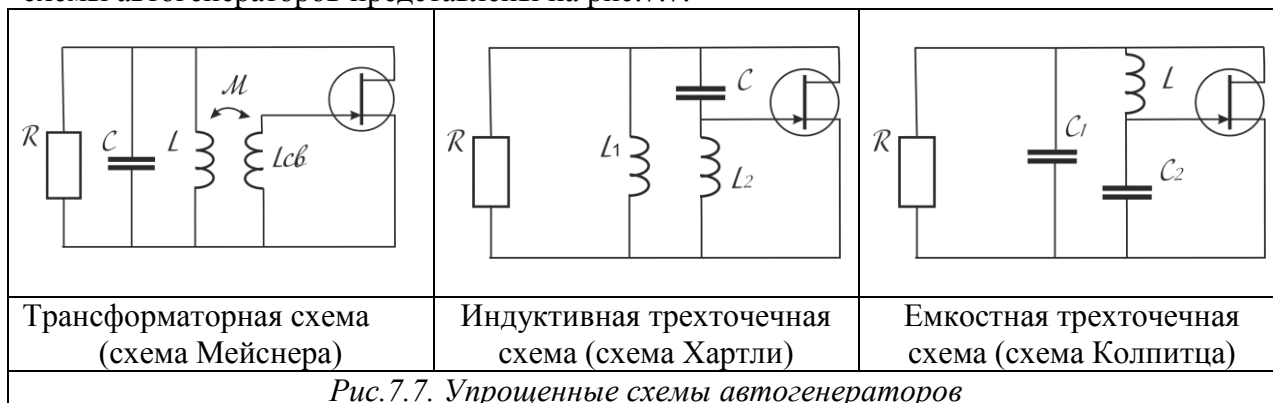


Рис.7.7. Упрощенные схемы автогенераторов

Трехточечная схема построения автогенератора использует применение принципа обратной связи, при котором эта связь реализуется через делитель с комплексными сопротивлениями, как показано на рис.7.8

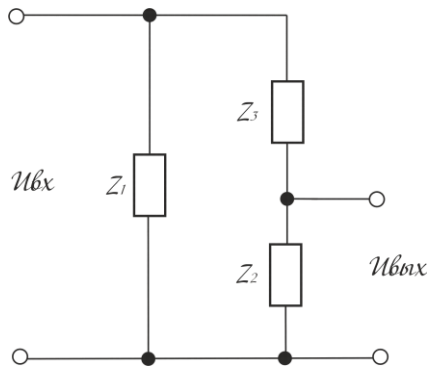


Рис.7.8. Делитель напряжения в цепи обратной связи трехточечных схем автогенератора

Поэтому условиями возникновения колебаний в этом случае являются (соблюдение баланса амплитуд и баланса фаз) в этом случае являются

$$|W_{oc}(j\omega_r)||K(j\omega_r)| = 1$$

и

$$\varphi_K(\omega_r) + \varphi_{oc}(\omega_r) = 2\pi k, k = 0, \pm 1, \pm 2, \dots$$

где $K(j\omega)$ – частотная передаточная функция (комплексный коэффициент передачи) усилительного элемента,

ω_r – частота автоколебаний генератора.

Для определения частоты колебаний автогенератора можно воспользоваться

уравнением для резонансного режима параллельного колебательного контура, в соответствии с которым,

$$X_1(\omega_0) + X_2(\omega_0) + X_3(\omega_0) = 0.$$

Для простоты рассмотрения активную составляющую параллельных колебательных контуров генератора положим равным 0.

Для индуктивной трехточечной схемы резонансная частота может быть определена в виде

$$\omega_0 = \frac{1}{\sqrt{(L_1 + L_2)C}}. \quad (7.20)$$

Для емкостной трехточечной схемы резонансная частота равна

$$\omega_0 = \sqrt{L \frac{C_1 C_2}{C_1 + C_2}}. \quad (7.21)$$

Для трансформаторной схемы

$$\omega_0 = \frac{1}{\sqrt{LC}}. \quad (7.22)$$

Когда цепь обратной связи находится в резонансном режиме, рассчитать соответствующее этому режиму значение коэффициента передачи обратной связи очень просто.

Для индуктивной трехточечной схемы кольцевой ток контура связан с напряжением на контуре соотношением

$$j\omega_0 L_1 I_k^\circ = U_k^\circ,$$

а напряжение снимается с индуктивности L_2 , поэтому

$$W_{oc}(j\omega_0) = -\frac{L_2}{L_1} = \frac{L_2}{L_1} e^{j\pi},$$

то есть $|W_{oc}(j\omega_0)| = \frac{L_2}{L_1}$, $\varphi_{oc}(\omega_0) = \pi$.

Аналогичным образом для емкостной трехточечной схемы можно записать

$$I_k^\circ = j\omega_0 C_1 U_k^\circ$$

и далее

$$W_{oc}(j\omega_0) = -\frac{C_1}{C_2} = \frac{C_1}{C_2} e^{j\pi},$$

где $|W_{oc}(j\omega_0)| = \frac{C_1}{C_2}$, $\varphi_{oc}(\omega_0) = \pi$.

Передаточная функция (комплексный коэффициент передачи) резонансного усилителя на резонансной частоте ω_0 в экспоненциальном виде записывается как

$$K(j\omega_0) = |K(j\omega_0)| e^{j\varphi(\omega_0)}$$

В резонансном режиме (7.20)- (7.22) ток в контуре и напряжение связаны через эквивалентное сопротивление R_3 и поэтому являются синфазными. Кроме того ток первой гармоники связан с амплитудой колебаний посредством средней крутизны $S_1(U)$.

Поэтому

$$|K(j\omega_0)| = \frac{U_k}{U} = \frac{R_3 I_1(U)}{U} = R_3 S_1$$

Поскольку трехточечная схема предполагает не полное, а только частичное включение контура в цепь обратной связи, то при расчете эквивалентных сопротивлений на резонансной частоте

$$R_{экв} = \alpha^2 R_3 = \alpha^2 \rho Q,$$

где α – коэффициент включения.

Для индуктивной трехточечной схемы коэффициент включения равен

$$\alpha = \frac{L_1}{L_1 + L_2}.$$

Для емкостной схемы

$$\alpha = \frac{C_2}{C_1 + C_2}.$$

Трехточечные схемы автогенераторов могут быть реализованы не только на представленных в рассмотренных примерах полевых транзисторах, но и на любых других активных усилительных элементах. Общий вид трехточечной схемы с активным элементом представлен на рис.7.9.

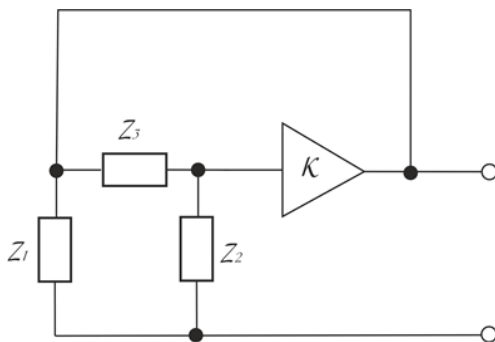


Рис.7.9. Трехточечная схема автогенератора с усилительным элементом.

В упрощенном виде трехточечная схема автогенератора состоит из трех реактивных элементов с комплексными сопротивлениями Z_1, Z_2, Z_3 и усилительного элемента, характеризуемого коэффициентом усиления K (для неинвертирующего варианта) или $-K$ (для инвертирующего включения).

Начнем рассмотрение со случая применения инвертирующего усилителя.

Для простоты будем полагать, что используемые в схеме реактивные элементы являются идеальными, и их сопротивления не имеют вещественной составляющей.

Частотная передаточная функция (комплексный коэффициент передачи) обратной связи для схемы, приведенной на рис.7.6 можно записать в виде

$$W_{oc}(j\omega) = \frac{Z_2}{Z_1 + Z_3} = \frac{X_2}{X_1 + X_3}.$$

В рассматриваемом случае усилительный элемент обладает инвертирующим свойством, поэтому для выполнения условия баланса фаз коэффициент обратной связи тоже должен быть отрицательным. Это условие можно осуществить только, если реактивные сопротивления X_2 и X_3 имеют разные знаки и $|X_2| < |X_3|$. Элемент X_1 в данном случае выбирается одного знака с X_2 , а его величина должна обеспечить получение резонанса в контуре на требуемой частоте генератора

$$X_1 + X_2 + X_3 = 0.$$

Отсюда понятно, что

$$W_{oc}(j\omega) = \frac{X_2}{X_2 + X_3} = -\frac{X_2}{X_1}.$$

С другой стороны для выполнения условия баланса амплитуд необходимо, чтобы

$$-\frac{X_2}{X_1} (-K) = 1, \quad \text{то есть } K = \frac{X_1}{X_2},$$

где K – величина коэффициента усиления при нагрузке на контур Z_1, Z_2, Z_3 .

Влево и вправо от резонансной частоты эквивалентное сопротивление контура резко падает, что приводит к уменьшению $K(\omega)$ на всех частотах, кроме резонансной.

Если в трехточечной схеме используется не инвертирующий усилительный элемент, то элементы X_2 и X_3 следует выбирать одного знака, а условие баланса фаз

$$X_1 + X_2 + X_3 = 0$$

обеспечивать соответствующим выбором X_1 .

Обратная связь формируется

$$W_{oc}(j\omega) = \frac{X_2}{X_2 + X_3}$$

и условие баланса амплитуд выполняется при

$$K = \frac{X_2 + X_3}{X_2},$$

где K – коэффициент усиления усилителя, нагруженного на контур Z_1, Z_2, Z_3 .

Пример 7.2.

Продemonстрируем работу генератора, собранного по емкостной трехточечной схеме, приведенной на рис.7.10. На этом же рисунке приведена осциллограмма возникающих автоколебаний, из которой хорошо видно, что вид полученных колебаний сильно отличается от синусоидального. В качестве упражнения рекомендуем предложить изменения в схеме, которые улучшат вид получаемого колебания.

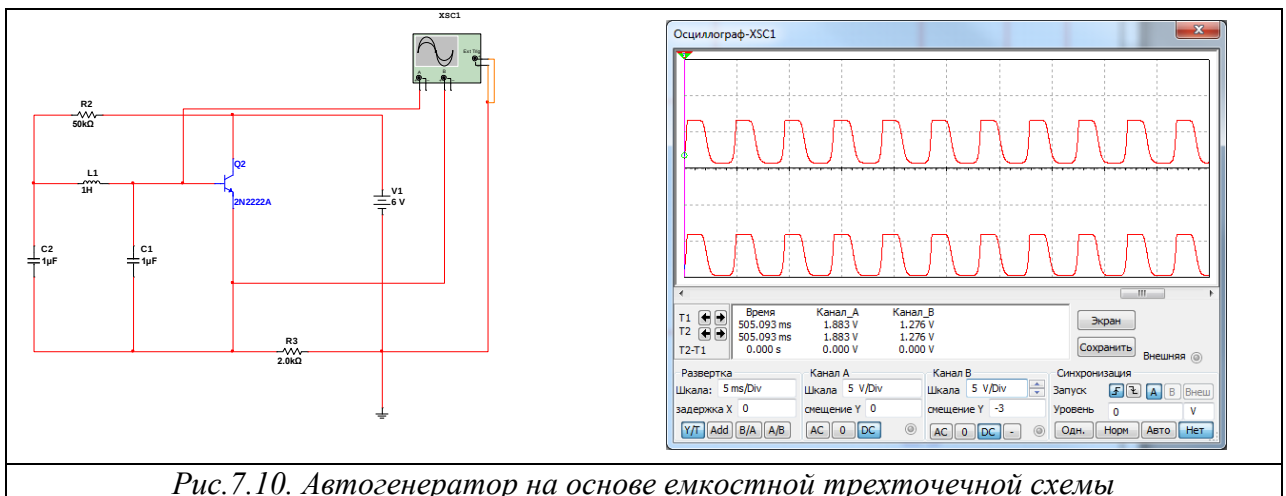


Рис.7.10. Автогенератор на основе емкостной трехточечной схемы

Пример 7.3.

Продemonстрируем работу генератора, собранного по емкостной трехточечной схеме с усилительным элементом в виде идеального операционного усилителя приведенной на рис.7.11. На этом же рисунке приведена осциллограмма возникающих автоколебаний, из которой хорошо видно, что вид полученных колебаний сильно отличается от синусоидального. В качестве упражнения рекомендуем предложить изменения в схеме, которые улучшат вид получаемого колебания.

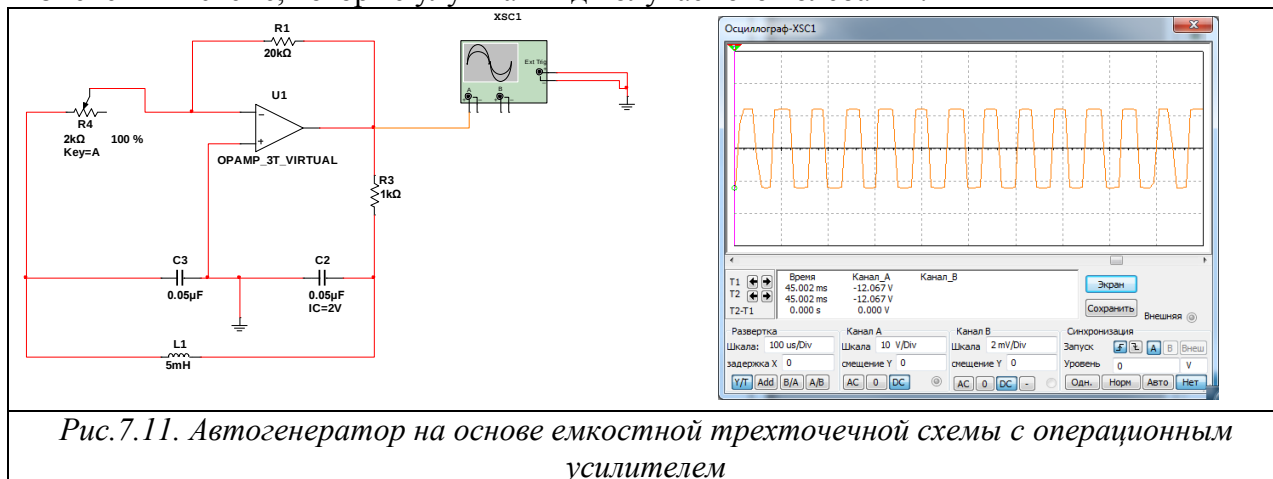


Рис.7.11. Автогенератор на основе емкостной трехточечной схемы с операционным усилителем

7.7.RC-автогенераторы и автогенераторы с внутренней обратной связью

Рассмотренные выше LC-автогенераторы применяются в основном на высоких частотах (свыше десятков кГц). На более низких частотах применение колебательных систем на основе LC-контуров сталкивается с проблемами, обусловленными габаритами необходимых катушек индуктивности. Иногда задача генерации низкочастотного колебания с помощью LC-контуров решается иначе. Два гетеродина (LC-автогенератора) вырабатывают напряжения с частотами примерно равными ω_1 и ω_2 . Эти напряжения усиливаются и подаются на смеситель, в токе которого наряду с другими гармониками есть сигнал с разностной частотой. Сигнал с разностной частотой выделяется с помощью фильтра низких частот и затем усиливается. Перестраивая один из гетеродинов, можно получать различные значения частоты выходного сигнала. Форма выходного колебания очень близка к синусоидальной, однако стабильность поддержания разностной частоты значительно меньше стабильности исходных высокочастотных колебаний.

По этой причине на данных частотах значительно чаще используются RC-автогенераторы, представляющие собой соединение активных четырехполюсников и пассивных RC-цепей, выполняющих роль обратной связи с заданными характеристиками.

7.7.1.Автогенераторы на фазосдвигающих цепочках

Так как широкополосный усилитель низкой частоты в пределах полосы пропускания имеет почти постоянный (не зависящий от частоты) коэффициент усиления и постоянный фазовый сдвиг между входным и выходным напряжениями, то при подключении цепи положительной обратной связи условие баланса фаз может выполняться в пределах полосы пропускания усилителя одновременно для всех частот от ω_1 до ω_2 . Поэтому для получения на выходе гармонических колебаний заданной частоты

нужно создать преимущественные условия именно для этой частоты ω_r . Следовательно, необходимо иметь такую частотно-зависимую обратную связь или колебательную систему, чтобы только для этой одной частоты ω_r выполнялся бы баланс фаз и амплитуд.

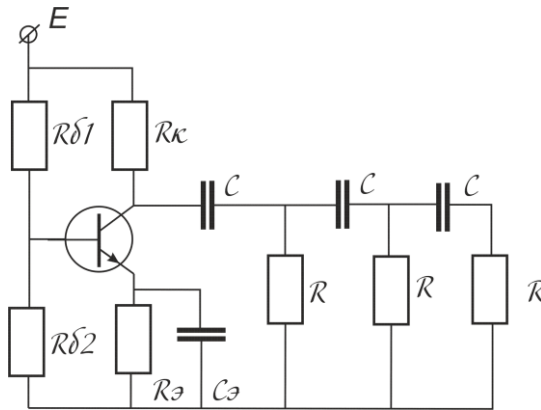


Рис.7.12. Схема автогенератора с RC-фазосдвигающей цепочкой

рис.7.12. Для определения частоты генератора примем следующие условия

$$R_k \ll R, \\ (R_{61} || R_{62}) \gg R.$$

тогда для контурных токов в каждом из звеньев RC-цепочки можно записать

$$\begin{cases} \left(R + \frac{1}{j\omega C}\right) I_1^\circ - R I_2^\circ = U_{кэ}^\circ \\ -R I_1^\circ + \left(2R + \frac{1}{j\omega C}\right) I_2^\circ - R I_3^\circ = 0 \\ -R I_2^\circ + \left(2R + \frac{1}{j\omega C}\right) I_3^\circ = 0 \end{cases}$$

Решая данную систему уравнений относительно I_3° , получаем

$$I_3^\circ = \frac{U_{кэ}^\circ / R}{-\left(5\left(\frac{1}{\omega CR}\right)^2 - 1\right) + j\left(\left(\frac{1}{\omega CR}\right)^3 - 6\frac{1}{\omega CR}\right)}.$$

Поскольку

то

$$K_{oc}^\circ = -\frac{U_{бэ}^\circ}{U_{кэ}^\circ} = \frac{1}{-\left(5\left(\frac{1}{\omega CR}\right)^2 - 1\right) + j\left(\left(\frac{1}{\omega CR}\right)^3 - 6\frac{1}{\omega CR}\right)}.$$

Для обеспечения на частоте генерации фазового сдвига 180° требуется, чтобы $K_{oc}^\circ(\omega_r)$ была действительной и отрицательной величиной. Следовательно,

$$\left(\frac{1}{\omega_r CR}\right)^3 - 6\frac{1}{\omega_r CR} = 0.$$

Откуда получаем

$$\omega_r = \frac{1}{\sqrt{6}RC}. \quad (7.23)$$

На частоте (7.23)

Простейшая RC-цепь, как было рассмотрено ранее, имеет фазо-частотную характеристику, изменяющуюся в пределах от 0 до 90° , поэтому для создания фазового сдвига 180° требуется последовательное соединение нескольких RC-цепочек. Например, при использовании трех цепочек с фазовым сдвигом 60° у каждой удастся получить необходимый суммарный угол сдвига фазы.

Схема автогенератора, реализованного с применением трех RC-цепочек, приведена на

$$K^{\circ}_{oc}(\omega_r) = \frac{1}{29}. \quad (7.24)$$

Для обеспечения баланса амплитуд (7.6) усилитель должен иметь коэффициент усиления

$$K = S_1 R_k = \frac{1}{K^{\circ}_{oc}(\omega_r)} = 29.$$

Таким образом, в данной схеме постоянная времени выбранной RC-цепочки $\tau = RC$ однозначно определяет и частоту генерируемых колебаний и требуемое усиление усилителя.

В рассмотренном примере для сдвига фазы была использована дифференцирующая RC-цепь. Эту же задачу можно решить с помощью интегрирующей RC-цепи.

В случае применения интегрирующей цепи частота колебаний определяется как

$$\omega_r = \frac{\sqrt{6}}{RC}, \quad (7.25)$$

а модуль коэффициента передачи, обеспечивающий выполнение условия баланса амплитуд остается неизменным

$$K = 29. \quad (7.26)$$

Между парами уравнений (7.23), (7.25) и (7.24), (7.26) хорошо заметно обратное симметричное соответствие.

Пример 7.4.

Продemonстрируем работу автогенератора, использующего фазо-сдвигающую цепочку, схема которого приведена на рис.7.13. На этом же рисунке приведена осциллограмма возникающих в автогенераторе колебаний.

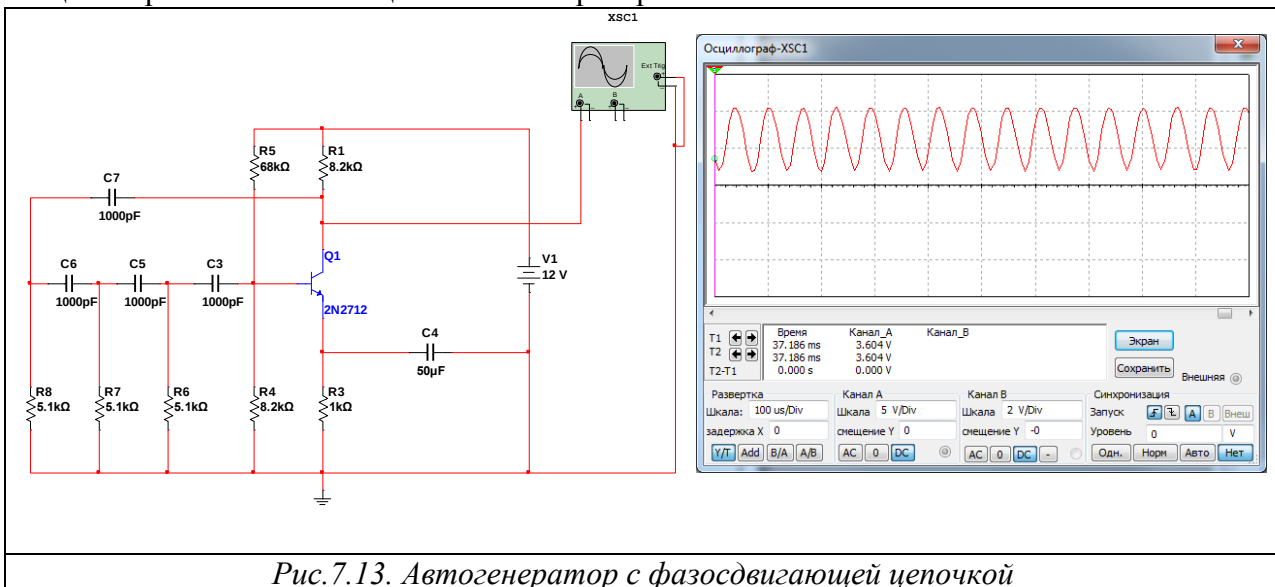


Рис.7.13. Автогенератор с фазосдвигающей цепочкой

Пример 7.5.

Продemonстрируем работу автогенератора, использующего фазосдвигающую цепь и идеальный операционный усилитель, схема которого приведена на рис.7.14. На этом же рисунке приведена осциллограмма возникающих в автогенераторе колебаний.

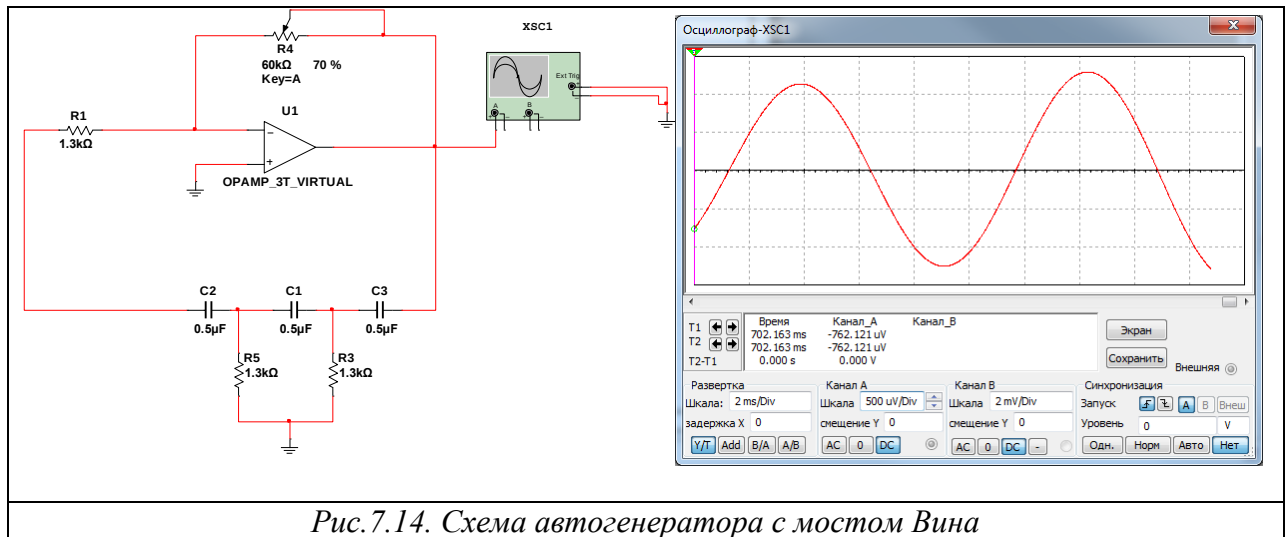


Рис.7.14. Схема автогенератора с мостом Вина

7.7.2.RC- автогенератор с мостом Вина

Автогенератор с мостом Вина представляет собой разновидность электронных генераторов синусоидальных колебаний. Схема основывается на электрической цепи (полосовом фильтре), первоначально разработанной Максом Вином в 1891 г. и известной, как мост Вина.

Автогенератор включает в себя электронный усилитель, охваченный частотно зависимой положительной обратной связью через мост Вина.

Особенность данного типа генератора состоит в том, что он может генерировать колебания в широком диапазоне частот и позволяет получить сигнал с очень малыми отличиями от идеальной синусоиды.

Принципиальная схема генератора приведена на рис. 7.15.

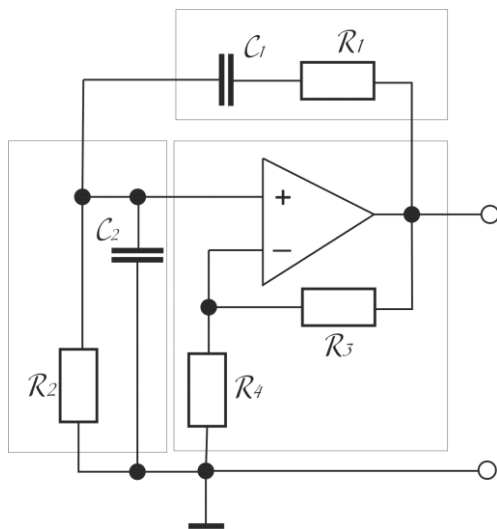


Рис.7.15. Автогенератор с мостом Вина

В обратной цепи данного генератора использовано последовательное соединение параллельной и последовательной RC-цепочек, которое на частоте генерации должно обеспечить выполнение условия баланса фаз. Поэтому для генерируемой частоты легко получить

$$\omega_r C_2 R_1 - \frac{1}{\omega_r C_1 R_2} = 0,$$

откуда

$$\omega_r = \frac{1}{\sqrt{C_2 C_1 R_2 R_1}}. \quad (7.27)$$

На практике чаще всего используются схемы, в которых $R_2 = R_1 = R$ и $C_2 = C_1 = C$, что приводит к упрощению уравнения (7.27) для генерируемой частоты в виде

$$\omega_r = \frac{1}{RC}.$$

Для обеспечения стабильности работы автогенератора используется цепь отрицательной обратной связи, глубина которой определяется соотношением резисторов R_3 и R_4 , в качестве последнего из которых иногда применяют термозависимый резистор.

Сопротивление терморезистора растет с ростом его температуры, что позволяет стабилизировать амплитуду генерируемых колебаний.

Пример 7.6.

Продemonстрируем работу автогенератора, использующего мост Вина и идеальный операционный усилитель, схема которого приведена на рис.7.16. На этом же рисунке приведена осциллограмма возникающих в автогенераторе колебаний. В качестве упражнения рекомендуем предложить изменения в схеме, которые улучшат вид получаемого колебания.

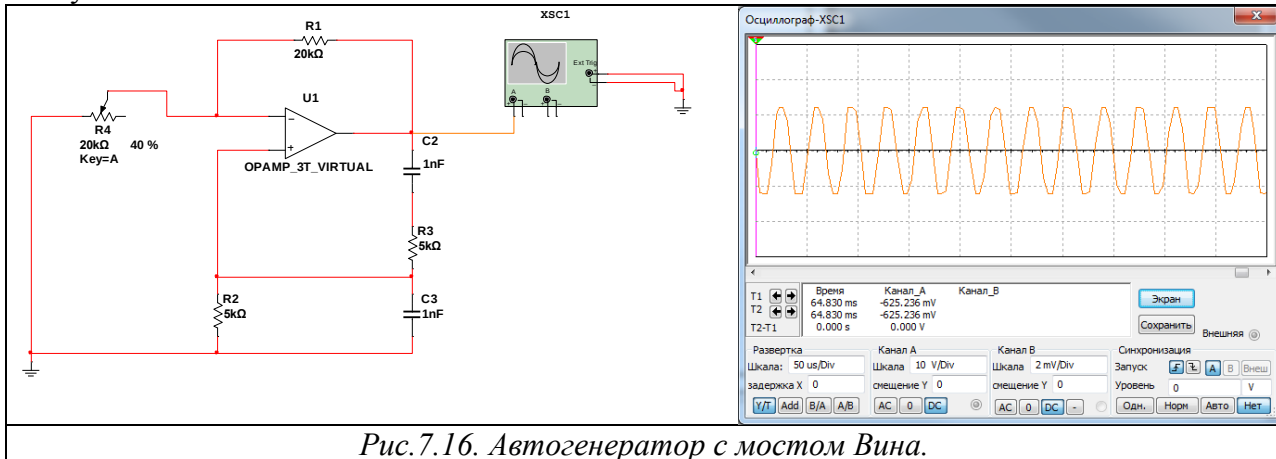


Рис.7.16. Автогенератор с мостом Вина.

7.7.3.Автогенераторы с внутренней обратной связью

При рассмотрении процесса возникновения автоколебаний в схеме автогенератора с трансформаторной связью, было получено условие генерации (7.9) в виде

$$\frac{r_k}{L_k} \pm \frac{MS}{L_k C_k} < 0,$$

которое можно выполнить при отрицательном знаке перед вторым слагаемым. В этом случае величина условие выполнялось за счет использования обратной связи, реализуемой посредством взаимной индуктивной связи M двух катушек и их соответствующего включения.

Существуют такие электрические и полупроводниковые приборы, которые позволяют получить отрицательную проводимость (сопротивление) без введения дополнительной ОС, а за счет падающих участков ВАХ. К такому типу элементов, например, можно отнести туннельные диоды. Типичная ВАХ элемента с отрицательным дифференциальным сопротивлением представлен на рис.7.17.

На падающем участке характеристики дифференциальные проводимость и сопротивление являются отрицательными

$$g_{d-} = \frac{1}{R_{d-}} = \frac{di}{du} < 0. \quad (7.28)$$

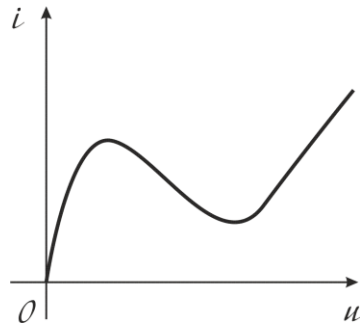


Рис.7.17. ВАХ элемента с отрицательным дифференциальным сопротивлением.

При включении элемента с отрицательным дифференциальным сопротивлением (7.28) в колебательный контур его эквивалентная схема может быть представлена в виде, изображенном на рис.7.18.

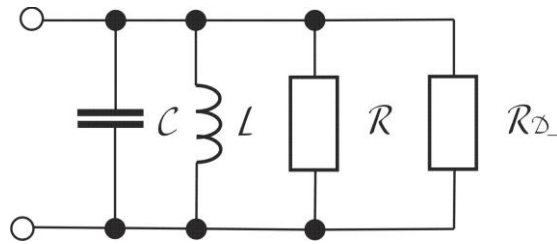


Рис.7.18. Эквивалентная схема контура с элементом, имеющим отрицательное дифференциальное сопротивление

В случае, если

$$R_{д-} = R_{\text{э}},$$

где $R_{\text{э}}$ – эквивалентное резонансное сопротивление контура, потери контура в контуре компенсируются и в нем

могут установиться незатухающие колебания.

Принципиальная схема автогенератора на туннельном диоде приведена на рис.7.19.

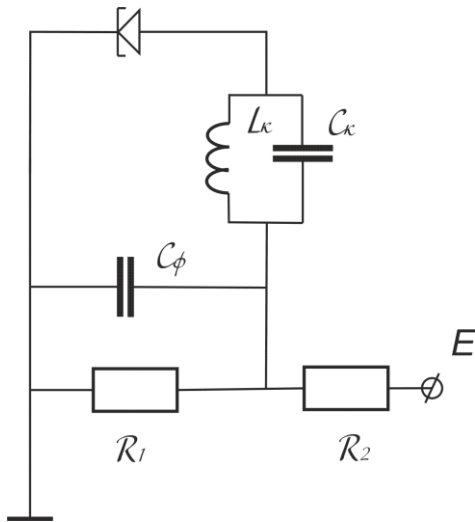


Рис.7.19. Схема автогенератора на туннельном диоде.

С помощью сопротивлений R_1 и R_2 устанавливается рабочая точка диода, соответствующая середине падающего участка ВАХ. Эквивалентная схема генератора примет вид, представленный на рис. 7.20.

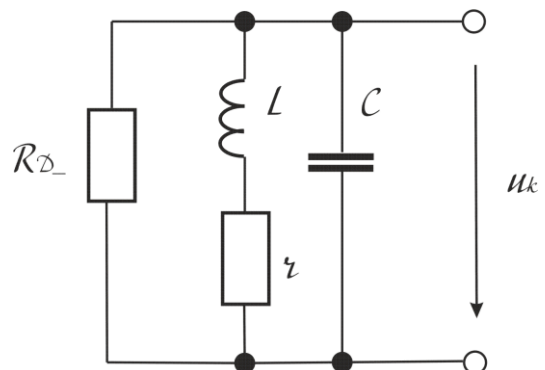


Рис.7.20. Эквивалентная схема автогенератора на туннельном диоде

Ток через диод определяется в виде

$$i_{\text{д}} = -u_k / R_{\text{д-}}$$

Так как ток в диоде равен сумме токов через емкость и индуктивность, каждый из которых в свою очередь связан с напряжением на контуре соотношениями

$$i_c = C \frac{du_k}{dt} \quad \text{и} \quad u_k = r_k i_L + L \frac{di_L}{dt}$$

общее уравнение для тока в диоде запишем в виде

$$i_d = i_c + i_L = i_L + r_k C \frac{di_L}{dt} + LC \frac{d^2 i_L}{dt^2}.$$

Объединяя два полученных для тока диода выражения в одно уравнение получаем

$$i_L + r_k C \frac{di_L}{dt} + LC \frac{d^2 i_L}{dt^2} = -\frac{1}{R_{д-}} \left(r_k i_L + L \frac{di_L}{dt} \right),$$

откуда итоговое дифференциальное уравнение имеет вид

$$\frac{d^2 i_L}{dt^2} + \left(\frac{r_k}{L} + \frac{1}{CR_{д-}} \right) \frac{di_L}{dt} + \frac{1 + \frac{r_k}{R_{д-}}}{LC} i_L = 0.$$

Из полученного уравнения по аналогии с уравнением автогенератора с трансформаторной связью получаем условие самовозбуждения

$$\frac{r_k}{L} + \frac{1}{CR_{д-}} < 0 \text{ или } \frac{r_k}{L} - \frac{1}{C|R_{д-}|} < 0. \quad (7.29)$$

Из неравенства (7.29) условие самовозбуждения можно записать иначе

$$|R_{д-}| < \frac{L}{r_k C} = \rho Q = \frac{Q}{\omega_0 C}. \quad (7.30)$$

При выполнении неравенства (7.30) в системе начнутся автоколебания, рост амплитуды которых прекратится в стационарном режиме при

$$R_{д-}(U_{ст}) = R_3.$$

7.8. Вопросы для самоконтроля и повторения

1. Приведите общую схему автоколебательной системы.
2. Какие устройства называют автогенераторами?
3. Какой режим работы автогенератора называют стационарным?
4. Сформулируйте условия баланса фаз.
5. Сформулируйте условие баланса амплитуд.
6. Получите уравнение для условий баланса фаз и амплитуд.
7. Сформулируйте условия возможности самовозбуждения автогенератора.
8. Каким образом определяются параметры стационарного режима работы автогенератора?
9. Поясните разницу между мягким и жестким режимами самовозбуждения автогенератора.
10. В чем состоит явление колебательного гистерезиса?
11. Запишите нелинейное уравнение автогенератора.
12. Перечислите основные схемы реализации LC-генераторов.
13. Поясните принципы построения RC-генераторов.
14. Поясните принцип работы генераторов на фазосдвигающих цепочках.
15. Поясните принцип работы RC-генератора на мосте Вина.
16. Поясните принцип построения генераторов с внутренней обратной связью

ГЛАВА 8. Выделение сигнала на фоне помех. Принципы оптимальной линейной фильтрации сигналов

8.1. Фильтрация и отношение сигнал-шум

Практически во всех радиотехнических системах возникает задача выделения полезного сигнала на фоне шума. Система должна быть спроектирована таким образом, чтобы противостоять мешающему воздействию помех и обеспечивать тем самым высокую помехоустойчивость. Для достижения данной цели используются два основных подхода. Первый подход повышения помехоустойчивости состоит в создании устройств обработки принятой информации, способных наилучшим образом выделить сигнал, искажённый присутствием помех. Другой подход заключается в совершенствовании структуры передаваемых сигналов, использовании помехоустойчивых методов кодирования и модуляции.

При проектировании радиотехнических систем особый интерес представляет вопрос ослабления вредного действия помехи с помощью линейной фильтрации, основанной на выборе линейных фильтров с соответствующей частотной характеристикой.

Исторически сложилось так, что начальном этапе развития радиотехники к фильтрам, используемым для устранения помехи, предъявлялось требование возможно более равномерного пропускания спектра сигнала и возможно более полного подавления частот вне этого спектра. Соответственно, идеальным считался полосовой фильтр с прямоугольной АЧХ. По мере развития теории информации и статистической теории обнаружения сигналов подход к разработке линейных фильтров существенно изменился.

При решении задачи выбора АЧХ фильтра стали принимать во внимание форму сигнала и статистические свойства помехи. В результате выяснилось, что полосовой фильтр с П-образной частотной характеристикой в тех случаях, когда имеется априорная информация о характеристиках сигнала и помехи не является оптимальным.

Переломный момент в теории и практике линейной фильтрации произошел после появления работ Н. Винера и А. Н. Колмогорова, которые поставили и решили задачу синтеза фильтра, оптимального в определенном смысле для приема заданного сигнала, действующего на фоне помехи с заданными статистическими характеристиками.

Среди множества областей применения согласованной фильтрации можно так же выделить задачи повышения разрешающей способности по дальности локационных систем и системы скрытой передачи информации.

В зависимости от специфики конкретной задачи (обнаружение сигнала, оценивание параметров сигнала, и т.д.) критерии оптимальности могут быть разными. В задачах обнаружения сигналов на фоне помех, например, наибольшее распространение получил критерий максимума отношения сигнал-шум на выходе фильтра.

Отношение сигнал-шум чаще всего выражают в децибелах

$$SNR1 = 20 \lg \frac{|s(t)|_{\max}}{\sigma} \quad (8.1)$$

или

$$SNR2 = 10 \lg \frac{\langle s^2 \rangle}{\sigma^2}, \quad (8.2)$$

где σ – среднеквадратическое значение шума,
 $\langle s^2 \rangle$ - средняя мощность полезного сигнала,
 $|s(t)|_{max}$ – максимальное абсолютное значение сигнала.

Наибольшее возможное пиковое значение сигнала (8.1) чаще всего используется в задачах обнаружения, при этом не ставится условие сохранения формы сигнала, так как для обнаружения его в шумах форма значения не имеет.

Если на входе линейного фильтра с передаточной функцией $W(s)$ сигнал присутствует вместе с аддитивной помехой в виде

$$u_{вх}(t) = s_1(t) + n_1(t),$$

а на выходе получается сигнал

$$u_{вых}(t) = s_2(t) + n_2(t),$$

то задача создания наилучшего фильтра в соответствии с (8.2) состоит в поиске такой $W(s)$, чтобы

$$SNR = 10 \lg \frac{\langle s_2^2 \rangle}{\sigma_2^2} \rightarrow \max.$$

Если значение сигнал-шум для выхода фильтра оказывается больше, чем для входного сигнала, то говорят, что фильтр дает выигрыш, величина которого оценивается отношением

$$M_\phi = \frac{SNR_{вых}}{SNR_{вх}}$$

Рассмотрим случай, когда сигнал на входе фильтра представлен в виде

$$s_1(t) = \sum_{k=0}^{\infty} A_k \cos(\omega_k + \theta_k),$$

а помеха имеет дисперсию, определяемую выражением

$$\sigma_1^2 = \int_0^{\infty} F_n(\omega) d\omega.$$

В соответствии с формулой Релея средняя мощность полезного сигнала на входе фильтра равна

$$\langle s_1^2 \rangle = \sum_{k=0}^{\infty} \frac{A_k^2}{2}.$$

Если частотная передаточная функция фильтра задана в виде

$$W(j\omega) = W(s)|_{s=j\omega},$$

то для выходного сигнала и выходной помехи можно записать

$$\langle s_2^2 \rangle = \sum_{k=0}^{\infty} \frac{A_k^2}{2} |W(j\omega)|^2,$$

$$\sigma_2^2 = \int_0^{\infty} F_n(\omega) |W(j\omega)|^2 d\omega.$$

Тогда отношение сигнал-шум на выходе фильтра будет определяться как

$$SNR_2 = \frac{\sum_{k=0}^{\infty} \frac{A_k^2}{2} |W(j\omega)|^2}{\int_0^{\infty} F_n(\omega) |W(j\omega)|^2 d\omega}. \quad (8.3)$$

Полученная формула (8.3) показывает, что располагая информацией о спектрах полезного сигнала и шума, можно таким образом подобрать параметры фильтра, чтобы получить значительное увеличение отношений сигнал-шум в выходном сигнале. Однако следует помнить, что данное уравнение полностью игнорирует фазовые сдвиги сигналов, происходящие в фильтре. Поэтому искажения полезного сигнала могут быть весьма существенными, что допустимо в задачах обнаружения, но не приемлемо, если полезный сигнал должен быть выделен с минимальными изменениями формы.

8.2. Согласованная фильтрация детерминированных сигналов

Рассмотрим случай, когда полезная составляющая входного сигнала $s_1(t)$ является полностью детерминированной. Искомый линейный фильтр в этом случае должен обеспечивать максимум пикового значения сигнал, чтобы

$$SNR = 20 \lg \frac{|s_2(t)|_{\max}}{\sigma} \rightarrow \max. \quad (8.4)$$

Тогда для выходного сигнала линейного фильтра справедливо уравнение

$$s_2(t) = \int_{-\infty}^{\infty} s_1(\tau) g(t - \tau) d\tau, \quad (8.5)$$

где $g(t)$ – импульсная характеристика линейного фильтра.

Для определения условия обеспечивающего максимум принятому критерию (8.4), а соответственно и записанному интегралу свертки исходного сигнала и импульсной характеристики фильтра (8.5) воспользуемся неравенством Коши-Буняковского, в соответствии с которым

$$\left| \int_a^b x(t)y(t)dt \right|^2 \leq \left| \int_a^b x(t)dt \right| \left| \int_a^b y(t)dt \right|.$$

При этом равенство в приведенном соотношении достигается только в случае, когда функции $x(t)$ и $y(t)$ связаны между собой соотношением

$$y(t) = kx(t),$$

где $k = \text{const}$.

По аналогии с этим для условия достижения выходным сигналом $s_2(t)$ максимума в некоторый момент времени t_0 запишем

$$g(t_0 - \tau) = As_1(\tau),$$

где $A = \text{const}$.

Откуда для импульсной характеристики искомого фильтра получаем

$$g(t) = As_1(t_0 - t). \quad (8.6)$$

Фильтр с импульсной характеристикой (8.6) называют согласованным с сигналом $s_1(t)$. Для физической реализуемости такого фильтра его импульсная характеристика должна равняться нулю при отрицательных значениях аргумента. При этом понятно, что значение времени t_0 не может быть меньше длительности полезного сигнала $s_1(t)$,

поскольку для формирования оптимального отклика необходима информация обо всем входном сигнале. Следовательно, такой фильтр может быть реализован только для сигналов с конечной длительностью.

Необходимое условие физической реализуемости четырехполюсника, имеющего необходимую импульсную характеристику согласованного фильтра, дает критерий Пэли-Винера

$$\int_0^{\infty} \frac{|\ln(|W(j\omega)|)|}{1 + \omega^2} d\omega < \infty.$$

На рис.8.1. продемонстрирован принцип получения необходимой импульсной характеристики согласованного фильтра: она представляет собой зеркально отраженную и сдвинутую на t_0 масштабную копию входного сигнала.

Выходной сигнал $s_2(t)$ при известном входном сигнале $s_1(t)$ и импульсной характеристике фильтра $g(t)$ можно получить по уравнению

$$s_2(t) = \int_{-\infty}^{\infty} s_1(\tau) A s_1(t_0 - t + \tau) d\tau = A \int_{-\infty}^{\infty} s_1(\tau) s_1(\tau - (t - t_0)) d\tau.$$

В полученном уравнении интеграл представляет собой корреляционную функцию сигнала $s_1(t)$, рассчитанную для значения аргумента $t - t_0$, то есть выходной сигнал фильтра $s_2(t)$ с масштабным коэффициентом A повторяет корреляционную функцию входного сигнала

$$s_2(t) = A R_{s_1}(t - t_0).$$

Поскольку корреляционная функция имеет максимальное значение в начале координат, то максимальное значение сигнала $|s_2(t)|_{max}$ достигается в момент времени t_0 .

Имея выражение для импульсной характеристики фильтра, можно найти его передаточную функцию (применив преобразование Лапласа) или частотную передаточную функцию (применив преобразование Фурье). Для частотной передаточной функции искомого фильтра запишем

$$W(j\omega) = \int_{-\infty}^{\infty} A s_1(t_0 - t) e^{-j\omega t} dt.$$

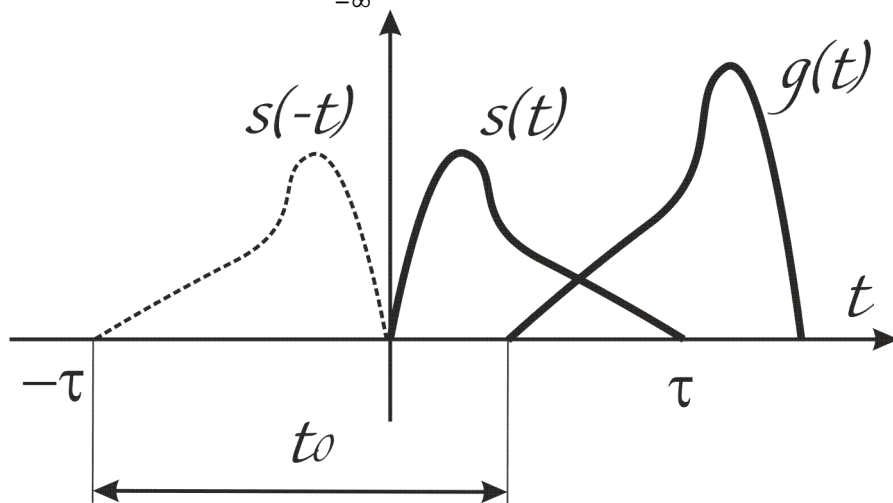


Рис.8.1. Соотношение сигнала и импульсной характеристики согласованного фильтра

Данное уравнение после замены переменных дает

$$W(j\omega) = \int_{-\infty}^{\infty} A s_1(\tau) e^{-j\omega(t_0-\tau)} d(-\tau) = A e^{-j\omega t_0} \int_{-\infty}^{\infty} s_1(\tau) e^{j\omega\tau} d\tau. \quad (8.7)$$

Полученный интеграл (8.7) представляет собой не что иное, как спектральную функцию входного сигнала $s_1(t)$, вычисленную для аргумента $-\omega$. Для спектра вещественного сигнала $s_1(t)$ справедливо соотношение

$$\mathbf{S}_1(-\omega) = \mathbf{S}_1^*(\omega),$$

где $\mathbf{S}_1^*(\omega)$ – комплексно сопряженная функция к $\mathbf{S}_1(\omega)$.

Таким образом, для частотной передаточной функции искомого фильтра окончательно получаем

$$W(j\omega) = A \mathbf{S}_1^*(\omega) e^{-j\omega t_0}.$$

Из этого уравнения следует, что модуль частотной передаточной функции (коэффициента передачи) масштабно повторяет модуль спектральной плотности сигнала

$$|W(j\omega)| = A |\mathbf{S}_1(\omega)|, \quad (8.8)$$

а фазовый сдвиг сигнала на выходе фильтра будет иметь вид

$$\varphi(\omega) = \theta_1(\omega) + [-\theta_1(\omega) - \omega t_0] = -\omega t_0. \quad (8.9)$$

Анализ выражения (8.9) для полученного фазового сдвига демонстрирует эффект, получивший название компенсации начальных фаз, состоящий в том, что в момент времени t_0 все спектральные составляющие сигнала складываются синфазно и образуют тем самым пик выходного отклика фильтра.

Если входной шум $n_1(t)$ является белым с постоянной спектральной плотностью мощности $S_{P0}(\omega) = S_N$, то для спектральной плотности мощности шумовой составляющей выходного сигнала в силу (8.8) справедливо соотношение

$$S_{P2}(\omega) = |W(j\omega)|^2 S_N = A^2 |\mathbf{S}_1(\omega)|^2 S_N.$$

Соответственно дисперсия случайной составляющей выходного сигнала определяется из соотношения

$$\sigma_2^2 = \frac{A^2 S_N}{2\pi} \int_{-\infty}^{\infty} \mathbf{S}_1(\omega) \mathbf{S}_1^*(\omega) d\omega = A^2 S_N E_1,$$

где E_1 – энергия входного сигнала.

Корреляционная функция выходного сигнала

$$R_2(\tau) = S_N B_g(\tau) = A^2 S_N R_1(\tau).$$

Выходной сигнал $s_2(t)$ достигает максимума в момент времени t_0

$$|s_2(t)|_{\max} = A R_{s1}(t - t_0)|_{t=t_0} = A R_{s1}(0) = A E_1.$$

Тогда отношение сигнал-шум на выходе фильтра можно определить как

$$SNR_2 = 20 \lg \frac{A E_1}{\sqrt{A^2 S_N E_1}} = 20 \lg \sqrt{\frac{E_1}{S_N}}. \quad (8.10)$$

Поскольку числитель в полученной формуле (8.10) представляет собой предельно возможный отклик, то согласованная фильтрация в данном случае обеспечивает максимально возможное отношение сигнал-шум. Поэтому согласованный фильтр является оптимальным для обработки аддитивной смеси детерминированного сигнала и белого гауссовского шума, то есть обеспечивающим максимально-возможное отношение сигнал-шум в выходном сигнале.

Полученная формула (8.10) показывает, что отношение сигнал-шум не зависит от формы сигнала, а определяется только его энергией и спектральной плотностью мощности шума. Выбор формы сигнала зависит от конкретной задачи. Так, например, для повышения скрытности сигнала можно увеличить его длительность (при сохранении ширины спектра) с одновременным уменьшением амплитуды. Это действие приводит к понижению отношения сигнал-шум и поэтому затруднит его обнаружение. Лишь согласованно настроенный приемник позволит получить максимально возможное отношение сигнал-шум.

Пример 8.1.

Рассмотрим вопрос создания согласованного фильтра для полезного сигнала имеющего вид прямоугольного импульса.

Пусть детерминированная составляющая входного сигнала имеет вид

$$s_1(t) = \begin{cases} 0, & \text{при } t < 0 \\ 1, & \text{при } t \in [0, T] \\ 0, & \text{при } t > T \end{cases}$$

Спектральная плотность этого сигнала определяется выражением

$$S_1(\omega) = \int_{-\infty}^{\infty} s_1(t) e^{-j\omega t} dt = \int_0^T e^{-j\omega t} dt = \frac{1}{j\omega} (1 - e^{-j\omega T})$$

Частотная передаточная функция (комплексный коэффициент передачи) искомого согласованного фильтра может быть найдена из уравнения

$$W(j\omega) = A S_1^*(\omega) e^{-j\omega T} = \frac{A}{-j\omega} (1 - e^{j\omega T}) e^{-j\omega T} = \frac{A}{j\omega} (1 - e^{-j\omega T})$$

Передаточная функция искомой цепи имеет вид

$$W(s) = \frac{A}{s} (1 - e^{-sT}).$$

Соответственно цепь, реализующая такую передаточную функцию, будет состоять из следующих основных блоков: усилителя, интегратора, блока задержки сигнала и сумматора. Структурная схема такой цепи приведена на рис.8.2.

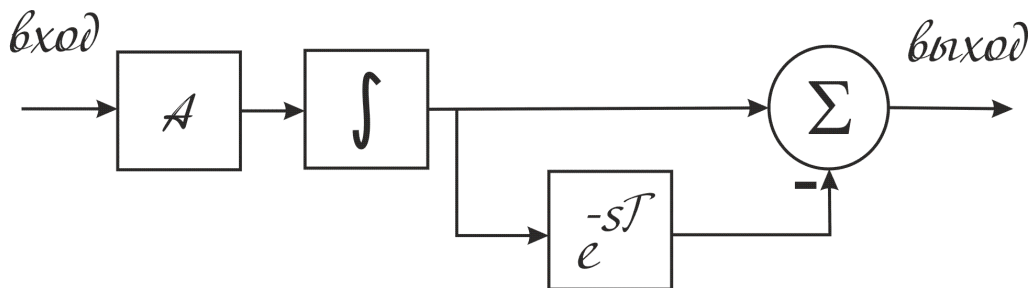


Рис.8.2. Структурная схема согласованного фильтра для одиночного импульса.

8.3. Оптимальная фильтрация при не белом шуме

В предыдущем разделе был рассмотрен согласованный фильтр для случая, когда входной сигнал представляет собой аддитивную смесь детерминированного сигнала и помехи в виде белого шума. В более общем случае помеха может иметь более сложные статистические характеристики и не являться белым шумом.

Допустим, что шумовой процесс на входе системы имеет известную спектральную плотность мощности $S_n(\omega)$.

При рассмотрении вопросов прохождения случайного сигнала через линейную цепь было получено уравнение, связывающее спектральные плотности мощности входного и выходного сигнала линейной системы с известной передаточной функцией. В случае, когда входным сигналом линейной системы является белый шум, это уравнение принимает простой вид

$$S_x(\omega) = S_N |W(j\omega)|^2,$$

где $S_N = \text{const}$.

Если использовать линейную цепь для обратной цели, то есть для формирования белого шума из случайного процесса с известной спектральной плотностью мощности, то для передаточной функции искомой линейной системы легко записать

$$|W_N(j\omega)|^2 = \frac{S_N}{S_x(\omega)}. \quad (8.11)$$

Применяя линейный фильтр с характеристиками (8.11) можно, как говорят "обелить" имеющийся случайный сигнал.

После прохождения отбеливающего фильтра $W_N(j\omega)$ спектральная плотность шума входного сигнала примет вид

$$S_{1N}(\omega) = S_n(\omega) W_N(j\omega).$$

Поскольку входной сигнал представляет собой сумму полезного сигнала и шума

$$u_{\text{вх}}(t) = s_1(t) + n_1(t),$$

то в процессе отбеливания шумовой составляющей произойдет искажение полезного сигнала, вызванное его прохождением через тот же отбеливающий фильтр. В результате на выходе фильтра получается сигнал

$$u_{\text{ф}}(t) = s'_1(t) + n_N(t),$$

где $n_N(t)$ – отбеленный вариант входного шума,

$s'_1(t)$ – результат искажения полезного сигнала, вызванный применением отбеливающего фильтра.

Теперь к сигналу $u_{\text{ф}}(t)$ может быть применена рассмотренная ранее процедура расчета согласованного фильтра (8.7), так как этот сигнал представляет собой сумму некоторого полезного детерминированного сигнала и белого шума.

Тогда для частотной передаточной функции согласованного (для $u_{\text{ф}}(t)$) фильтра получим

$$W_{\text{сф}}(j\omega) = AS_1^*(\omega) e^{-j\omega t_{0N}} = AS_1^*(\omega) W_N^*(j\omega) e^{-j\omega t_{0N}},$$

где t_{0N} – длительность сигнала $s'_1(t)$ на выходе отбеливающего фильтра.

Итоговая передаточная функция описанной системы преобразований может быть записана в виде

$$\begin{aligned} W(j\omega) &= W_N(j\omega) W_{\text{сф}}(j\omega) = W_N(j\omega) AS_1^*(\omega) W_N^*(j\omega) e^{-j\omega t_{0N}} = \\ &= AS_1^*(\omega) |W_N(j\omega)|^2 e^{-j\omega t_{0N}} = AS_N \frac{S_1^*(\omega)}{S_n(\omega)} e^{-j\omega t_{0N}} \end{aligned} \quad (8.12)$$

Структурная схема, соответствующая описанному преобразованию сигналов и (8.12) представлена на рис.8.3.

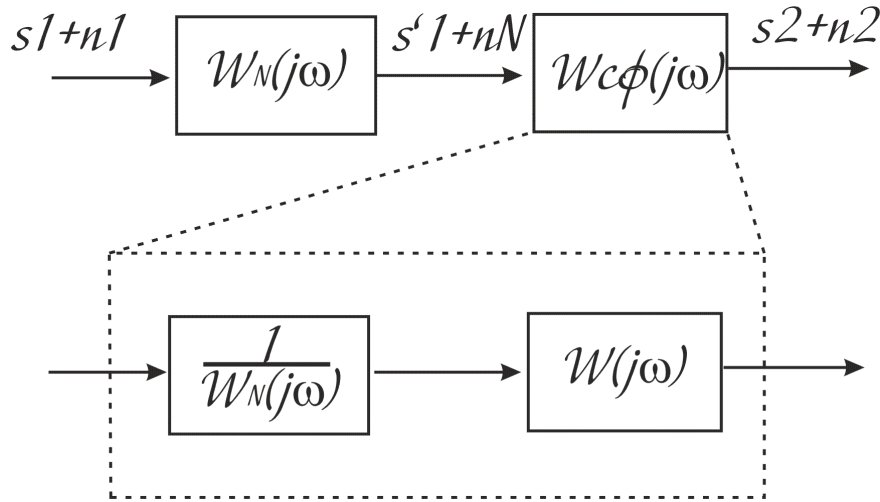


Рис.8.3. Структурная схема согласованной фильтрации с отбеливающим фильтром

Сформированный указанным выше способом согласованный фильтр является оптимальным для сигнала

$$u_{\phi}(t) = s'_1(t) + n_N(t),$$

что, вообще говоря, не означает его оптимальность по отношению к входному сигналу

$$u_{\text{вх}}(t) = s_1(t) + n_1(t).$$

Доказательство данного факта можно провести методом от противного. Допустим, что существует фильтр с частотной передаточной функцией (коэффициентом передачи) $W_{\text{и}}(j\omega)$, которая для сигнала $u_{\text{вх}}(t)$ обеспечивает лучшее соотношение сигнал-шум в выходном сигнале.

Представим данную передаточную функцию в виде

$$W_{\text{и}}(j\omega) = W_N(j\omega) \frac{W_{\text{и}}(j\omega)}{W_N(j\omega)},$$

где $W_N(j\omega)$ – передаточная функция отбеливающего фильтра.

В соответствии со сделанным предположением отношение SNR на выходе этого идеального фильтра должно быть больше, чем у полученного ранее фильтра

$$W_{\text{сф}}(j\omega) = \frac{W(j\omega)}{W_N(j\omega)},$$

то есть меньше, чем у согласованного фильтра, что невозможно. Следовательно, полученный ранее фильтр является оптимальным для обнаружения детерминированного сигнала на фоне небелого шума.

Используя полученную в предыдущем разделе формулу (8.10) для отношения сигнал-шум, запишем

$$SNR_2 = 20 \lg \sqrt{\frac{E'_1}{S_N}},$$

где E'_1 – энергия сигнала $s'_1(t)$,

S_N – спектральная плотность отбеленного шума $n_N(t)$.

Энергию сигнала $s'_1(t)$, как результата прохождения исходного сигнала через отбеливающий фильтр можно вычислить с использованием равенства Парсеваля

$$E'_1 = \frac{1}{2\pi} \int_{-\infty}^{\infty} |S_1(\omega)|^2 |W_N(j\omega)|^2 d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{S_N}{S_n(\omega)} |S_1(\omega)|^2 d\omega.$$

Откуда окончательно можно записать

$$SNR_2 = 20 \lg \sqrt{\frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{|S_1(\omega)|^2}{S_n(\omega)} d\omega}. \quad (8.13)$$

В случае, когда шум входного сигнала является белым, полученное уравнение (8.13) упрощается и повторяет ранее полученный результат (8.10) для согласованной фильтрации детерминированного сигнала на фоне белого шума.

8.4. Квазиоптимальная фильтрация детерминированных сигналов

Согласованные фильтры являются оптимальными с точки зрения обеспечения максимума отношения сигнал-шум, однако их реализация зачастую либо приводит к слишком громоздким схемам, либо сталкивается с проблемой физической или конструктивной реализации. В то же время в ряде случаев можно существенно упростить схему фильтра при незначительном ухудшении показателя отношения сигнал-шум по сравнению с оптимальным фильтром. Подобные фильтры называют квазиоптимальными.

Ранее нами был рассмотрен пример создания оптимального фильтра для сигнала, представляющего собой прямоугольный импульс. Схема реализации данного фильтра приведена на рис. 8.2. Попробуем вместо такого оптимального фильтра использовать простую RC-цепь, подобрав соответствующим образом ее постоянную времени $\tau = RC$. Схема RC-цепи и результат прохождения через нее прямоугольного импульсного сигнала представлен на рис.8.4.

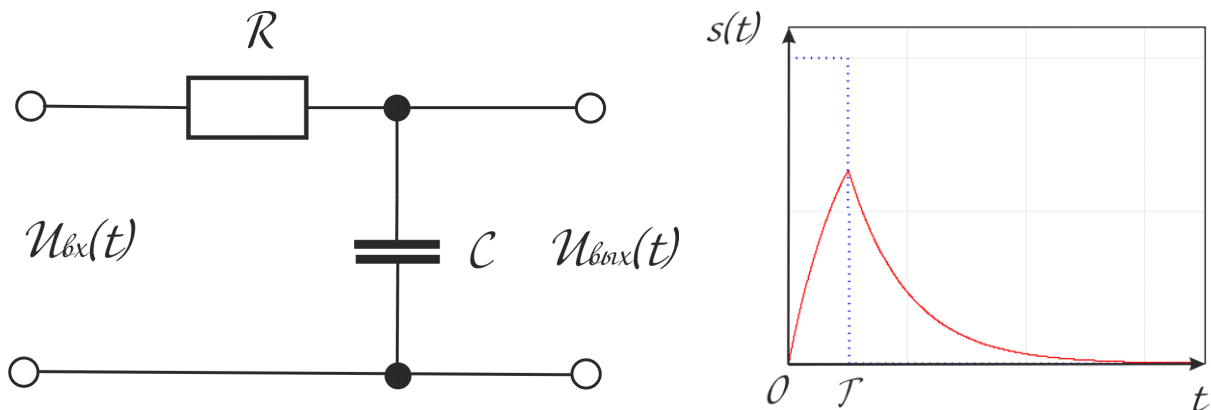


Рис.8.4. RC-цепь и результат прохождения через нее прямоугольного импульса

В соответствии с ранее рассмотренным материалом передаточная функция данной цепи имеет вид

$$W(s) = \frac{1}{1 + sRC} = \frac{1}{1 + \tau s},$$

а импульсная и переходная характеристики соответственно определяются выражениями

$$g(t) = \begin{cases} \frac{1}{\tau} e^{-t/\tau}, & t \geq 0, \\ 0, & t < 0. \end{cases}$$

и

$$h(t) = \begin{cases} 1 - e^{-t/\tau}, & t \geq 0, \\ 0, & t < 0. \end{cases}$$

Если входной сигнал представляет собой импульс длительностью T , то из характеристик цепи ясно, что максимальное значение выходного сигнала достигается в момент времени $t = T$ и может быть выражено уравнением

$$|s_2(t)|_{max} = A \left(1 - e^{-\frac{T}{\tau}} \right),$$

где $A = const$ - амплитуда прямоугольного импульса.

При подаче на вход рассматриваемой цепи сигнала, являющегося белым шумом с равномерной спектральной плотностью мощности S_N , на выходе будет наблюдаться сигнал со спектральной плотностью мощности соответствующей уравнению

$$S_{n2}(\omega) = S_N |W(j\omega)|^2 = S_N \left| \frac{1}{1 + j\omega\tau} \right|^2 = \frac{S_N}{1 + \omega^2\tau^2}.$$

Дисперсия шумовой составляющей выходного сигнала может быть определена по теореме Винера-Хинчина

$$\sigma_{вых}^2 = \sigma_{n2}^2 = R_{n2}(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{n2}(\omega) d\omega = \frac{S_N}{2\pi} \int_{-\infty}^{\infty} \frac{1}{1 + \omega^2\tau^2} d\omega.$$

Воспользуемся равенством $\int \frac{1}{1+x^2} dx = \arctg x$, произведем соответствующую замену переменных и получим

$$\sigma_{вых}^2 = \frac{S_N}{2\pi\tau} \int_{-\infty}^{\infty} \frac{1}{1+x^2} dx = \frac{S_N}{2\pi\tau} \arctg x \Big|_{-\infty}^{\infty} = \frac{S_N}{2\tau}.$$

Полученный результат позволяет записать выражение для отношения сигнал-шум

$$SNR_2 = 20 \lg \frac{A \left(1 - e^{-\frac{T}{\tau}} \right)}{\sqrt{\frac{S_N}{2\tau}}} = 20 \lg \left(\sqrt{\frac{A^2 T}{S_N}} \sqrt{\frac{2\tau}{T}} \left(1 - e^{-\frac{T}{\tau}} \right) \right). \quad (8.14)$$

Для оптимального согласованного фильтра ранее было получено

$$SNR_{2opt} = 20 \lg \sqrt{\frac{E_1}{S_N}}. \quad (8.15)$$

Сравнивая два последних уравнения (8.14) и (8.15), становится понятно, что они отличаются множителем $\sqrt{\frac{2\tau}{T}} \left(1 - e^{-\frac{T}{\tau}} \right)$, который понижает отношение сигнал-шум в квазиоптимальном фильтре по сравнению с оптимальным согласованным фильтром.

Выбор параметров цепи квазиоптимального фильтра сводится к нахождению такого значения $\tau = RC$, чтобы максимизировать отношение SNR_2 , а, следовательно, получить максимально возможное значение понижающего коэффициента $\sqrt{\frac{2\tau}{T}} \left(1 - e^{-\frac{T}{\tau}} \right)$.

Переходя к нормированной переменной $z = \tau/T$, понижающий коэффициент квазиоптимального фильтра можно представить в виде функции одной переменной

$$K_{SOPT}(z) = \sqrt{2z} \left(1 - e^{-\frac{1}{z}} \right).$$

К сожалению, аналитическое решение задачи поиска экстремума этой функции является невозможным, но вполне применимо численное решение. График функции $K_{SOPT}(z)$ приведен на рис. 8.5.

Численное решение дает соотношение $\tau \approx 0,796 T$, при котором $K_{SOPT} \approx 0,903$.

Таким образом, в случае с прямоугольным импульсом потери при переходе от оптимальной к квазиоптимальной фильтрации составляют около 10%. При более сложной форме информационного сигнала потери могут быть значительно весомее.

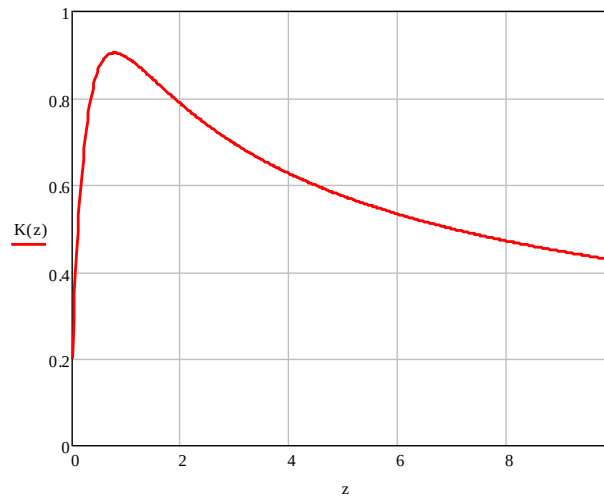


Рис. 8.5. Коэффициент снижения эффективности квазиоптимального фильтра

Пример 8.2.

На рис. 8.6. приведен результат демонстрационного моделирования работы оптимального согласованного фильтра и квазиоптимального фильтра на основе RC-цепи для обнаружения единичного импульса. Для целей увеличения наглядности одиночные импульсы представлены периодической последовательностью идентичных импульсов.

```
clear figure;
%Параметры сигнала
T=0.5; % период
Tint=5;%интервал наблюдения
tau=0.001; % интервал дискретизации при моделировании
% Параметры фильтра 1-го порядка
K=2; % Коэффициент усиления
Tl=0.1; % Постоянная времени
W=tf(K,[Tl 1]); % Передаточная функция фильтра 1-го порядка
%Генерирование исходного сигнала
[u,t]=gensig('square',T,Tint,tau);
%Генерирование аддитивного шума
SS=0.5;% сигма
Am=1;% амплитуда полезного сигнала
y=Am*u+SS*randn(size(t));
subplot(4,1,1); plot(t,u,'-r','LineWidth',2);
grid on;
axis([0 5 -1 2]);
subplot(4,1,2); plot(t,u,'-r',t,y);
grid on;
%axis([0 5 -7 7]);
y1=lsim(W,y,t);
subplot(4,1,3); plot(t,y); hold on;
plot(t,y1,'-m','LineWidth',2);
hold off;
%axis([0 5 -7 7]);
grid on;
```

```
%Формирование оптимального фильтра
A=10; %Коэффициент усиления
s=tf('s');
W2=-A*exp(-T/2*s)/s;
W1=A/s;
W3=parallel(W1,W2);
y3=lsim(W3,y,t);
subplot(4,1,4); plot(t,y1,'-m',t,y3,'-k','LineWidth',2);
grid on;
```

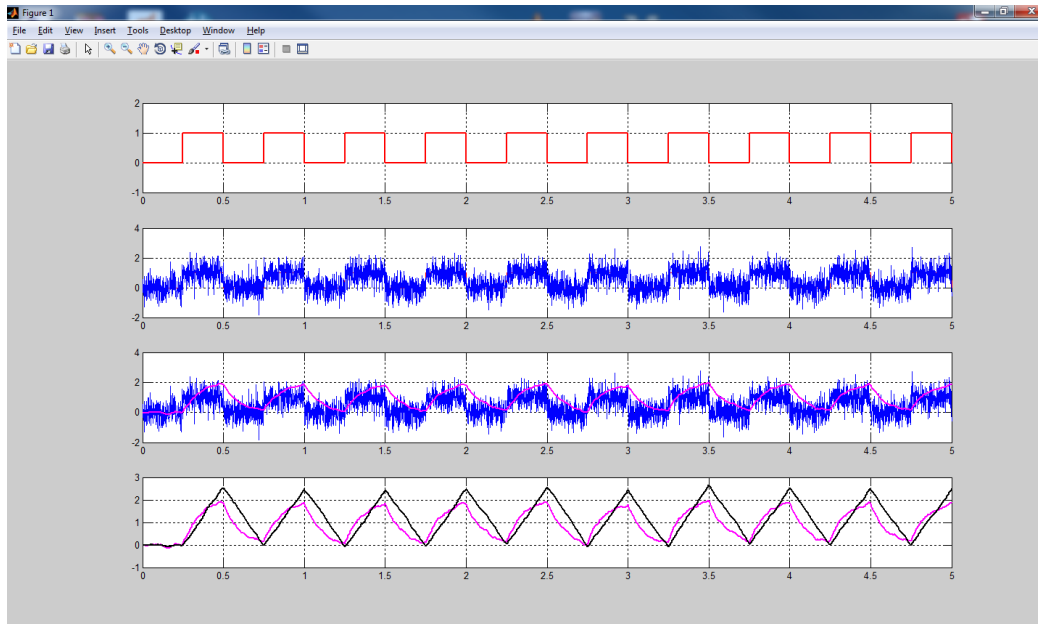


Рис.8.6. Иллюстрация работы оптимального и квазиоптимального фильтров

8.5.Оптимальная фильтрация случайных сигналов. Фильтр Винера

В рассмотренных ранее вариантах обработки сигналов предполагалось, что исходный сигнал является суммой детерминированной и случайной составляющей

$$u_{\text{вх}}(t) = s_1(t) + n_1(t).$$

В более общем случае необходимо предположить, что информационный сигнал $s_1(t)$ является не детерминированным, а тоже случайным. Для того, чтобы придать этому случаю отличие в обозначениях примем следующую модель входного сигнала

$$u_{\text{вх}}(t) = \xi(t) + n(t),$$

где $\xi(t)$ – полезный (информационный) сигнал,
 $n(t)$ – шумовая помеха.

Будем также полагать, что оба процесса $\xi(t)$ и $n(t)$ являются центрированными стационарными, нормальными и имеют известные спектральные плотности мощности $S_{\xi}(\omega)$ и $S_n(\omega)$.

Задачей обработки входного сигнала $u_{\text{вх}}(t)$ является создание такого фильтра, чтобы его выходной сигнал с максимально возможной точностью воспроизводил полезный случайный сигнал $\xi(t)$ или другими словами, чтобы ошибка воспроизведения сигнала $\xi(t)$ на выходе искомого фильтра была минимальной.

Ошибку на выходе фильтра естественно определить в виде

$$e(t) = u_{\text{вых}}(t) - \xi(t).$$

В качестве критерия оценки малости ошибки $e(t)$, которая в силу линейности фильтра, аналогично входным сигналам будет стационарной, логично принять величину дисперсии

$$D_e = \sigma_e^2 = M\{e^2(t)\} - M^2\{e\}.$$

Центрированность входных процессов и линейность используемого фильтра обеспечивают нулевое математическое выходного сигнала, поэтому выражение для дисперсии можно упростить

$$\sigma_e^2 = M\{e^2(t)\} = R_e(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_e(\omega) d\omega,$$

где $R_e(\tau)$ – корреляционная функция сигнала $e(t)$,

$S_e(\omega)$ – спектральная плотность мощности сигнала $e(t)$.

Поиск требуемого фильтра направлен на минимизацию дисперсии

$$\sigma_e^2 \rightarrow \min, \quad (8.16)$$

что в геометрической интерпретации означает минимизацию площади под кривой спектральной плотности мощности $S_e(\omega)$.

Структурная схема, поясняющая процесс формирования сигнала ошибки приведена на рис.8.7

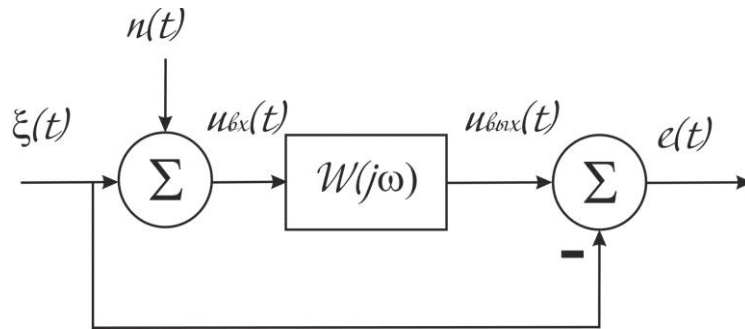


Рис.8.7. Структурная схема формирования ошибки оптимальной фильтрации

В общем виде задача оптимальной линейной фильтрации в смысле (8.16) была решена Н.Винером для коррелированных стационарных случайных процессов. Импульсная характеристика оптимального линейного фильтра $g(\beta)$, удовлетворяющая условию физической реализуемости, в соответствии с решением Н.Винера должна удовлетворять интегральному уравнению Винера – Хопфа

$$\int_{-\infty}^{\infty} g(\beta) R_u(\tau - \beta) d\beta - R_{\xi u}(\tau) = 0, \quad (8.17)$$

где $R_u(\tau) = R_{\xi}(\tau) + R_n(\tau) + R_{\xi n}(\tau) + R_{n\xi}(\tau)$ – корреляционная функция суммарного входного сигнала.

Решение уравнения (8.17) не входит в задачи нашего курса, поэтому мы ограничимся качественным рассмотрением простейшего случая отсутствия статистической связи между входными сигналами.

Поскольку искомый фильтр $W(j\omega)$ предполагается линейным, а процессы $\xi(t)$ и $n(t)$ статистически независимыми, то спектральную плотность мощности сигнала ошибки $e(t)$ можно представить как спектр суммы двух выходных сигналов фильтра

$$S_e(\omega) = S'_{\xi}(\omega) + S'_n(\omega),$$

где $S'_{\xi}(\omega) = S_{\xi}(\omega) |W(j\omega) - 1|^2$,

$$S'_n(\omega) = S_n(\omega)|W(j\omega)|^2.$$

Разница в коэффициентах передачи для сигналов $\xi(t)$ и $n(t)$ определяется способом их включения в схему формирования ошибки $e(t)$ и легко определяется из приведенного рисунка.

После раскрытия скобок и суммирования полученных выражений получаем

$$\begin{aligned} S_e(\omega) &= S_\xi(\omega)(W(j\omega) - 1)(W^*(j\omega) - 1) + S_n(\omega)|W(j\omega)|^2 = \\ &= S_\xi(\omega)\{W(j\omega)W^*(j\omega) - W(j\omega) - W^*(j\omega) + 1\} + S_n(\omega)|W(j\omega)|^2 \end{aligned}$$

Представим искомую передаточную функцию в виде

$$W(j\omega) = W(\omega)e^{j\varphi(\omega)},$$

где $W(\omega) = |W(j\omega)|$,
получаем

$$S_e(\omega) = S_\xi(\omega)\{W^2(\omega) - 2W(\omega)\cos\varphi(\omega) + 1\} + S_n(\omega)W^2(\omega).$$

Из полученного уравнения видно, что для минимизации $S_e(\omega)$ слагаемое $2W(\omega)\cos\varphi(\omega)$ должно быть максимально. Это условие выполняется только при $\cos\varphi(\omega) = 1$, то есть $\varphi(\omega) = 2\pi k, k = 0, 1, 2, \dots$ Следовательно, искомый фильтр не должен вносить фазовой задержки сигнал.

Подстановка $\cos\varphi(\omega) = 1$ в последнее уравнение приводит к его упрощению

$$S_e(\omega) = S_\xi(\omega)(W(\omega) - 1)^2 + S_n(\omega)W^2(\omega).$$

Чтобы найти уравнение для искомой передаточной функции, продифференцируем по ней полученное выражение (при условии $\omega = \text{const}$) и приравняем полученную производную нулю

$$2S_\xi(\omega)(W(\omega) - 1) + 2S_n(\omega)W(\omega) = 0.$$

Откуда находим

$$W_{\text{опт}}(\omega) = \frac{S_\xi(\omega)}{S_\xi(\omega) + S_n(\omega)}. \quad (8.18)$$

Фильтр с характеристикой $W_{\text{опт}}(\omega)$, соответствующей полученному выражению (8.18), обеспечивает минимальную спектральную плотность сигнала ошибки на всех частотах ω , что в свою очередь приводит к минимизации дисперсии сигнала ошибки

$$\sigma_e^2|_{W_{\text{опт}}(\omega)} \rightarrow \min.$$

Фильтр с коэффициентом передачи $W_{\text{опт}}(\omega)$ называют оптимальным линейным фильтром для случайного сигнала или фильтром Винера.

Задержку во времени данный фильтр по указанным выше причинам не производит, поэтому фазовые соотношения входных сигналов не изменяются.

Качественный анализ полученного уравнения для коэффициента передачи оптимального фильтра показывает, что в областях частот, где

$$\begin{aligned} S_\xi(\omega) \gg S_n(\omega) &\rightarrow S_\xi(\omega) + S_n(\omega) \approx S_\xi(\omega) \rightarrow W_{\text{опт}}(\omega) \approx 1, \\ S_\xi(\omega) \ll S_n(\omega) &\rightarrow S_\xi(\omega) + S_n(\omega) \approx S_n(\omega) \rightarrow W_{\text{опт}}(\omega) \approx 0, \\ S_\xi(\omega) \approx S_n(\omega) &\rightarrow S_\xi(\omega) + S_n(\omega) \approx 2S_\xi(\omega) \rightarrow W_{\text{опт}}(\omega) \approx 1/2. \end{aligned}$$

То есть данный фильтр пропускает те частоты, на которых преобладает полезный информационный сигнал и задерживает частоты, где сильнее преобладание шума. Понятно, что чем ближе друг к другу будут спектральные плотности полезного сигнала и шума, тем меньше будет эффект фильтрации.

К сожалению, во многих практических задачах использование полученной формулы для оптимального фильтра приводит к физически нереализуемым результатам, требующим бесконечной протяженности импульсной характеристики фильтра в обе стороны.

Для того чтобы реализовать передаточную функцию, наиболее близкую к оптимальной из $W_{\text{опт}}(\omega)$, необходимо выделить физически реализуемую часть с полюсами, находящимися в верхней полуплоскости корней, а остальные члены отбросить. Данная методика была предложена Г.Бодом и К.Шенноном. В соответствии с ней знаменатель выражения для оптимального фильтра подвергается операции, получившей название "факторизация"

$$S_{\xi}(\omega) + S_n(\omega) = |\Psi(j\omega)|^2 = \Psi(j\omega)\Psi(-j\omega), \quad (8.18)$$

где $\Psi(j\omega)$ – некоторая функция, все полюсы и нули которой находятся в верхней комплексной полуплоскости,

$\Psi(-j\omega)$ – функция комплексно сопряженная к $\Psi(j\omega)$, все полюсы и нули которой, соответственно, расположены в нижней полуплоскости.

Затем производится разделение $W_{\text{опт}}(\omega)$ на реализуемые и нереализуемые слагаемые

$$W_{\text{опт}}(\omega) = \frac{1}{\Psi(j\omega)} \left[\frac{S_{\xi}(\omega)}{\Psi(-j\omega)} \right]_+ + \frac{1}{\Psi(j\omega)} \left[\frac{S_{\xi}(\omega)}{\Psi(-j\omega)} \right]_-$$

где реализуемая часть отмечена знаком "+", а нереализуемая знаком "-".

Отбросив члены, соответствующие нереализуемой части получают условие оптимальности с учетом физической реализуемости

$$W_{\text{опт}}(\omega) = \frac{1}{\Psi(j\omega)} \left[\frac{S_{\xi}(\omega)}{\Psi(-j\omega)} \right]_+ \quad (8.19)$$

Для практики особенно важным и распространенным случаем является ситуация, когда помеха представляет собой белый шум со спектральной плотностью

$$S_n(\omega) = S_N = \text{const},$$

а спектральная плотность мощности полезного сигнала описывается дробно-рациональной функцией

$$S_{\xi}(\omega) = \frac{G_1(\omega^2)}{G_2(\omega^2)},$$

где порядок $G_1(\omega^2)$ превышает порядок $G_2(\omega^2)$.

В данном случае искомая передаточная функция определяется по простой формуле

$$W_{\text{опт}}(j\omega) = 1 - \frac{\sqrt{S_N}}{\Psi(j\omega)}$$

Пример.8.3.

Рассмотрим случай, когда входной сигнал представляет собой сумму полезного случайного стационарного центрированного сигнала $\xi(t)$ со спектральной плотностью мощности

$$S_{\xi}(\omega) = \frac{S_{\xi}(0)}{1 + \omega^2 T^2}$$

и нормального белого шума $n(t)$ с постоянной спектральной плотностью мощности $S_n(\omega) = S_N = \text{const}$.

Требуется найти передаточную функцию фильтра $W_{\text{опт}}(j\omega)$, обеспечивающего минимум дисперсии ошибки $e(t) = u_{\text{вых}}(t) - \xi(t)$.

Вычислим суммарную спектральную плотность входного сигнала

$$S_u(\omega) = S_N + \frac{S_{\xi}(0)}{1 + \omega^2 T^2} = \frac{S_N(1 + \omega^2 T^2) + S_{\xi}(0)}{1 + \omega^2 T^2}$$

и произведем ее "факторизацию" по (8.18)

$$S_u(\omega) = \Psi(j\omega)\Psi(-j\omega).$$

Для этого проведем преобразование

$$\frac{S_N(1 + \omega^2 T^2) + S_\xi(0)}{1 + \omega^2 T^2} = |S_N + S_\xi(0)| \frac{(1 + j\omega\alpha)(1 + j\omega\beta)}{(1 + j\omega T)(1 - j\omega T)},$$

где $\alpha = \frac{T}{\sqrt{1+\gamma}}$, $\gamma = \frac{S_\xi(0)}{S_N}$

Тогда

$$\begin{aligned}\Psi(j\omega) &= \sqrt{S_N + S_\xi(0)} \frac{(1 + j\omega\alpha)}{(1 + j\omega T)} \\ \Psi(-j\omega) &= \sqrt{S_N + S_\xi(0)} \frac{(1 - j\omega\alpha)}{(1 - j\omega T)}.\end{aligned}$$

Теперь для искомой передаточной функции можно написать

$$\begin{aligned}W_{\text{опт}}(j\omega) &= 1 - \frac{\sqrt{S_N}}{\Psi(j\omega)} = 1 - \frac{\sqrt{S_N}}{\sqrt{S_N + S_\xi(0)}} \frac{(1 + j\omega T)}{(1 + j\omega\alpha)} = \\ &= \frac{\sqrt{S_N + S_\xi(0)} - \sqrt{S_N}}{(1 + j\omega\alpha)\sqrt{S_N + S_\xi(0)}} + j\omega \frac{\alpha\sqrt{S_N + S_\xi(0)} - T\sqrt{S_N}}{(1 + j\omega\alpha)\sqrt{S_N + S_\xi(0)}}.\end{aligned}$$

Так как

$$\alpha = \frac{T}{\sqrt{1+\gamma}} = \frac{T\sqrt{S_N}}{\sqrt{S_N + S_\xi(0)}}$$

второе слагаемое в уравнении искомой передаточной функции равно нулю, то есть

$$W_{\text{опт}}(j\omega) = \frac{\sqrt{S_N + S_\xi(0)} - \sqrt{S_N}}{(1 + j\omega\alpha)\sqrt{S_N + S_\xi(0)}}$$

или

$$W_{\text{опт}}(j\omega) = \left(1 - \frac{\sqrt{S_N}}{\sqrt{S_N + S_\xi(0)}}\right) \frac{1}{(1 + j\omega\alpha)} = \left(1 - \frac{1}{\sqrt{1+\gamma}}\right) \frac{1}{(1 + j\omega\alpha)}.$$

8.6. Общие сведения о фильтре Калмана-Бьюси

Основные результаты Н. Винера были получены для случая линейных систем и стационарных центрированных случайных процессов и рассматривались на основе анализа в частотной области.

В 1960 году Р. Калман и Р. Бьюси рассмотрели задачу линейной фильтрации во временной области и на основе концепции пространства состояний предложили новый и эффективный метод синтеза оптимальных фильтров и регуляторов по критерию

минимума математического ожидания квадрата случайной ошибки, применимый как к стационарным, так и нестационарным системам.

В настоящее время фильтр Калмана чаще всего реализуется в цифровом виде, поэтому к его рассмотрению мы вернемся в разделе курса, посвященном цифровой обработке сигналов, а сейчас ограничимся лишь самыми общими замечаниями для простейшего скалярного случая.

Допустим, что создаваемая система должна воспроизводить некоторый сигнал $z(t)$, который в общем случае является нестационарным случайным процессом. Входной сигнал системы представляет собой сумму этого сигнала и некоторой помехи, которая также может быть нестационарной

$$u(t) = z(t) + f(t). \quad (8.20)$$

В рамках предложенной Калманом теории математическая модель формирования сигнала $Z(t)$ представляется в виде переменных состояния

$$\frac{d}{dt}z(t) = A(t)z(t) + v(t), \quad (8.21)$$

где $A(t)$ – некоторая функция времени,

$v(t)$ – нестационарный случайный процесс типа белого шума.

Такой способ описания случайных сигналов (8.20), (8.21) получил название метода формирующего фильтра, который по своей сути уже рассматривался в нашем курсе применительно к задаче формирования стационарного случайного сигнала с заданными статистическими характеристиками из белого шума.

Корреляционные функции процессов, участвующих в формировании входного сигнала следующие

$$\begin{aligned} R_{vv}(t, \tau) &= L(t)\delta(t - \tau), \\ R_{ff}(t, \tau) &= N(t)\delta(t - \tau), \\ R_{vf}(t, \tau) &= 0, \end{aligned}$$

где $L(t), N(t)$ – непрерывные и непрерывно дифференцируемые неотрицательные функции.

Для стационарных процессов

$$\begin{aligned} R_{vv}(\tau) &= L\delta(t - \tau), \\ R_{ff}(\tau) &= N\delta(t - \tau), \\ R_{vf}(t, \tau) &= 0. \end{aligned}$$

Если на выходе системы наблюдается случайный процесс $X(t)$, то ошибка системы может быть охарактеризована разностью

$$e(t) = z(t) - x(t).$$

Калман и Бьюси рассмотрели решение задачи создания фильтра, обеспечивающего минимальное значение математического ожидания квадрата данной ошибки, то есть

$$M\{e(t)^2\} = M\{(z(t) - x(t))^2\} \rightarrow \min. \quad (8.21)$$

В результате решения задачи (8.21) было показано, что оптимальная система должна описываться дифференциальным уравнением

$$\frac{d}{dt}x(t) = Q(t)x(t) + G(t)u(t), \quad (8.22)$$

где $Q(t), G(t)$ – определяются исходя из свойств входного сигнала и помехи.

Первым шагом в нахождении функций $Q(t), G(t)$ из (8.22) является поиск вспомогательной функции

$$r(t) = M\{e(t)^2\},$$

которая удовлетворяет уравнению Риккати

$$\frac{d}{dt}r(t) = L(t) + 2A(t)r(t) - \frac{r^2(t)}{N(t)}.$$

Для решения этого уравнения требуется начальное значение $r(t_0)$. Обычно полагают $X(t_0) = 0$, что дает

$$e(t_0) = z(t_0) - x(t_0) = z(t_0)$$

и

$$r(t_0) = M\{e(t_0)^2\} = M\{z(t_0)^2\} = R_{zz}(t_0 t_0).$$

После нахождения $r(t)$ определяются

$$G(t) = \frac{r(t)}{N(t)}$$

и

$$Q(t) = A(t) - G(t).$$

В результирующем виде уравнение оптимального фильтра можно записать как

$$\frac{d}{dt}x(t) = A(t)x(t) + G(t)\{u(t) - x(t)\}.$$

Для стационарных процессов

$$A(t) = A$$

и в установившихся режимах

$$G(t) = G, Q(t) = Q = A - G.$$

В стационарном случае фильтр также является стационарным и удовлетворяет уравнению

$$\frac{d}{dt}x(t) = (A - G)x(t) + Gu(t). \quad (8.23)$$

Естественно, что для стационарного случая результаты Калмана и Бьюси (8.23) совпадают с методом Н.Винера и полученное уравнение эквивалентно оптимальному фильтру Винера (8.18), определяемому в частотной области.

Структурная схема формирования случайного сигнала и оптимального фильтра для стационарного случая представлена на рис.8.8.

Полученные результаты легко обобщаются на случай векторных случайных сигналов.

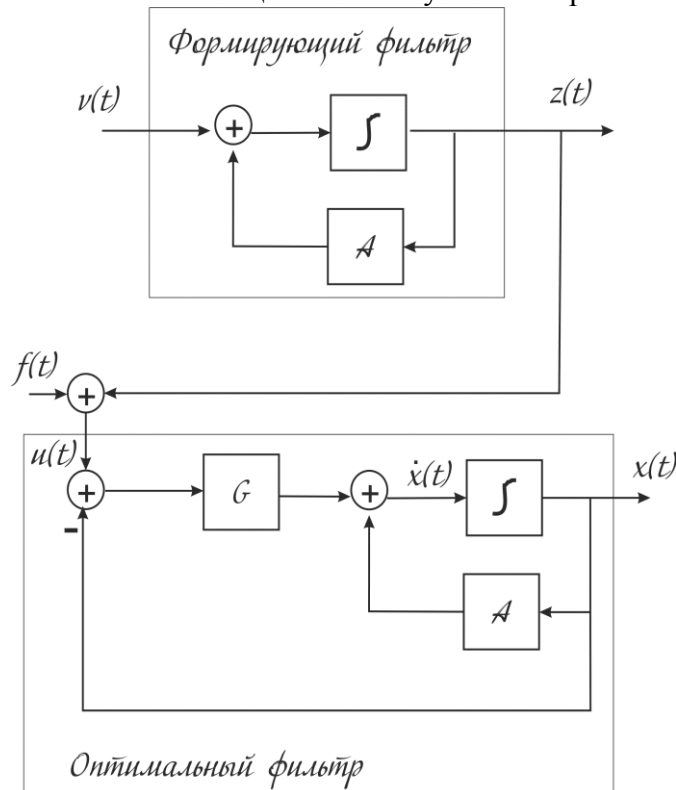


Рис.8.8. Оптимальный фильтр Калмана-Бьюси

8.7. Корреляторы сигналов

При решении задачи обнаружения сигнала приемная система должна содержать два основных блока: согласованный фильтр и компаратор, обеспечивающий сравнение выходного сигнала фильтра с заданным пороговым значением. Упрощенная схема такого обнаружителя сигнала приведена на рис.8.9.

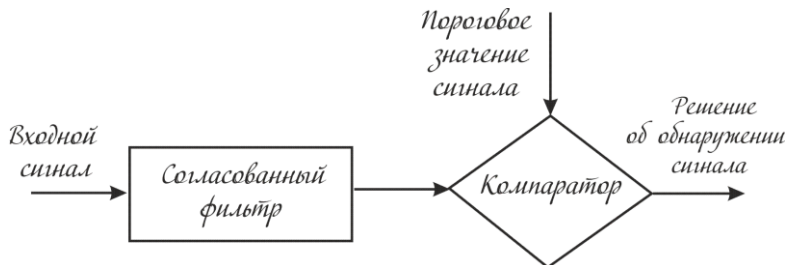


Рис.8.9. Упрощенная схема обнаружителя сигнала

Усложнение формы используемого сигнала приводит к усложнению структуры согласованного с ним фильтра. Для упрощения технической реализации обнаружителя сигнала вместо согласованного фильтра часто используются устройства, называемые

корреляторами.

Коррелятор включает в себя умножитель, на один вход которого подается принимаемый сигнал, а на второй – специально сформированный в приемном устройстве **опорный** сигнал.

Принимаемый сигнал предполагается состоящим из полезного сигнала и помехи

$$u_{\text{вх}}(t) = s(t) + n(t),$$

а опорный сигнал формируется в виде

$$u_0(t) = s(t - \tau).$$

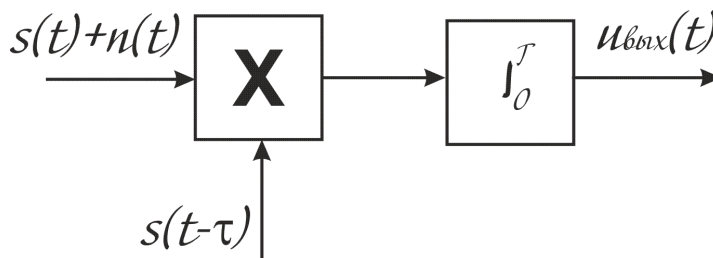


Рис.8.10. Упрощенная схема коррелятора.

Таким образом, опорный сигнал по форме полностью повторяет полезный сигнал, но отстает от него на интервал времени τ .

Далее сигнал, формируемый умножителем, подвергается интегрированию, как показано на схеме,

изображенной на рис.8.10. После проведения интегрирования получаем сигнал

$$u_{\text{вых}}(t) = A \int_0^T (s(t) + n(t)) s(t - \tau) dt = AB_s(\tau) + A \int_0^T s(t - \tau) n(t) dt, \quad (8.24)$$

где A – некоторая константа.

Если интервал интегрирования в (8.24) выбрать равным длительности сигнала $s(t)$, то при $\tau = 0$ получаем

$$u_{\text{вых}}(t) = AE + A \int_0^T s(t) n(t) dt,$$

что совпадает с сигналом согласованного фильтра в момент окончания длительности полезного сигнала.

Следовательно, основная проблема при практическом использовании коррелятора состоит в необходимости обеспечения условия $\tau = 0$, то есть отсутствие сдвига между принимаемым и опорным сигналами. Для этого необходимо точно знать время прихода полезного сигнала, что противоречит сути самой задачи его обнаружения.

Для преодоления этой трудности на практике используют набор корреляторов с разными значениями задержки $\tau_k, k = 1, 2, 3 \dots, n$. Затем производится сравнение полученных результатов, из которых выбирается тот, который имеет наибольшее значение. Именно этот наибольший сигнал далее используется в компараторе для сравнения с заданным пороговым значением.

На практике реализация схем с корреляторами часто оказывается проще создания согласованного фильтра.

8.8. Вопросы для самоконтроля и повторения

1. Сформулируйте задачу обнаружения сигнала на фоне помех.
2. Сформулируйте задачу фильтрации полезного информационного сигнала, наблюдаемого на фоне помех.
3. Сформулируйте критерий нахождения параметров оптимального линейного фильтра для детерминированного сигнала, наблюдаемого на фоне помех.
4. Получите уравнение для импульсной характеристики оптимального линейного фильтра.
5. Какую передаточную функцию должен иметь оптимальный линейный фильтр?
6. Изобразите структурную схему согласованного фильтра для одиночного импульса.
7. Каким образом проводится оптимальная фильтрация при не белом шуме?
8. В чем состоит принцип квазиоптимальной фильтрации детерминированных сигналов?
9. Поясните работу фильтра Винера.
10. Поясните разницу между фильтром Винера и фильтром Калмана.
11. Что представляют собой корреляторы сигналов?

ГЛАВА 9. Основы цифровой обработки сигналов

9.1. Квантование аналогового сигнала

Развитие компьютерных технологий и микропроцессорной техники привело к тому, что электронные вычислительные устройства, первоначально преимущественно использовавшиеся для моделирования различных процессов, стали все чаще использоваться в управляющих и информационно-измерительных системах для обработки сигналов в реальном масштабе времени.

В большинстве цифровых систем возникает необходимость обработки как аналоговых, так и цифровых сигналов, поэтому практически все они используют преобразования сигналов, позволяющие обеспечить взаимодействие аналоговых и цифровых элементов систем.

К числу основных преобразовательных операций можно отнести аналогово-цифровое и цифро-аналоговое преобразование сигналов.

Аналогово-цифровое преобразование, выполняемое с помощью устройства, называемого аналого-цифровой преобразователь (АЦП), состоит в преобразовании аналогового сигнала в цифровой код. АЦП является необходимым интерфейсным элементом любой аналоговой подсистемы, выходные сигналы которой предполагается обрабатывать в цифровом виде.

Цифро-аналоговый преобразователь осуществляет операцию декодирования цифровых данных, формируя на своем выходе аналоговый сигнал в виде тока или напряжения. ЦАП, иногда также называемый декодером, является необходимым интерфейсным средством между цифровым каналом и аналоговым устройством.

Современные вычислительные машины и процессоры построены на основе двоичной системы счисления. Поэтому цифровой сигнал в них представляется как последовательность нулей и единиц. Каждый разряд двоичного числа называют битом, а объединение 8 бит носит название байта. Несколько байтов могут быть объединены в машинное слово. Слова могут быть любой длины от 4 до 128 бит и более в зависимости от задач решаемых конкретной вычислительной системой. Точность вычислительного устройства определяется длиной слов, которыми это устройство оперирует.

При изучении аналогово-цифровых и цифроаналоговых преобразований важно правильно понимать вес каждого разряда машинного слова. На практике АЦП и ЦАП базируются на двоичном представлении чисел с использованием двоичного кода дробного числа. В этом случае вес каждого разряда определяется как доля от максимального значения числа, представляемого машинным словом выбранной длины.

Так, например, если двоичное слово имеет n разрядов, то вес младшего разряда определяется как 2^{-n} . Вне зависимости от того используется ли код целого числа или дробного n -разрядное слово определяет 2^n различных состояний с разрешением $\frac{1}{2^n}$.

Разница между соседними значениями цифрового отсчета часто называют квантом, величина которого определяется как количеством разрядов используемых при квантовании, так и максимальным значением цифрового отсчета. Если максимальное значение равно единице, то квант определяется как

$$\Delta = \frac{1}{2^n}.$$

Заметим, что цифровой код в абсолютном значении не соответствует своему аналоговому сигналу, а только используется для его представления в виде, удобном для цифровой обработки. Повышение длины машинного слова приводит к снижению веса младшего разряда в машинном слове, что повышает разрешающую способность такого представления.

В любом случае на практике количество разрядов машинного слова конечно, поэтому любое аналогово-цифровое преобразование может обеспечить только конечное разрешение. Одной из главных операций при аналогово-цифровом преобразовании сигналов является процедура квантования сигнала по уровню. Поскольку цифровой выход может иметь только конечное число состояний, то аналоговое значение сигнала должно быть проквантовано или округлено до ближайшего возможного цифрового значения. В результате график определяющий соответствие исходного аналогового сигнала и его представления машинным словом будет иметь ступенчатый характер, как это показано на рис.9.1.

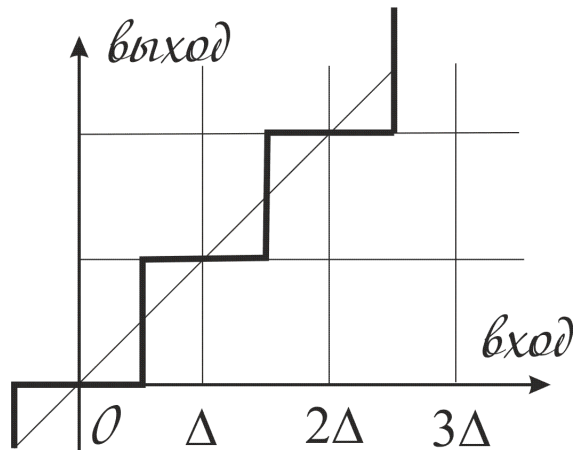


Рис.9.1. Характеристика квантователя

Обычно перед прохождением операции квантования аналоговый сигнал проходит операцию дискретизации во времени, рассмотренную нами ранее. В результате квантованию по уровню подвергается не сам аналоговый сигнал, а его выборочные отсчеты, взятые с заданным интервалом времени T_0 . Данный процесс схематично представлен на рис.9.2.

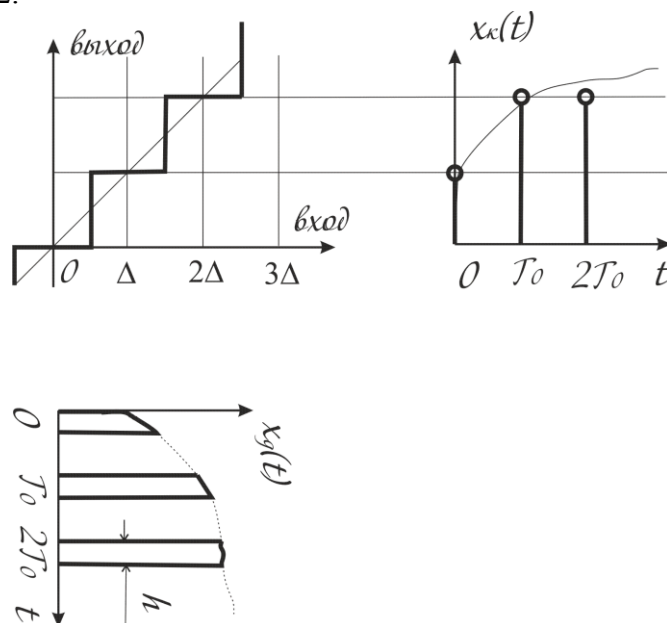


Рис.9.2. Объединение процессов дискретизации и квантования аналогового сигнала

Разница между значениями цифрового отсчета квантователя $x_k(t)$ и дискретного отсчета $x_d(t)$ представляет собой ошибку квантования

$$e(t) = x_d(t) - x_k(t),$$

часто называемую шумом квантования.

Исходя из способа получения квантованного сигнала, понятно, что величина ошибки квантования лежит в диапазоне

$$-\frac{\Delta}{2} \leq e(t) \leq \frac{\Delta}{2}.$$

В большинстве случаев ошибку квантования $e(t)$ можно считать случайным процессом с равномерным распределением плотности вероятности в указанных выше пределах. В этом случае математическое ожидание процесса $e(t)$ равняется нулю, а дисперсия равна $\Delta^2/12$.

При переходе к дискретному времени шум квантования становится дискретным случайным процессом, причем в большинстве случаев его значения можно полагать некоррелированными между собой.

Принципиальная возможность уменьшения шумов квантования до сколь угодно малого уровня путем увеличения разрядности машинного слова позволяет для аналитических целей идеализировать квантователь и существенно упростить процедуру анализа.

Дальнейшее рассмотрение мы будем проводить для идеализированного квантователя и идеального цифрового фильтра, удовлетворяющих следующим допущениям:

- выходной сигнал преобразователя считается идеализированной цифровой последовательностью, абсолютно точно воспроизводящей уровни дискретных выборочных значений аналогового сигнала;
- цифровой фильтр (цифровой вычислитель) является идеализированным устройством, которое абсолютно точно выполняет все математические операции, предусмотренные алгоритмом его действия;
- шумы квантования и другие искажающие эффекты, связанные с конечной разрядностью используемых цифровых устройств рассматриваются отдельно.

9.2. Разностные уравнения цифрового фильтра. Понятие БИХ и КИХ фильтров

Для широкого класса цифровой техники выбор алгоритма работы базируется на основе заданных аналоговых прототипов. Поэтому при проектировании цифровых фильтров, часто используется понятие фильтра-прототипа, то есть аналогового фильтра, который реализует тот же алгоритм обработки сигналов.

При расчетах линейных аналоговых фильтров в качестве одного из базовых методов математического описания используется интеграл свертки, определяющий связь выходного сигнала фильтра $y(t)$ с входным сигналом $u(t)$ и импульсной характеристикой фильтра $g(t)$

$$y(t) = \int_0^t g(t - \tau)u(\tau)d\tau = \int_0^t g(\tau)u(t - \tau)d\tau. \quad (9.1)$$

Однако эта же операция может быть выполнена и цифровым вычислителем путем замены интеграла суммой разумно выбранного количества слагаемых

$$y(kT_0) = \sum_{n=0}^N u(nT_0)g(kT_0 - nT_0)T_0. \quad (9.2)$$

Получившийся сигнал $y(kT_0)$ отличается от своего аналогового прототипа (9.1) только дискретной формой и при правильно выбранной частоте дискретизации может быть восстановлен (с учетом допущений об идеальности квантования по уровню). Поэтому цифровой вычислитель, выполнивший те же действия, что и аналоговый фильтр с заданной импульсной характеристикой, вполне уместно называть цифровым фильтром, сохраняющим характеристики фильтра прототипа.

Опуская в (9.2) множитель T_0 , данное уравнение записывается в более простом виде

$$y(kT_0) = \sum_{n=0}^N u(nT_0)g(kT_0 - nT_0)$$

или при переходе к дискретному времени

$$y(k) = \sum_{n=0}^N u(n)g(k - n). \quad (9.3)$$

Рассмотрим в качестве аналогового фильтра-прототипа простейшую RC-цепь, схема которой приведена на рис.9.3.

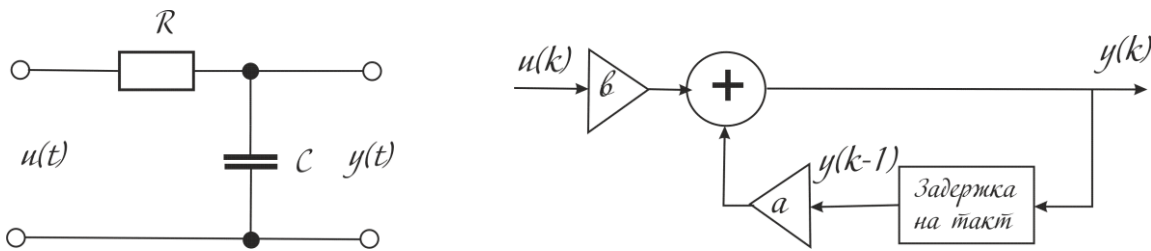


Рис.9.3. RC-фильтр прототип и его цифровой эквивалент

Получение цифрового фильтра, обладающего свойствами аналогового прототипа, базируется на дискретизации временных процессов, протекающих в фильтре-аналоге. При этом производная по времени заменяется выражением

$$\frac{dy(t)}{dt} \approx \frac{y(kT_0) - y((k-1)T_0)}{T_0} = \frac{y(k) - y(k-1)}{T_0}.$$

Дифференциальное уравнение, описывающее RC-цепь, как уже было многократно показано, имеет вид

$$RC \frac{dy(t)}{dt} + y(t) = u(t)$$

или

$$\tau \frac{dy(t)}{dt} + y(t) = u(t),$$

где $\tau = RC$ – постоянная времени цепи.

Переходя к дискретному времени и подставляя вместо производной аппроксимирующее выражение, получаем

$$y(k) = ay(k-1) + bu(k),$$

где $a = \frac{\tau}{\tau + T_0}$, $b = \frac{T_0}{T_0 + \tau} = 1 - a$.

Полученное уравнение называют разностным уравнением цифрового фильтра. Структурная схема, соответствующая данному уравнению приведена на рис. 9.3.

Из приведенной схемы видно, что цифровая РС-цепь состоит из комбинации трех типов функциональных блоков, выполняющих операции умножения, суммирования и задержки сигнала на один такт квантования T_0 .

Величины a и b называют коэффициентами цифрового фильтра.

Особенность полученной схемы цифрового фильтра является наличие петли обратной связи обеспечивающей использование предшествующего значения сигнала $y(k-1)$ при вычислении значения текущего значения сигнала $y(k)$. Фильтры такого типа получили название рекурсивных.

Для сравнения характеристик исходного фильтра-аналога и полученного цифрового фильтра сравним их временные характеристики. Из предыдущего материала известно, что импульсная характеристика интегрирующей РС-цепи имеет вид

$$g(t) = \frac{1}{\tau} e^{-\frac{t}{\tau}}.$$

Если данную характеристику подвергнуть идеализированному квантованию по уровню и дискретизации во времени с интервалом T_0 , то для отсчетов получаемой дискретной импульсной характеристики (ДИХ) можно записать

$$g(k) = \frac{1}{\tau} e^{-\frac{T_0 k}{\tau}}.$$

Такой характеристикой должен обладать цифровой фильтр, заменяющий РС-прототип. Для проверки этого факта получим функцию $g(k)$ на основе решения разностного уравнения цифрового фильтра при условии, что на его вход подан единичный импульс $\delta(t)$ и начальные условия нулевые, то есть $y(k-1) = 0$.

В этом случае, нетрудно подсчитать, что последовательность отсчетов входного сигнала будет иметь вид

$$u(k) = 1, 0, 0, 0, \dots,$$

а выходной сигнал

$$y(k) = b, ba, ba^2, \dots, ba^k, \dots$$

Тогда дискретную импульсную характеристику можно представить в виде

$$g(k) = ba^k = \frac{T_0}{T_0 + \tau} \left(\frac{\tau}{\tau + T_0} \right)^k.$$

Полученный результат отличается по своему виду от ожидаемой экспоненциальной зависимости. Поэтому произведем некоторые преобразования.

Представим коэффициент a в виде

$$a = \frac{\tau}{\tau + T_0} = \left(1 + \frac{T_0}{\tau} \right)^{-1}.$$

Из курса математики известно, что

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \dots$$

При малых значениях x можно ограничиться первыми двумя членами разложения, поэтому при условии малости отношения $\frac{T_0}{\tau}$, получаем

$$a \approx e^{-\frac{T_0}{\tau}}.$$

Коэффициент b выполняет роль масштабирующего множителя и поэтому его выбор в достаточной степени является произвольным. В частности, его часто рекомендуют выбирать $b = 1$, (однако для полного совпадения интересующих характеристик необходимо учитывать соответствующую масштабную поправку). С учетом сделанных допущений итоговое уравнение для дискретной импульсной характеристики принимает вид

$$g(k) \approx be^{-\frac{T_0}{\tau}k}.$$

Причины возникших отклонений состоят в использованном приближенном методе дискретизации дифференциального уравнения фильтра-прототипа, что не позволило точно определить коэффициенты цифрового фильтра. Причины этого явления будут подробно рассмотрены позже.

На рис.9.4. приведен пример моделирования реакции аналоговой и дискретной системы на единичный ступенчатый импульс. Из данного примера видно, каким образом пришлось изменять коэффициенты дискретной системы при изменении интервала дискретизации.

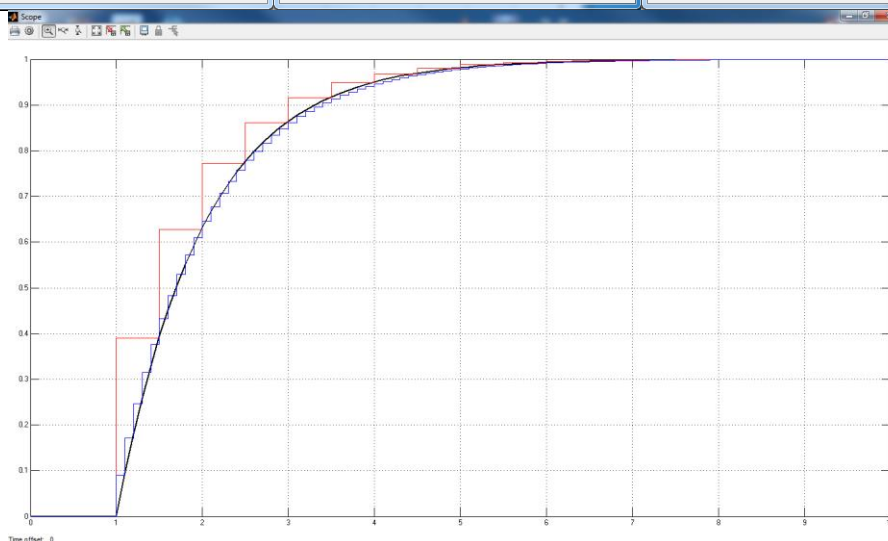
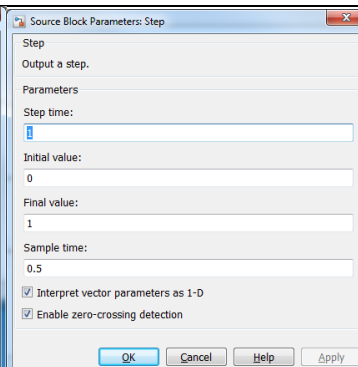
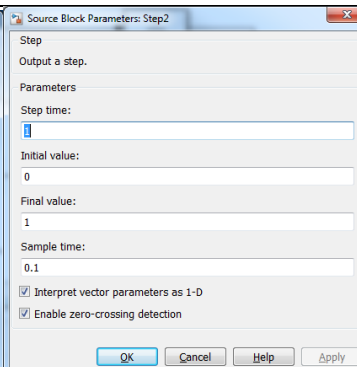
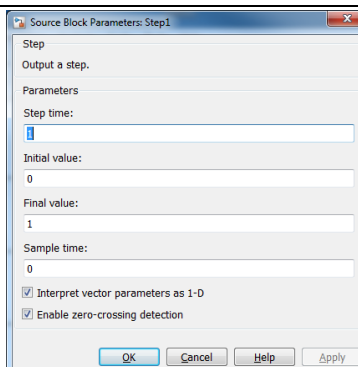
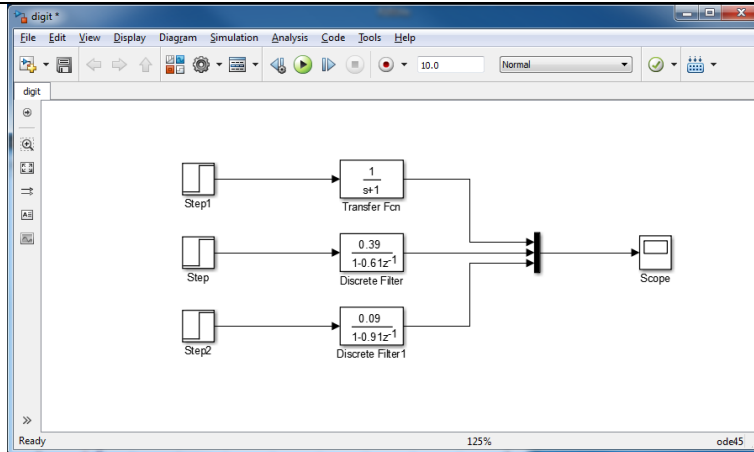


Рис.9.4. Сравнительное моделирование аналоговой и дискретных систем при различных интервалах дискретизации

Обратим внимание на то, что ДИХ полученного цифрового фильтра теоретически имеет бесконечную протяженность во времени, поэтому фильтры с таким типом ДИХ называют фильтрами с бесконечной импульсной характеристикой или БИХ-фильтрами.

Помимо рассмотренного варианта дискретного описания РС-цепи, возможен и другой подход. Вернемся к уравнению свертки в виде

$$y(k) = \sum_{n=0}^{\infty} u(n)g(k-n). \quad (9.4)$$

Из уравнения (9.4) видно, что текущее значение выходного сигнала $y(k)$ может быть рассчитано как взвешенная сумма текущего и задержанных во времени значений входного сигнала.

В общем случае данная сумма бесконечна, но для целей физической реализуемости ДИХ можно использовать конечное число слагаемых

$$y(k) = u(k)g(0) + u(k-1)g(1) + \dots + u(k-N)g(N). \quad (9.5)$$

Реализация уравнения (9.5) приводит к структуре цифрового фильтра, изображенной на рис.9.4.

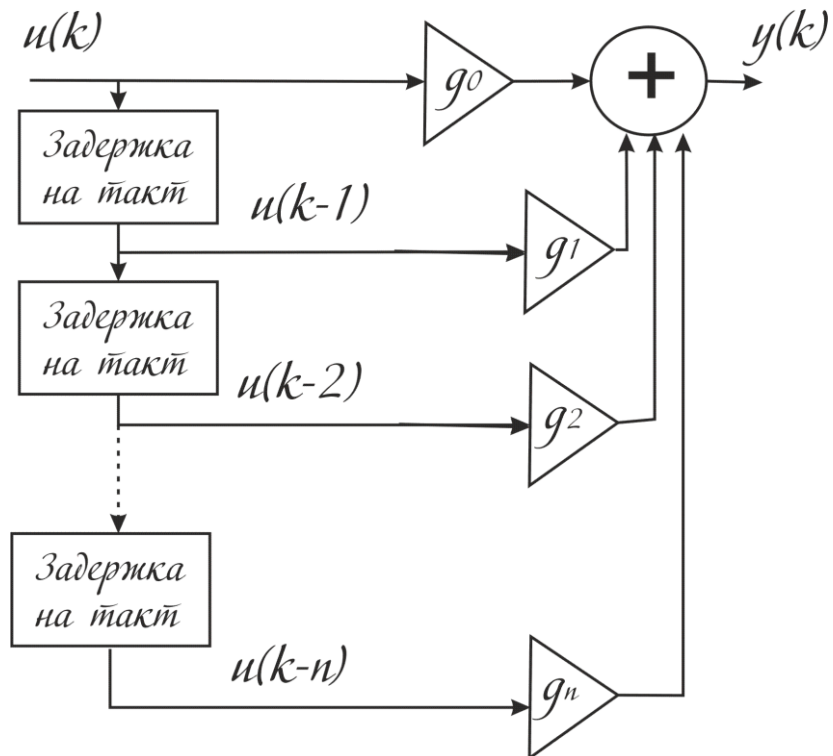


Рис.9.4. Схема нерекурсивного цифрового фильтра.

В структуре данного варианта реализации цифрового фильтра отсутствуют петли обратной связи, поэтому он называется нерекурсивным, а ограниченность во времени ДИХ данного типа фильтров позволяет отнести их к классу фильтров с конечной импульсной характеристикой или просто КИХ-фильтров.

В результате для одного и того же аналогового прототипа (РС-цепи) получено два варианта цифрового представления: БИХ-фильтр и КИХ-фильтр.

Сравнение этих вариантов показывает, что КИХ-фильтр имеет значительно более сложную форму, причем точность его соответствия аналоговому прототипу зависит от числа использованных коэффициентов. Погрешности воспроизведения выходного сигнала из-за отсечения части импульсной характеристики в ряде случаев могут очень существенными.

К достоинствам КИХ-фильтров следует отнести их абсолютную устойчивость и возможность реализации ДИХ любой заданной формы. Кроме того при определенных условиях они обладают линейной фазовой характеристикой, что особенно важно для сигналов, информационными параметрами которых являются частота и фаза.

9.3. Z- преобразование и его свойства

Для описания динамических свойств линейных систем в непрерывном времени мы широко использовали математический аппарат, основанный на преобразовании Лапласа, переводящего оригинал вещественного сигнала $x(t)$ в комплексное изображение $X(s)$

$$X(s) = \int_0^{\infty} x(t)e^{-st} dt, \quad (9.6)$$

где $s = \sigma + j\omega$.

Для дискретного времени преобразование Лапласа (9.6) принимает вид

$$X(s) = T_0 \sum_{k=0}^{\infty} x(kT_0)e^{-skT_0}, \quad (9.7)$$

где T_0 – временной интервал дискретизации аналогового сигнала $x(t)$.

Если в уравнении (9.7) опустить нормирующий множитель T_0 и ввести замену переменных

$$z = e^{sT_0},$$

то при переходе к дискретным отсчетам времени получаем

$$X(z) = \sum_{k=0}^{\infty} x(k)z^{-k}. \quad (9.8)$$

Полученное уравнение (9.8) носит название Z-преобразования.

Рассмотрим примеры Z-преобразования для некоторых простых функций.

1. Пусть сигнал $x(t)$ представляет собой масштабированный с коэффициентом A единичный импульс. Тогда

$$x(kT_0) = \begin{cases} A\delta(k), & \text{при } k = 0, \\ 0, & \text{при } k \neq 0. \end{cases}$$

используя формулу Z-преобразования в этом случае получаем

$$X(z) = \sum_{k=0}^{\infty} x(k)z^{-k} = A.$$

Следовательно, Z-преобразование одиночного импульса равно A .

2. Пусть $x(t)$ представляет собой масштабированную с коэффициентом A единичную ступенчатую функцию, то есть

$$x(kT_0) = \begin{cases} A, & \text{при } kT_0 \geq 0, \\ 0, & \text{при } kT_0 < 0. \end{cases}$$

Применяя формулу Z-преобразования, для данного случая запишем

$$X(z) = \sum_{k=0}^{\infty} x(k)z^{-k} = A \sum_{k=0}^{\infty} z^{-k} = A(1 + z^{-1} + z^{-2} + \dots) = \frac{A}{1 - z^{-1}} = \frac{Az}{z - 1},$$

для $|z^{-1}| < 1$.

3. Пусть $x(t)$ представляет собой линейную функцию времени ($x(t) = At$). Тогда $x(kT_0) = AkT_0$ и

$$X(z) = \sum_{k=0}^{\infty} x(k)z^{-k} = \sum_{k=0}^{\infty} AkT_0z^{-k} = AT_0z^{-1}(1 + 2z^{-1} + 3z^{-1} + \dots) = \frac{AT_0z^{-1}}{(1 - z^{-1})^2}$$

для $|z^{-1}| < 1$.

4. Пусть $x(t)$ представляет собой экспоненциальную функцию $x(t) = Ae^{-\alpha t}$. Тогда $x(kT_0) = Ae^{-\alpha kT_0}$ и

$$X(z) = \sum_{k=0}^{\infty} x(k)z^{-k} = \sum_{k=0}^{\infty} Ae^{-\alpha kT_0}z^{-k} = A(1 + e^{-\alpha T_0}z^{-1} + e^{-\alpha 2T_0}z^{-2} + \dots) =$$

$$= \frac{A}{1 - e^{-\alpha T_0}z^{-1}} = \frac{Az}{z - e^{-\alpha T_0}}.$$

Теперь рассмотрим основные свойства Z-преобразования. (Операция получения Z-преобразования обозначается $\mathbb{Z}$).

1. Свойство линейности

$$\mathbb{Z}(\alpha x(k) + \beta y(k)) = \alpha \mathbb{Z}(x(k)) + \beta \mathbb{Z}(y(k)) = \alpha X(z) + \beta Y(z).$$

2. Свойство смещения аргумента в области оригинала

$$\mathbb{Z}(x(k \pm m)) = z^{\pm m} \mathbb{Z}(x(k)) = z^{\pm m} X(z)$$

3. Свойство смещения независимой переменной в области изображения

$$X(e^{\pm \alpha T_0} z) = \mathbb{Z}(e^{\mp \alpha T_0} x(k)).$$

4. Правило дифференцирования изображения

$$\mathbb{Z}(kT_0 x(k)) = -T_0 z \frac{dX(z)}{dz}.$$

5. Связь начальных и конечных значений

$$\lim_{k \rightarrow 0} x(k) = \lim_{z \rightarrow \infty} X(z).$$

Если z-преобразование решетчатой функции не имеет полюса на единичной окружности или вне ее, то конечное значение определяется выражением

$$\lim_{k \rightarrow \infty} x(k) = \lim_{z \rightarrow 1} \left[\frac{z-1}{z} X(z) \right].$$

6. Свойство свертки.

Если $X(z) = \mathbb{Z}(x(k))$ и $Y(z) = \mathbb{Z}(y(k))$, то

$$X(z)Y(z) = \mathbb{Z}\left\{ \sum_{m=0}^{\infty} x(m)y(k-m) \right\}$$

7. Обратное Z-преобразование ставит в соответствие функции комплексной переменной $X(z)$ последовательность $x(k)$ в соответствии с уравнением

$$x(k) = \frac{1}{2\pi j} \oint X(z)z^{n-1} dz,$$

где интеграл берется по контуру, расположенному в области сходимости $X(z)z^{n-1}$ и охватывающему все полюсы функции $X(z)$.

9.4. Передаточная функция цифрового фильтра и формы ее представления

Рассмотренное в предыдущем разделе Z-преобразование дискретных последовательностей, позволят подойти к описанию цифровых систем с позиций, близких аппарату передаточной функции, используемому при анализе и синтезе систем непрерывного времени.

Рассмотрим цифровой фильтр (систему) с импульсной характеристикой $g(k)$, на вход которого поступает дискретный сигнал $u(k)$. Тогда выходной сигнал цифрового фильтра может быть определен с использованием дискретной свертки

$$y(k) = \sum_{m=0}^{\infty} u(m)g(k-m). \quad (9.9)$$

Из свойств Z-преобразования известно, что операции свертки (9.9) во временной области соответствует умножение изображений сворачиваемых функций, то есть

$$Y(z) = U(z)G(z).$$

Поэтому функцию

$$G(z) = \frac{Y(z)}{U(z)}, \quad (9.10)$$

являющуюся отношением изображений выходного и входного сигнала можно назвать дискретной передаточной функцией рассматриваемого фильтра (системы).

Нетрудно заметить, что по аналогии с передаточной функцией непрерывных систем, дискретная передаточная функция (9.10) связана с импульсной характеристикой системы Z-преобразованием

$$G(z) = \sum_{k=0}^{\infty} g(k)z^{-k}. \quad (9.11)$$

Возвращаясь к рассмотренному ранее примеру RC-цепи, найдем ее дискретную передаточную функцию.

Для импульсной характеристики RC-цепи ранее было получено уравнение

$$g(k) = be^{-\frac{T_0}{\tau}k}.$$

Дискретная передаточная функция цепи с такой импульсной характеристикой в соответствии с (9.10) и (9.11) может быть найдена как

$$G(z) = \sum_{k=0}^{\infty} g(k)z^{-k} = b \sum_{k=0}^{\infty} e^{-\frac{T_0}{\tau}k} z^{-k} = b \sum_{k=0}^{\infty} q^k,$$

где $q = e^{-\frac{T_0}{\tau}} z^{-1}$.

Применяя к полученному выражению формулу для суммы членов убывающей геометрической последовательности, получим

$$G(z) = \frac{b}{1-q} = \frac{b}{1-az^{-1}}.$$

Структурная схема, соответствующая этому уравнению представлена на рис.9.5.

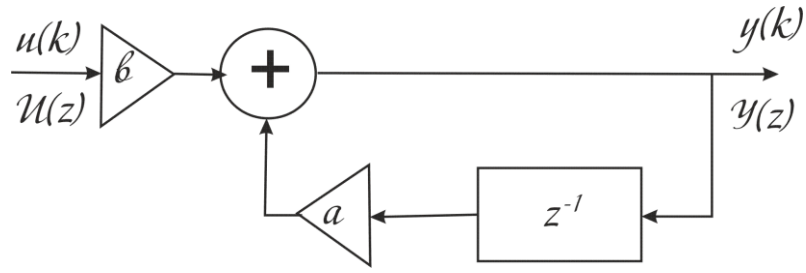


Рис.9.5. Структурная схема передаточной функции цифрового аналога RC-цепи

С точностью до обозначений эта схема соответствует схеме, приведенной на рис. 9.3. Блок, обозначенный символом z^{-1} , производит задержку сигнала на один такт, что совпадает со свойством Z-преобразования для сдвига аргумента.

Для произвольной цифровой системы передаточная функция в общем виде может быть представлена как

$$W_z(z) = \frac{\sum_{n=0}^M b_n z^{-n}}{1 - \sum_{n=1}^L a_n z^{-n}}, \quad (9.12)$$

где a_k, b_k – коэффициенты цифровой системы (фильтра),

W_z – обозначение дискретной передаточной функции (индекс z введен, чтобы не путать ее с передаточной функцией непрерывных систем).

Из уравнения (9.12) с помощью обратного Z-преобразования легко записать разностное уравнение, связывающее между собой отсчеты входного и выходного сигналов

$$y(k) - \sum_{n=1}^L a_n y(k-n) = \sum_{n=0}^M b_n u(k-n)$$

или

$$y(k) = \sum_{n=1}^L a_n y(k-n) + \sum_{n=0}^M b_n u(k-n). \quad (9.13)$$

Необходимо отметить, что в литературе для коэффициентов a_k в (9.12), (9.13) очень часто выбирают обратные знаки, чтобы в выражение для передаточной функции имело только форму суммирования и в числителе, и в знаменателе. Естественно, что никакой принципиальной разницы в результатах от такой замены не возникает, но на этот нюанс следует обращать внимание при использовании каких-либо формул из справочников, а также при использовании пакетов компьютерных программ.

Поскольку переменная z связана с переменной Лапласа s соотношением

$$z = e^{sT_0},$$

то при соответствующей подстановке по дискретной передаточной функции можно определить частотную передаточную функцию цифрового фильтра.

Чаще всего для цифровых фильтров частотные характеристики строятся либо в так называемом масштабе цифровой частоты $\Phi = \omega T_0$, либо в относительном масштабе частоты дискретизации f_0 , то есть $\Phi = 2\pi f / f_0$.

Подставив в уравнение дискретной передаточной функции цифрового эквивалента RC-цепи $z = e^{j\omega T_0} = e^{j\Phi}$, для частотной передаточной функции получаем

$$W_z(j\omega) = \frac{b}{1 - a e^{-j\omega T_0}} = \frac{b}{1 - a e^{-j\Phi}}$$

Воспользовавшись равенством

$$e^{-j\Phi} = \cos\Phi - j\sin\Phi,$$

получаем

$$W_z(j\Phi) = \frac{b}{(1 - a\cos\Phi) + j\sin\Phi}$$

откуда для амплитудно-частотной характеристики данного фильтра получим

$$|W_z(j\Phi)| = \frac{b}{\sqrt{1 + a^2 - 2a\cos\Phi}}.$$

Или в нормированном варианте

$$|W_{z\text{norm}}(j\Phi)| = \frac{(1 - a)}{\sqrt{1 + a^2 - 2a\cos\Phi}},$$

где коэффициент нормировки выбран равным $W_z(0) = \frac{b}{1-b}$.

Сравнивая полученное выражение с амплитудно-частотной характеристикой аналоговой RC-цепи

$$|W(j\omega)| = \frac{1}{\sqrt{1 + \omega^2\tau^2}},$$

видно, что АЧХ дискретного фильтра в отличие от АЧХ прототипа является периодической функцией.

Графики АЧХ прототипа и цифрового аналога RC-цепи для значений $a = 0.8, T_0 = 10$ мс приведены на рис.9.6. На правом графике приведен укрупненный участок АЧХ, соответствующий диапазону частот, удовлетворяющих теореме Котельникова.

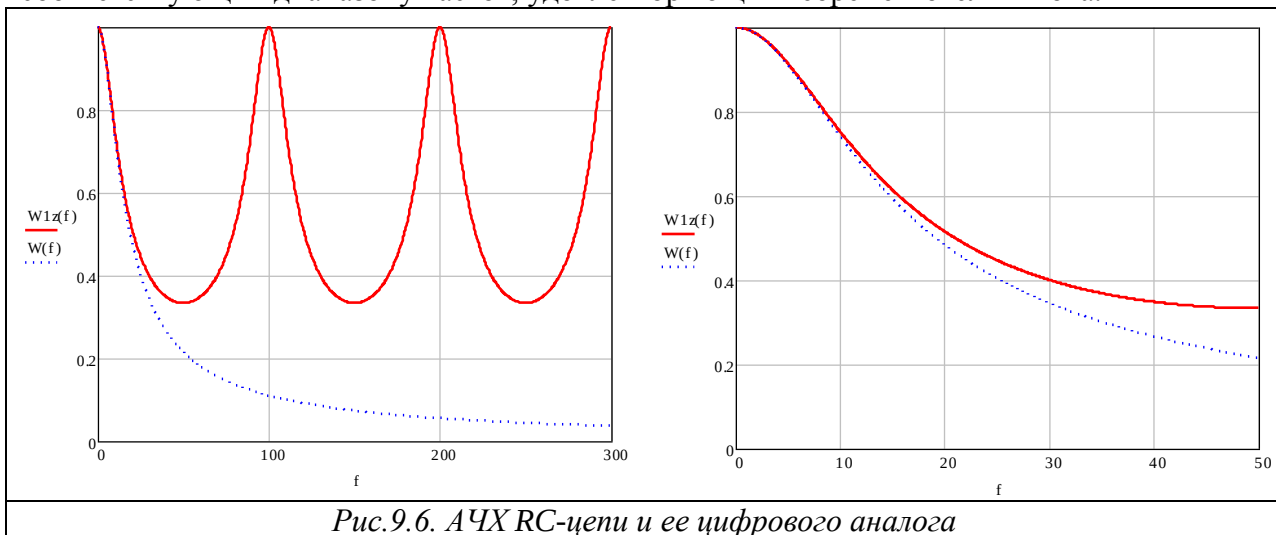


Рис.9.6. АЧХ RC-цепи и ее цифрового аналога

Из приведенных графиков хорошо видно, что близость АЧХ достигается только в области низких частот. При создании цифровых фильтров необходимо учитывать специфические особенности присущие цифровому способу обработки сигналов, правильно выбирать частоту дискретизации и при необходимости устанавливать на входе системы необходимые ограничивающие фильтры.

Поскольку дискретная передаточная функция

$$W_z(z) = \frac{\sum_{n=0}^M b_n z^{-n}}{1 - \sum_{n=1}^L a_n z^{-n}}$$

является рациональным полиномом относительно переменной z^{-k} , ее по аналогии с формами представления передаточной функции непрерывного времени можно представить в форме нулей и полюсов, полюсов и вычетов, а также в форме переменных состояния.

Поскольку аналогичная тема была подробно рассмотрена при изучении линейных цепей непрерывного времени, сейчас отграничимся приведением только общей формы соответствующих уравнений.

В форме нулей и полюсов дискретная передаточная функция имеет вид

$$W_z(z) = K \frac{(1 - z_m z^{-1})(1 - z_{m-1} z^{-1}) \dots (1 - z_1 z^{-1})}{(1 - p_n z^{-1})(1 - p_{n-1} z^{-1}) \dots (1 - p_1 z^{-1})}. \quad (9.14)$$

В форме полюсов и вычетов дискретная передаточная функция имеет вид

$$W_z(z) = \frac{k_1}{s - p_1 z^{-1}} + \frac{k_2}{s - p_2 z^{-1}} + \dots + \frac{k_n}{s - p_n z^{-1}}. \quad (9.15)$$

Представление цифровой системы в переменных состояния соответственно примет вид

$$\begin{aligned} \mathbf{x}(k+1) &= \mathbf{A}\mathbf{x}(k) + \mathbf{B}u(k), \\ y(k) &= \mathbf{C}\mathbf{x}(k) + Du(k), \end{aligned} \quad (9.16)$$

где $y(k)$ – дискретная последовательность выходного сигнала,

$u(k)$ – дискретная последовательность входного сигнала,

$\mathbf{x}(k)$ – n -мерный вектор состояния,

$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}$ – квадратная матрица размерности $n \times n$,

$\mathbf{B} = \begin{bmatrix} b_1 \\ b_2 \\ \dots \\ b_n \end{bmatrix}$ – вектор-столбец размерности $n \times 1$,

$\mathbf{C} = [c_1 \ c_2 \ \dots \ c_n]$ – вектор-строка размерности $1 \times n$

Найти уравнения связи дискретной передаточной функции в различных формах представления (9.14) – (9.16) рекомендуем провести самостоятельно в качестве дополнительного упражнения.

Пример 9.1.

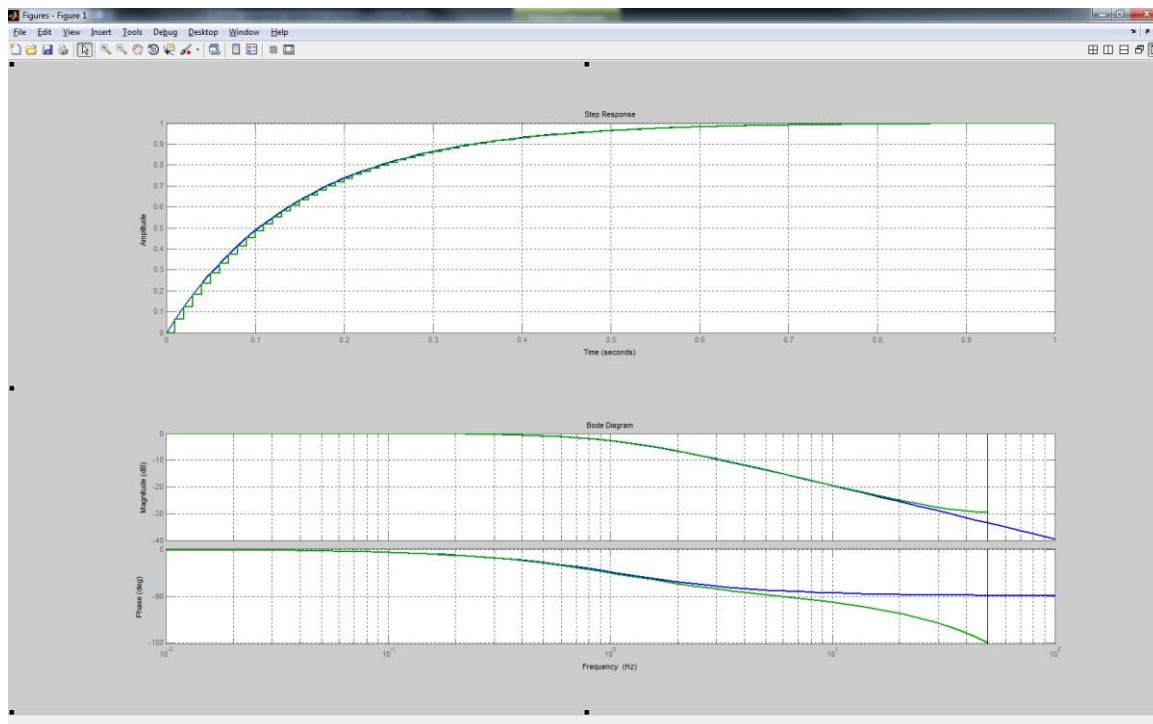
Рассмотрим процесс нахождения цифрового эквивалента для аналоговой RC-цепи с использованием программы MATLAB. Листинг рабочей сессии приведен на рис.9.7

```
T0=0.01; %Задание периода дискретизации непрерывного сигнала
RC=0.15;%Задание постоянной времени аналогового фильтра
sys=tf(1, [RC,1]) % Задание передаточной функции аналогового фильтра
subplot(2,1,1); step(sys); % построение графики переходной характеристики RC-цепи
hold on;
sysz1=c2d(sys,T0) % Расчет дискретной передаточной функции RC-цепи
step(sysz1) %Построение графика переходной характеристики дискретной системы
subplot(2,1,2); bode(sys); % Построение диаграммы Боде аналогового фильтра
hold on;
bode(sysz1);% Построение диаграммы Боде цифрового эквивалента
sys2=d2c(sysz1) % Проверка обратного преобразования цифрового эквивалента
sys =

    1
-----
0.15 s + 1
Continuous-time transfer function.
```

```
sysz1 =
    0.06449
    -----
    z - 0.9355
Sample time: 0.01 seconds
Discrete-time transfer function.
sys2 =
    6.667
    -----
    s + 6.667
Continuous-time transfer function.
```

Обратим внимание, что при обратном преобразовании цифрового фильтра в непрерывный аналог форма записи передаточной функции несколько отличается от первоначальной, однако поделив числитель и знаменатель полученной передаточной функции на 6,667 легко получить исходный вариант представления передаточной функции RC- цепи



Полученные графики временных и частотных характеристик наглядно показывают различия аналогового прототипа и его цифрового эквивалента.

Рис.9.7. Построение цифрового аналога для RC-цепи.

9.5. Основы проектирования дискретных фильтров

Задача проектирования цифрового фильтра (9.12)

$$W_z(z) = \frac{\sum_{n=0}^M b_n z^{-n}}{1 - \sum_{n=1}^L a_n z^{-n}},$$

где a_k, b_k – коэффициенты цифровой системы (фильтра), состоит в отыскании коэффициентов фильтра a_k, b_k , обеспечивающих соответствие характеристик фильтра заданным требованиям. Поскольку выбор оптимального значения коэффициентов a_k, b_k проводится для цифрового фильтра заданной структуры, то строго говоря, задача синтеза распадается на два этапа:

- выбор структуры цифрового фильтра (определение значений M и L),
- подбор параметров для фильтра выбранной структуры.

В ряде случаев в процессе синтеза приходится опробовать различные структуры фильтра и принимать окончательное решение как компромисс между возможностями технической реализации и отклонениями от желаемых характеристик.

В соответствии с введенной ранее классификацией цифровых фильтров методы синтеза можно разделить на методы синтеза рекурсивных фильтров и методы синтеза нерекурсивных фильтров.

Проектирование цифровых фильтров можно вести двумя способами, первый из которых основан на использовании аналоговых прототипов, а второй, получивший название прямого, проводится без использования прототипа.

При синтезе цифровых фильтров с использованием аналоговых прототипов предполагается, что передаточная функция аналогового фильтра $W(s)$, удовлетворяющего поставленным требованиям, известна. Задача синтеза в данном случае состоит в нахождении эквивалентной дискретной передаточной функции $W_z(z)$. Однако, синтезируемый дискретный фильтр не может быть полностью идентичен непрерывному прототипу по своим характеристикам, в частности и потому, что, как уже отмечалось, частотные характеристики цифрового фильтра имеют периодичность. Поэтому при синтезе цифрового фильтра по аналоговому прототипу можно говорить лишь об определенном соответствии их характеристик в заданном частотном диапазоне.

Для синтеза рекурсивных цифровых фильтров по заданному аналоговому прототипу используются два основных метода:

- метод билинейного Z-преобразования,
- метод инвариантной импульсной характеристики.

9.5.1. Синтез рекурсивных цифровых фильтров по аналоговому прототипу. Метод билинейного Z-преобразования

Метод билинейного Z-преобразования основан на таком преобразовании дробно-рациональной передаточной функции аналогового прототипа $W(s)$, которое отображает мнимую ось переменной s на единичную окружность в z -области, трансформируя таким образом только частотную ось характеристики аналогового фильтра.

Простейшей из функций, удовлетворяющих поставленной задаче является билинейное Z-преобразование, определяемое соотношением

$$s = \frac{2}{T_0} \frac{z - 1}{z + 1}. \quad (9.17)$$

Поясним происхождение формулы (9.17). Из предыдущего материала известно, что связь между аргументами непрерывной и дискретной передаточными функциями имеет вид

$$z = e^{sT_0},$$

откуда

$$s = \frac{1}{T_0} \ln z.$$

Эти уравнения не могут быть использованы при синтезе БИХ-фильтров, так как не позволяют представить требуемую передаточную функцию в дробно-рациональном виде. Для преодоления данного затруднения логарифмическую функцию можно представить рядом следующего вида

$$\ln z = 2 \left[\frac{z-1}{z+1} + \frac{1}{3} \left(\frac{z-1}{z+1} \right)^3 + \frac{1}{5} \left(\frac{z-1}{z+1} \right)^5 + \dots \right].$$

Ограничиваясь в этом разложении только первым членом получаем

$$s = \frac{2}{T_0} \frac{z-1}{z+1} = \frac{2}{T_0} \frac{1-z^{-1}}{1+z^{-1}}.$$

После несложных преобразований из этого уравнения нетрудно получить

$$z = \frac{\frac{2}{T_0} + s}{\frac{2}{T_0} - s},$$

где $s = \sigma + j\omega$.

Движение по оси $j\omega$ предполагает выполнение условия $\sigma = 0$, что дает возможность записать

$$z = \frac{\frac{2}{T_0} + j\omega}{\frac{2}{T_0} - j\omega} = \frac{Ae^{j\theta}}{Ae^{-j\theta}},$$

где $\theta = \arctg \frac{\omega T_0}{2}$.

Окончательное выражение принимает вид

$$z = e^{j2\arctg \frac{\omega T_0}{2}}.$$

Анализ полученного уравнения показывает, что при таком способе перехода к комплексному аргументу z , его годограф в z -плоскости является окружностью единичного радиуса, но при линейном изменении ω происходит не циклическое движение по окружности, а ограниченное перемещение от точки с координатами (1,0) в точку с координатами (-1,0).

Понятно, что при таком преобразовании происходит деформация АЧХ фильтра-прототипа, при его пересчете в цифровую форму, но это не недостаток билинейного преобразования, а его особенность, которая к тому же легко прогнозируется и может быть скомпенсирована при расчетах.

Кратко поясним процедуру синтеза цифрового фильтра на примере проектирования БИХ-фильтра нижних частот по заданной частотной характеристике аналогового прототипа. На рис.9.8. приведены основные параметры АЧХ, задаваемые перед процессом синтеза фильтра.

Таковыми параметрами являются

ω_c – частота среза, определяющая полосу пропускания среза,

ω_z – граница области затухания,

K_c – уровень АЧХ, определяющий неравномерность передачи сигнала в полосе пропускания,

K_z – уровень АЧХ, определяющий ослабление в области затухания.

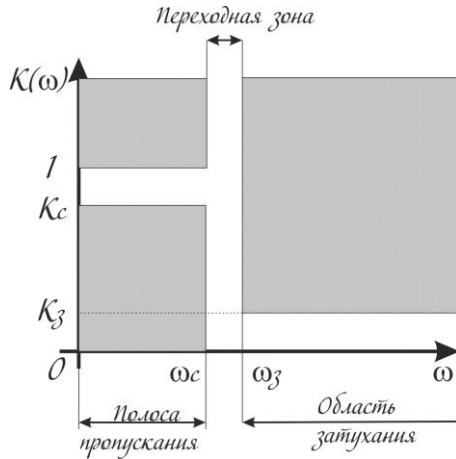


Рис.9.8. Основные параметры АЧХ фильтра нижних частот

Поведение АЧХ в полосе пропускания и области затухания, в частности допустимость пульсаций в данных областях является дополнительными условиями.

В общем случае возможны четыре основных варианта поведения АЧХ (монотонная, с пульсацией в полосе пропускания, с пульсацией в полосе задержания, с пульсациями в полосе пропускания и полосе задержания). Схематичный вид всех этих характеристик приведен на рис.9.9.

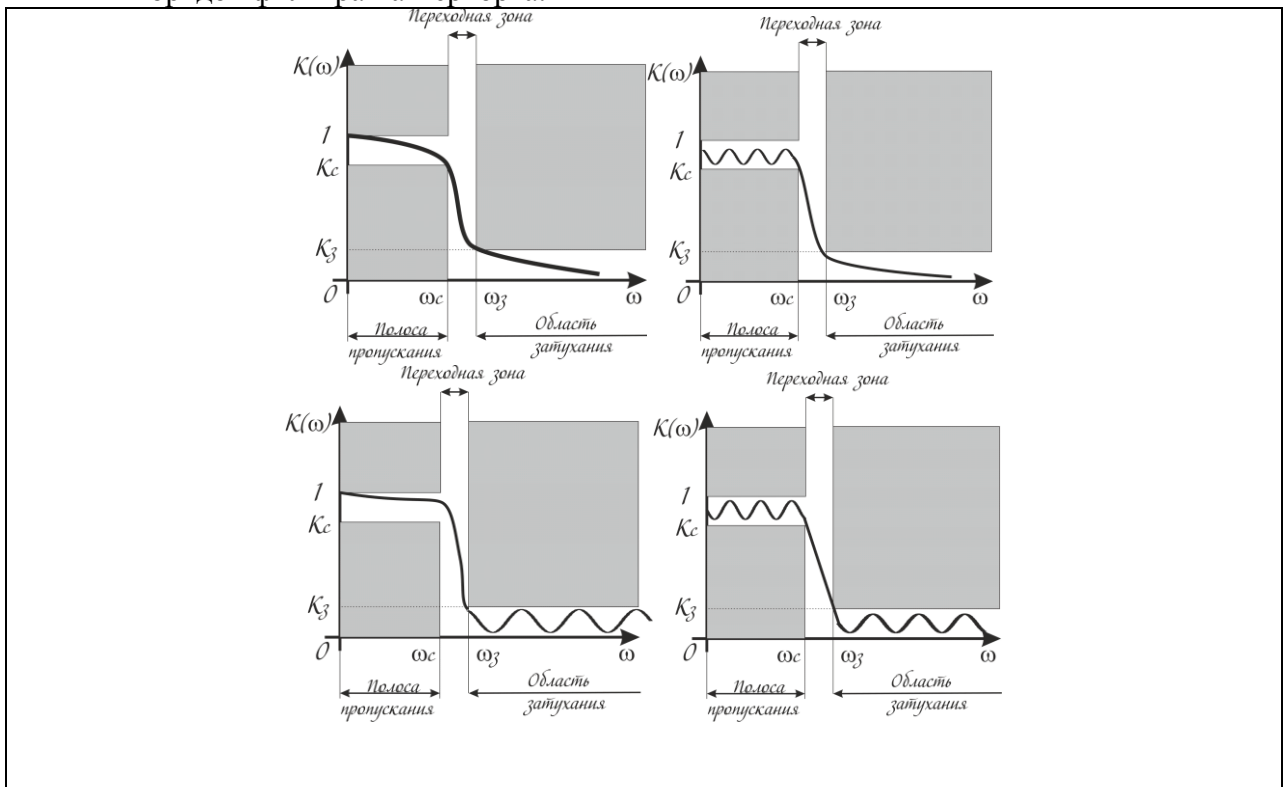
Для различных форм АЧХ используются различные

аппроксимации характеристик аналогового прототипа. Наиболее часто применяются аппроксимации Баттерворта, Чебышева и Кауэра. Фильтры, получаемые на основе данных аппроксимаций называют соответствующими именами – фильтр Баттерворта, фильтр Чебышева, Фильтр Кауэра. Также применяются эллиптические фильтры и фильтры Бесселя.

Так, например, АЧХ с монотонно убывающей в зоне нижних частот характеристикой имеет фильтр Баттерворта,

$$K(\omega) = \frac{1}{\sqrt{1 + \left(\frac{\omega}{\omega_0}\right)^N}}, \quad (9.18)$$

где ω_0 – характерная частота фильтра Баттерворта,
 N – порядок фильтра Баттерворта.



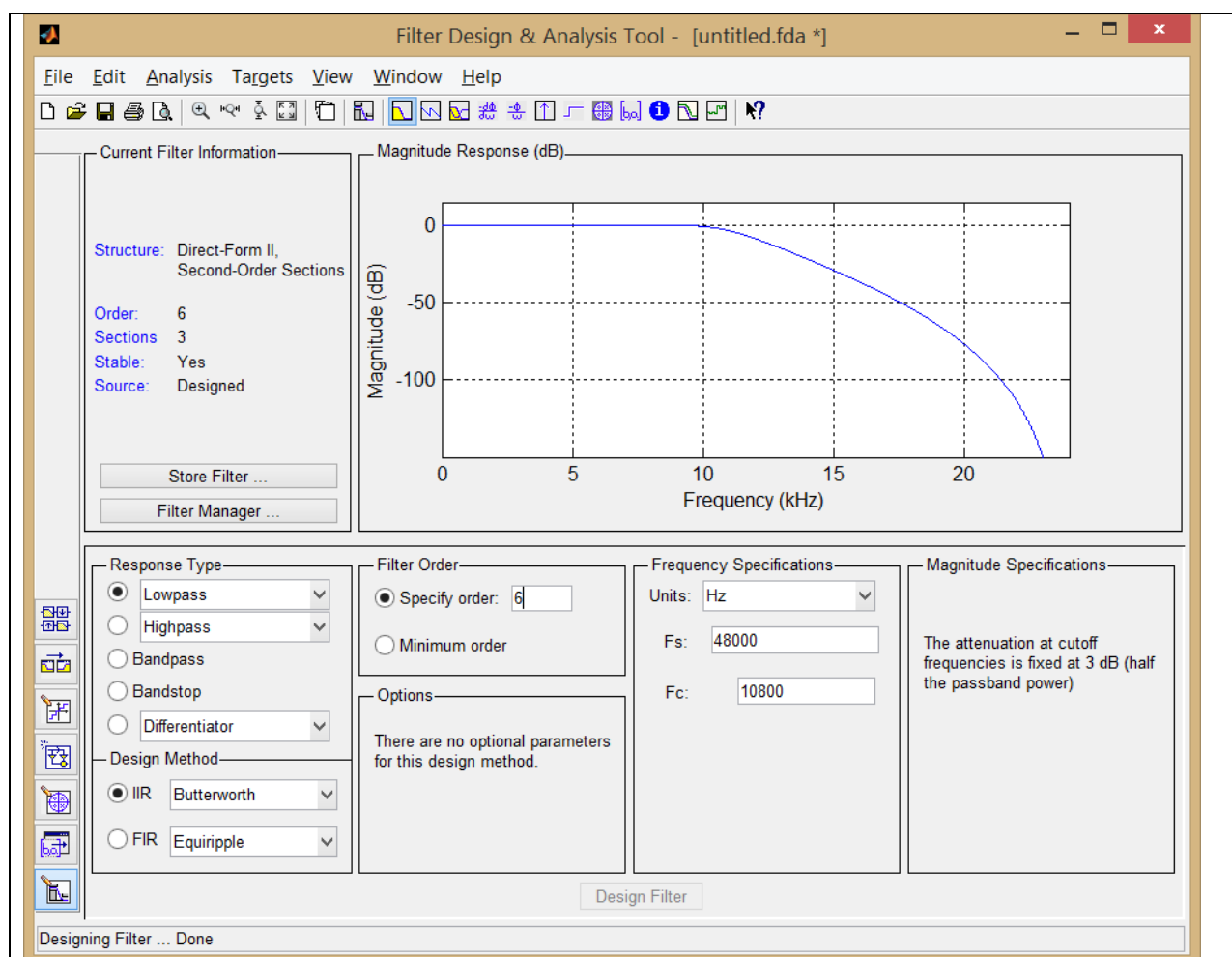


Рис.9.9. Возможные варианты АЧХ фильтров нижних частот и диалоговое окно для расчета параметров цифровых фильтров в программе Matlab

Амплитудно-частотные характеристики фильтра Баттерворта (9.18), построенные в относительном масштабе частоты, приведены на рис.9.10.

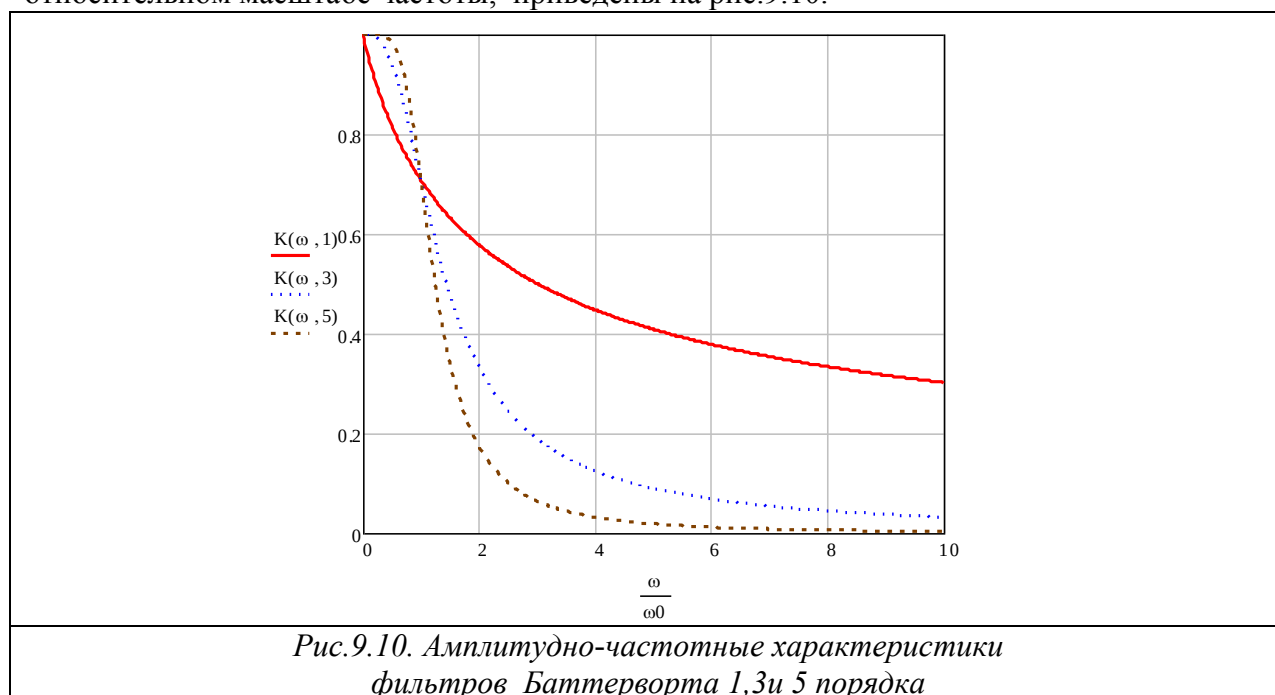


Рис.9.10. Амплитудно-частотные характеристики фильтров Баттерворта 1,3и 5 порядка

Фильтр Чебышева первого рода имеет характеристику

$$K(\omega) = \frac{1}{\sqrt{1 + \frac{\omega}{\omega_0} \varepsilon^2 T_N^2}} \quad (9.19)$$

где ω_0 – частота среза,

T_N – полином Чебышева N-го порядка от аргумента $\frac{\omega}{\omega_0}$,

ε – параметр, определяющий частоту пульсаций АЧХ в полосе пропускания.

Фильтр Кауэра имеет характеристику

$$K(\omega) = \frac{1}{\sqrt{1 + \varepsilon^2 R_N^2(\omega, \gamma)}}, \quad (9.20)$$

где $R_N(\omega, \gamma)$ – эллиптическая функция Якоби,

ε, γ – параметры, определяющие пульсации функции $R_N(\omega, \gamma)$ и АЧХ фильтра.

С точки зрения получения более "качественного" фильтра желательно иметь монотонный характер АЧХ, что свойственно фильтрам Баттерворта, однако во многих практических применениях некоторая колебательность характеристики фильтра вполне допустима, поэтому предпочтение отдается более простым в реализации фильтрам Чебышева или Кауэра (9.18) – (9.20).

В настоящее время синтез цифровых фильтров чаще всего проводится с помощью специализированных компьютерных программ. Эти вопросы подробно изучаются в курсах цифровой обработки сигналов и не входят в задачи нашего курса. Поэтому мы на них останавливаться не будем.

9.5.2. Общие сведения о методе инвариантной импульсной характеристики и прямых методах синтеза цифровых фильтров

Представление передаточной функции аналогового прототипа в форме полюсов и вычетов (9.15)

$$W(s) = \frac{k_1}{s - p_1} + \frac{k_2}{s - p_2} + \dots + \frac{k_n}{s - p_n}$$

позволяет представить импульсную характеристику стационарной цепи с постоянными параметрами в виде суммы экспоненциальных слагаемых, которые в случае простых корней имеют вид

$$g_i(t) = Ae^{p_i t}$$

или

$$g_m(t) = At^m e^{p_m t}$$

для корней с кратностью m .

Каждое из слагаемых может быть дискретизировано во времени, что дает $g_i(kT_0) = Ae^{p_i kT_0}$.

Если теперь произвести Z-преобразование последовательности, соответствующей такому экспоненциальному ряду, то можно получить

$$W_{zi}(z) = A(1 + e^{p_i T_0} z^{-1} + e^{2p_i T_0} z^{-2} + e^{3p_i T_0} z^{-3} + \dots) = \frac{A}{1 - e^{p_i T_0} z^{-1}}, \quad (9.21)$$

что соответствует дискретному рекурсивному фильтру первого порядка.

Аналогичные преобразования для кратных корней приводят к выражению

$$W_{zm}(z) = A \frac{d^m}{dp_m^m} \left(\frac{1}{1 - e^{p_m T_0} z^{-1}} \right), \quad (9.22)$$

которое соответствует (после вычисления m -ой производной) рекурсивному фильтру m -го порядка.

После выполнения аналогичных действий (9.21) или (9.22) для всех полюсов исходной передаточной функции аналогового прототипа можно получить дискретную версию импульсной характеристики, которая и используется для определения параметров цифрового эквивалентного фильтра.

Прямые методы синтеза цифровых фильтров не используют в качестве основы для расчетов аналоговые прототипы. Исходными данными в этом случае служат какие-либо параметры фильтра, например его амплитудно-частотная характеристика, которые могут задаваться с большей степенью произвольности.

Поиск структуры и параметров цифрового фильтра при таком подходе проводится численными методами оптимизации, обеспечивающими минимум (или максимум) заданной функции качества. Если обеспечивается нахождение точного экстремума показателя качества, то методы синтеза называют оптимальными, если решение ищется со значительными упрощающими допущениями, то в результате получается субоптимальный фильтр.

9.5.3. Частотные преобразования АЧХ фильтров нижних частот

В предыдущих разделах были рассмотрены вопросы синтеза цифрового БИХ-эквивалента фильтров нижних частот с различными вариантами аппроксимации АЧХ.

С помощью так называемых частотных преобразований фильтр нижних частот может быть использован как базовый для синтеза других типов фильтров (фильтры верхних частот, полосовые фильтры, заграждающие фильтры).

Под частотными преобразованиями фильтров понимаются такие преобразования их передаточной функции, при которых изменение происходит только вдоль частотной оси.

Такие преобразования осуществляются путем замены аргумента исходной передаточной функции на некоторую функцию в виде

$$W_2(s) = W_1(f_s(s)).$$

При этом применяемая функция преобразования должна сохранять дробно-рациональную форму передаточной функции и при $s = j\omega$ давать чисто мнимый результат. Каждой функции преобразования $f_s(s)$ соответствует функция преобразования частоты АЧХ $f_\omega(\omega)$. На практике для описанных целей преобразования оказывается достаточным использование дробно-рациональных функций $f_s(s)$ не выше второго порядка.

Простейшее преобразование ФНЧ состоит в переносе частоты среза при сохранении типа фильтра. Это преобразование легко произвести путем замены переменных

$$f_s(s) = \frac{s}{\omega_0}, \quad (9.23)$$

где ω_0 – требуемая частота среза.

Функция преобразования частоты в этом простейшем случае является линейной и приводит лишь к изменению масштаба частотной оси.

Преобразование ФНЧ в ФВЧ удобно производить с помощью преобразующей функции

$$f_s(s) = \frac{\omega_0}{s}, \quad (9.24)$$

что соответствует частотному преобразованию

$$f_\omega(s) = -\frac{\omega_0}{\omega}$$

где ω_0 – частота среза.

Преобразование ФНЧ в полосовой фильтр проводится с помощью более сложного преобразования вида

$$f_\omega(s) = \frac{\omega^2 - \omega_0^2}{\omega \Delta \omega}, \quad (9.25)$$

где $\Delta \omega$ - ширина полосы пропускания,

ω_0 - центральная частота полосы пропускания.

Преобразование ФНЧ в режекторный фильтр проводится аналогично преобразованию в полосовой фильтр по выражению

$$f_\omega(s) = \frac{\omega \Delta \omega}{\omega^2 - \omega_0^2}, \quad (9.26)$$

где $\Delta \omega$ - ширина полосы задержания,

ω_0 - центральная частота полосы задержания.

Все уравнения преобразований ФНЧ в фильтр заданного типа(9.23) – (9.26) приведены в нашем курсе только как демонстрация существующей возможности таких преобразований и не являются предметом глубокого рассмотрения.

Частотные преобразования фильтров, как и синтез цифровых эквивалентов для аналоговых прототипов удобно проводить с использованием специализированных компьютерных программ. Большое количество удобных инструментов для этой цели содержится, например, в программе MATLAB.

Пример 9.2.

Рассмотрим пример нахождения частотной характеристики ФНЧ-прототипа, преобразование ее к другим типам фильтров и синтез соответствующего цифрового фильтра эквивалента с использованием программы MATLAB. Листинг рабочей сессии представлен на рис.9.11.

```
% Расчет характеристики НЧ фильтра-прототипа
% Для расчета используем фильтр Баттерворта 3-го порядка
[z,p,k]=buttap(3) % сформирован аналоговый прототип на основе фильтра Б. 3-го порядка
[b,a]=zp2tf(z,p,k) % произведен пересчет модели полюсов-нулей к дробно-рациональной функции
W=tf(b,a) % задан линейный объект с передаточной функцией прототипа
p =
-0.5000 + 0.8660i
-0.5000 - 0.8660i
-1.0000
k =
1.0000
b =
```

```
0 0 0 1.0000
```

```
a =
```

```
1.0000 2.0000 2.0000 1.0000
```

```
W =
```

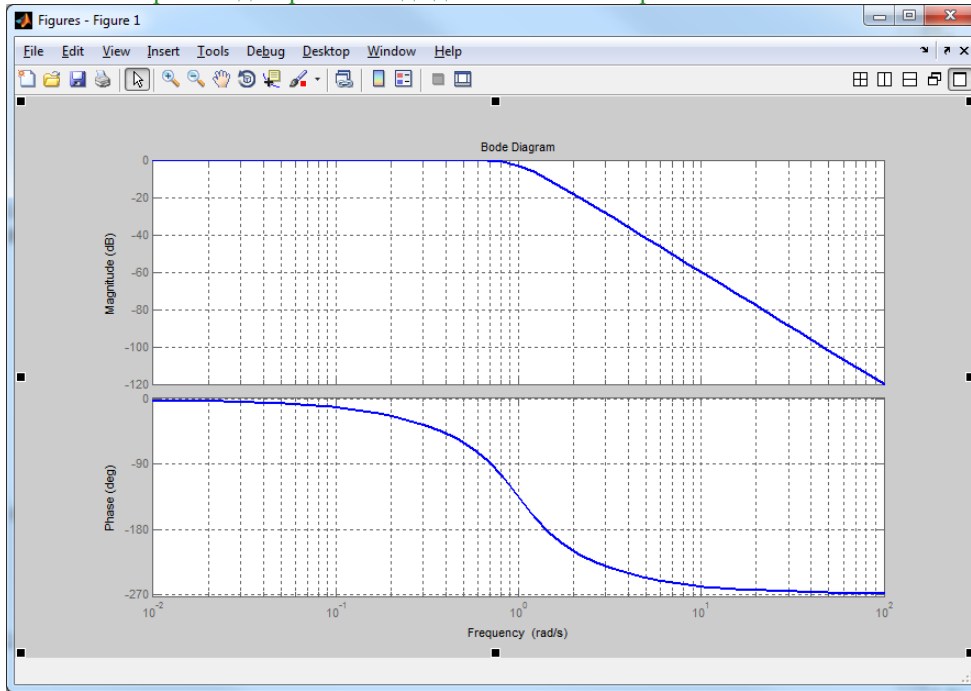
```
1
```

```
s^3 + 2 s^2 + 2 s + 1
```

Continuous-time transfer function.

```
bode(W)
```

% построение диаграммы Боде для аналогового прототипа



```
F0=1000
```

% задание новой частоты среза

```
[b1,a1]=lp2lp(b,a,2*pi*F0) % произведен сдвиг частоты среза в аналоговом прототипе
```

```
W1=tf(b1,a1) % создан объект с передаточной функцией для НЧ фильтра с новой частотой среза
```

```
bode(W1)
```

% построение диаграммы Боде для аналогового прототипа со сдвигом частоты среза

```
F0 =
```

```
1000
```

```
b1 =
```

```
2.4805e+11
```

```
a1 =
```

```
1.0e+11 *
```

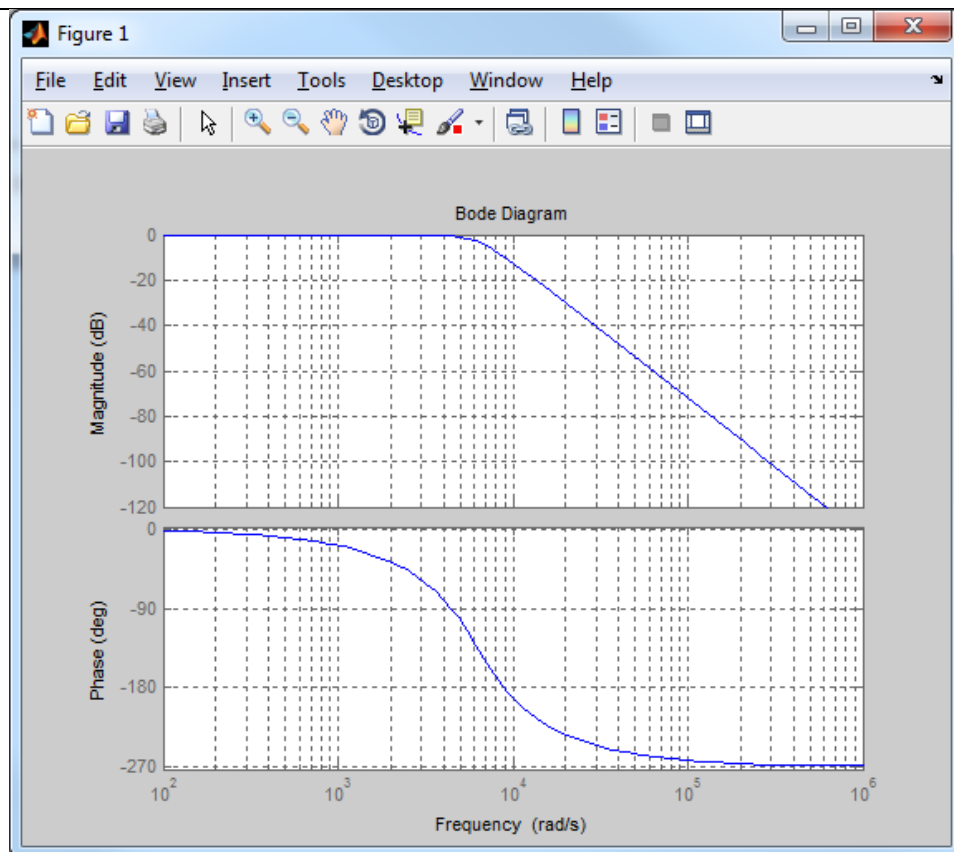
```
0.0000 0.0000 0.0008 2.4805
```

```
W1 =
```

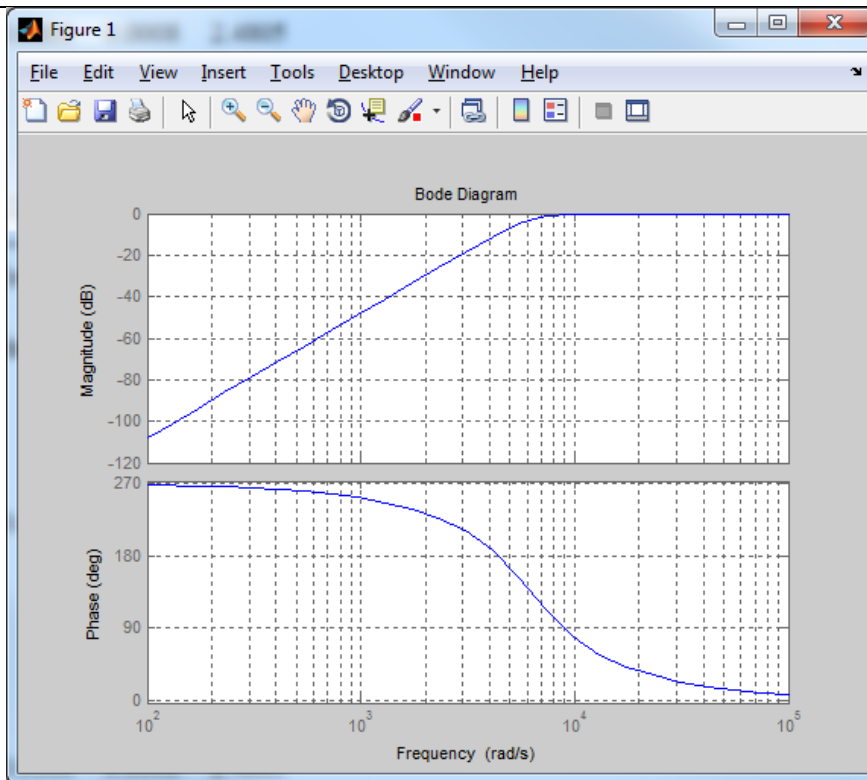
```
2.481e11
```

```
s^3 + 1.257e04 s^2 + 7.896e07 s + 2.481e11
```

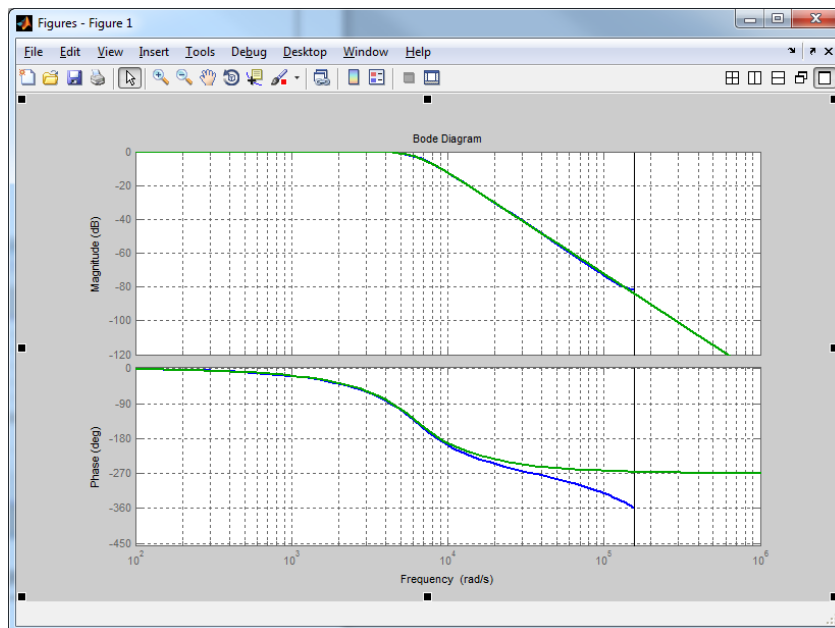
Continuous-time transfer function.



```
[b2,a2]=lp2hp(b,a,2*pi*F0) % преобразование аналогового ФНЧ в аналоговый ФВЧ
W2=tf(b2,a2) % создан объект с передаточной функцией для НЧ фильтра с новой частотой среза
bode(W2) % построение диаграммы Бode для полученного ФВЧ
b2 =
    1.0000 -0.0000 -0.0000    0
a2 =
    1.0e+11 *
    0.0000  0.0000  0.0008  2.4805
W2 =
    s^3 - 1.819e-12 s^2 - 1.143e-08 s
    -----
    s^3 + 1.257e04 s^2 + 7.896e07 s + 2.481e11
Continuous-time transfer function.
```



```
T0=0.00002 % задание интервала дискретизации
WZ1=c2d(W1,T0) % расчет дискретного эквивалента для аналоговой ПФ W1
bode(WZ1); hold on; bode(W1) % построение на одном графике частотных характеристик аналога и эквивалента
T0 =
    2.0000e-05
WZ1 =
    0.0003105 z^2 + 0.001166 z + 0.0002738
-----
    z^3 - 2.749 z^2 + 2.529 z - 0.7778
Sample time: 2e-05 seconds
Discrete-time transfer function.
```



```
WZ2=c2d(W2,T0) % расчет дискретного эквивалента для аналоговой ПФ W2
bode(WZ2); hold on; bode(W2) % построение на одном графике частотных характеристик аналога и эквивалента
WZ2 =
```

$$z^3 - 2.985 z^2 + 2.971 z - 0.9858$$

$$z^3 - 2.749 z^2 + 2.529 z - 0.7778$$

Sample time: 2e-05 seconds

Discrete-time transfer function.

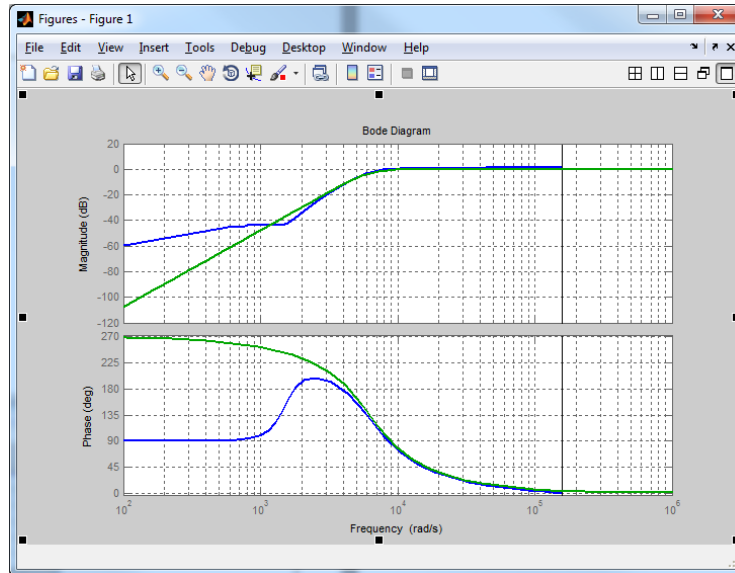


Рис.9.11. Моделирование характеристик ФНЧ-прототипа и его цифрового эквивалента

Пример 9.3.

На рис.9.12 представлен результат моделирования режекторного фильтра и его применение для фильтрации голосового сигнала, искаженного гармонической помехой известной частоты. На этом же рисунке приведена АЧХ рассчитанного режекторного фильтра.

```
Fs=44100;
s= repmat(GR822,3,1);
subplot(3,1,1);
plot(s);
title('Исходный голосовой сигнал');
xlabel('Отсчеты времени'); ylabel('Величина сигнала');
soundsc(s,44100);
disp('Произведен вывод исходного голосового сигнала. ');
disp('Для продолжения нажмите любую клавишу после окончания воспроизведения звука');
disp(' ');
pause;
%noise=randn(length(s),1);
for t=1:length(s)
sn(t)=s(t)+10.0*sin(2*pi*500*(t-1)*1/Fs);
end;
subplot(3,1,2); plot(sn);
title('Сигнал, искаженный шумом');
xlabel('Отсчеты времени'); ylabel('Величина сигнала');
soundsc(sn,44100);
disp('Произведен вывод искаженного голосового сигнала. ');
disp('Для продолжения нажмите любую клавишу после окончания воспроизведения звука');
disp(' ');
pause;
d = designfilt('bandstopiir','FilterOrder',4, ...
'HalfPowerFrequency1',490,'HalfPowerFrequency2',510, ...
'DesignMethod','butter','SampleRate',Fs);
```

```

e=filtfilt(d,sn);
%fvtool(d,Fs);
subplot(3,1,3); plot(e);
title('Сигнал после режекторной фильтрации ');
xlabel('Отсчеты времени'); ylabel('Величина сигнала');
ylim([-0.5,0.5]);
sound(e,44100);
pause;
fvtool(d);
    
```

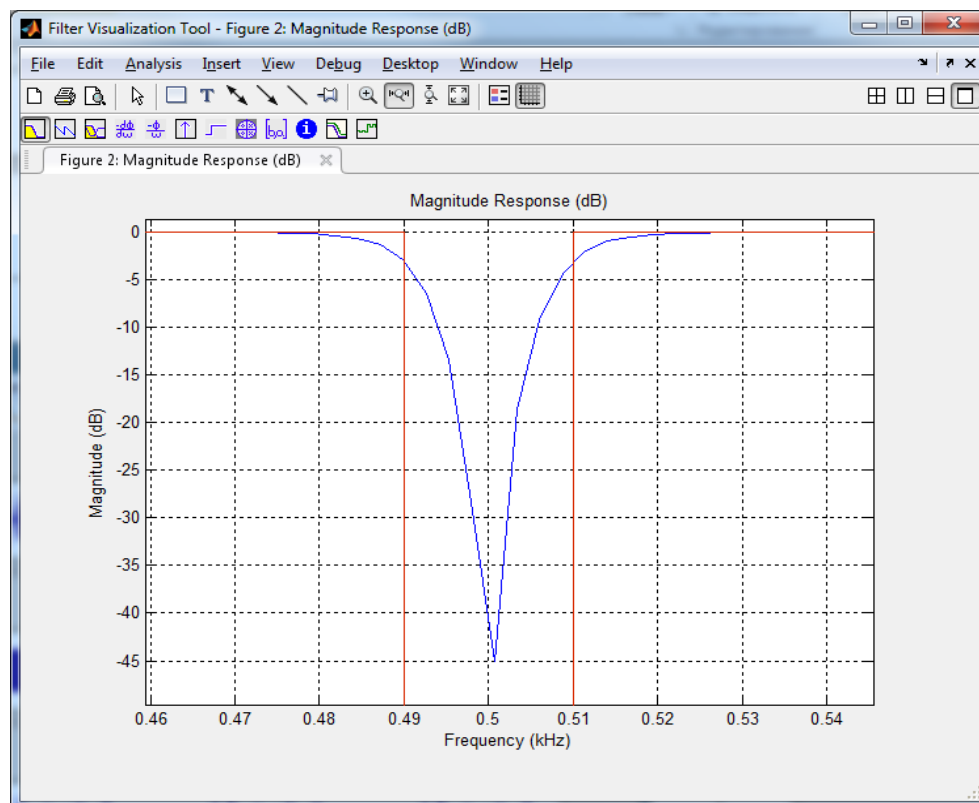
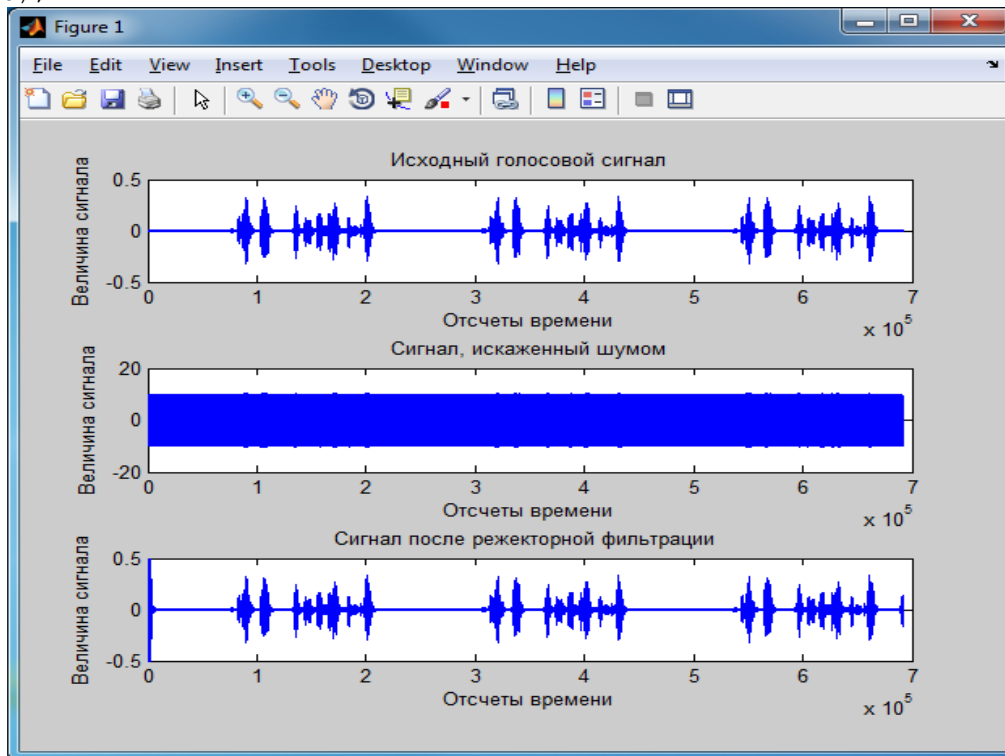


Рис.9.12. Фильтрация речевого сигнала на фоне гармонической помехи

9.6. Оценивание параметров сигналов по данным экспериментальных наблюдений

9.6.1. Принцип максимального правдоподобия и метод наименьших квадратов

Оценивание параметров различных процессов и отдельных величин составляет основу всех научных и большинства практических инженерных задач.

Базовым принципом для получения таких оценок в условиях присутствия случайных воздействий является метод (или принцип) максимального правдоподобия. Суть метода состоит в построении функции правдоподобия, зависящую от статистических характеристик оцениваемой величины, в которую оцениваемый параметр входит в качестве аргумента. Далее находится максимум этой функции правдоподобия, который соответствует наиболее вероятному значению оцениваемого параметра.

Продемонстрируем этот метод на простом примере.

Допустим, что случайная величина x имеет непрерывное распределение плотности вероятности $p(x|a)$, где a единственный параметр, подлежащий оцениванию. Тогда для выборки объема N при условии независимости каждого испытания плотность вероятности имеет вид

$$p(x_1, x_2, \dots, x_N) = \prod_{i=1}^N p(x_i|a) d x_i. \quad (9.27)$$

Если наблюдения x_i получены (следовательно, известен набор наблюдений- $\mathbf{x}$), а величина a – неизвестна, то можно ввести функцию правдоподобия

$$l(a|\mathbf{x}) = \prod_{i=1}^N p(x_i|a) \quad (9.28)$$

или логарифмическую функцию правдоподобия

$$L(a|\mathbf{x}) = \sum_{i=1}^N \ln p(x_i|a). \quad (9.29)$$

Рассмотрим простейший случай (9.27) – (9.29), когда величина x распределена по нормальному закону с единичной дисперсией и неизвестным математическим ожиданием, то есть

$$p(x|\mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}}. \quad (9.30)$$

Единственный параметр μ требуется оценить по выборке объемом N значений x . Опуская множитель $(2\pi)^{-N/2}$, функция правдоподобия в данном случае может быть записана в простом виде

$$l(\mu|\mathbf{x}) = e^{-\sum_{i=1}^N \frac{(x_i - \mu)^2}{2}}.$$

Соответственно логарифмическая функция правдоподобия представляется как

$$L(\mu|\mathbf{x}) = -\sum_{i=1}^N \frac{(x_i - \mu)^2}{2}. \quad (9.31)$$

Полученная в (9.31) функция $L(\mu)$ представляет собой квадратичную параболу по переменной μ для любого значения N и достигает максимума при таком значении μ , когда

$$\sum_{i=1}^N \frac{(x_i - \mu)^2}{2} \rightarrow \min.$$

Откуда получаем хорошо известный результат для оценки математического ожидания по выборке случайной величины

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^N x_i$$

Для случая неизвестной дисперсии параметров, подлежащих определению, становится два, и плотность распределения записывается как

$$p(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Соответственно функция правдоподобия принимает вид

$$l(\mu, \sigma^2|x) = (\sigma^2)^{-\frac{N}{2}} e^{-\frac{\sum_i (x_i - \mu)^2}{2\sigma^2}}$$

и логарифмическая функция правдоподобия записывается как

$$L(\mu, \sigma^2|x) = -\frac{N}{2} \ln \sigma^2 - \sum_{i=1}^N \frac{(x_i - \mu)^2}{2\sigma^2}.$$

Теперь необходимо найти такую пару значений (μ, σ^2) , которая с наибольшей вероятностью соответствовала бы имеющейся выборке наблюдений. Эта задача решается составлением системы из двух уравнений

$$\begin{cases} \frac{\partial}{\partial \mu} L(\mu, \sigma^2|x) = 0, \\ \frac{\partial}{\partial \sigma^2} L(\mu, \sigma^2|x) = 0. \end{cases}$$

Решение этой системы уравнений приводит к следующим результатам.

$$\frac{\partial}{\partial \mu} L(\mu, \sigma^2|x) = \frac{1}{\sigma^2} \sum_{i=1}^N (x_i - \mu) = 0,$$

откуда

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^N x_i. \quad (9.32)$$

Далее

$$\frac{\partial}{\partial \sigma^2} L(\mu, \sigma^2|x) = -\frac{N}{2} \frac{1}{\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^N (x_i - \mu)^2,$$

что дает

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \hat{\mu})^2, \quad (9.33)$$

Полученная таким образом оценка дисперсии является несколько смещенной, так как

$$M\{\hat{\sigma}^2\} = \frac{N-1}{N} \sigma^2,$$

поэтому на практике для оценки выборочной дисперсии чаще используется формула, учитывающее данное смещение

$$\hat{\sigma}^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \hat{\mu})^2.$$

В случае увеличения количества оцениваемых параметров и изменения вида закона распределения плотности вероятности оцениваемых величин, формулы для вычисления и процедура их вывода могут существенно усложниться, но принцип получения уравнений для оцениваемых параметров (9.32) и (9.33) остается таким же.

9.6.2. Метод наименьших квадратов

В инженерной практике наиболее широкое распространение получил оценивания параметров, называемый методом наименьших квадратов. История возникновения данного подхода началась с работ Лежандра в 1805 году и Гаусса в 1809 году, посвященным методам определения орбит комет. Хотя Гаусс оспаривал приоритет использования данного метода начиная еще с 1795 года. Значительный вклад в развитие метода внесли Лаплас, Коши, Бессель и Энке. В наиболее общем виде теоремы, касающиеся данного метода были доказаны А.Марковым в 1912 году, а в 1934 году Эйткен расширил результаты на случай коррелированных наблюдений с различными дисперсиями.

Общую идею метода удобно продемонстрировать на основе задачи регрессии, линейной относительно оцениваемых параметров.

Пусть некоторая функция $Q(t)$ является полиномиальной суммой вида

$$Q(t) = a_0 + a_1x + a_2x^2 + \cdots + a_{p-1}x^{p-1}, \quad (9.34)$$

а наблюдения за функцией $Q(t)$ представляют собой

$$y_i = Q(x_i) + \varepsilon_i,$$

где ε_i – ошибки наблюдения, представляющие собой случайные величины с одинаковыми дисперсиями, причем распределение ε_i не обязательно имеет нормальный характер. Проведя N наблюдений и получив N пар (x_i, y_i) , можно в соответствии с (9.34) связать их следующими уравнениями

$$\begin{cases} \varepsilon_1 = y_1 - a_0 + a_1 x_1 + a_2 x_1^2 + \cdots + a_{p-1} x_1^{p-1} \\ \varepsilon_2 = y_2 - a_0 + a_1 x_2 + a_2 x_2^2 + \cdots + a_{p-1} x_2^{p-1} \\ \vdots \\ \varepsilon_N = y_N - a_0 + a_1 x_N + a_2 x_N^2 + \cdots + a_{p-1} x_N^{p-1} \end{cases} \quad (9.35)$$

Оценку неизвестных параметров $a_1, k = 0, 1, \dots, p$ из (9.35) произведем таким образом, чтобы сумма квадратов ошибок была минимальна, то есть

$$S = \sum_{i=1}^N \varepsilon_i^2 = \sum_{i=1}^N (y_i - a_0 + a_1 x_i + a_2 x_i^2 + \dots + a_{p-1} x_i^{p-1})^2 \rightarrow \min.$$

Экстремальное значение S можно найти путем решения системы линейных уравнений

[illegible]

связана с дисперсией последовательности y_i соотношением

$$M\{R\} = (N - p)\sigma^2.$$

При этом оценку дисперсии можно получить по выражению

$$\hat{\sigma}^2 = \frac{R}{N - p}.$$

Полученные формулы легко обобщаются на случай когда вместо полиномиального представления

$$Q(t) = a_0 + a_1x + a_2x^2 + \dots + a_px^{p-1}$$

используется представление по системе функций в виде

$$Q(t) = a_0f_0(x) + a_1f_1(x) + a_2f_2(x) + \dots + a_{p-1}f_{p-1}(x).$$

Поскольку выделение константы в данном случае не является принципиальным и член $a_0f_0(x)$ по своей сути ничем не отличается от остальных слагаемых, то полагая, что модель p -го порядка содержит p –параметров, без потери общности матрицу $\mathbf{X}$ можно задать в виде

$$\mathbf{X} = \begin{bmatrix} f_0(x_1) & f_1(x_1) & f_2(x_1) & \cdot & \cdot & f_{p-1}(x_1) \\ f_0(x_2) & f_1(x_2) & f_2(x_2) & \cdot & \cdot & f_{p-1}(x_2) \\ f_0(x_3) & f_1(x_3) & f_2(x_3) & \cdot & \cdot & f_{p-1}(x_3) \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ f_0(x_N) & f_1(x_N) & f_2(x_N) & \cdot & \cdot & f_{p-1}(x_N) \end{bmatrix}.$$

9.6.3. Рекурсивный метод наименьших квадратов

Полученное в предыдущем разделе уравнение (9.38) для оценки параметров по методу наименьших квадратов

$$\hat{\mathbf{A}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

может быть использовано только для одновременной обработки всего набора наблюдаемых данных.

На практике часто бывает желательно обновлять оценку параметров процесса по мере поступления новых наблюдений. Эта задача может быть решена путем применения рекурсивного варианта метода наименьших квадратов.

Допустим, что по имеющемуся набору наблюдений $y_k, k = 1, \dots, N$, была получена оценка параметров в виде

$$\hat{\mathbf{A}}_N = (\mathbf{X}_N'\mathbf{X}_N)^{-1}\mathbf{X}_N'\mathbf{y}_N \quad (9.39)$$

При получении нового наблюдения y_{N+1} к вектору $\mathbf{y}$ добавляется один элемент, а к матрице $\mathbf{X}$ – одна строка, то есть

$$\mathbf{X}_{N+1} = \begin{bmatrix} \mathbf{X}_N \\ \mathbf{x}_{N+1} \end{bmatrix}, \mathbf{y}_{N+1} = \begin{bmatrix} \mathbf{y}_N \\ y_{N+1} \end{bmatrix} \quad (9.40)$$

Тогда

$$\hat{\mathbf{A}}_{N+1} = [\mathbf{X}_N'\mathbf{X}_N + \mathbf{x}_{N+1}'\mathbf{x}_{N+1}]^{-1}[\mathbf{X}_N'\mathbf{y}_N + \mathbf{x}_{N+1}'y_{N+1}]. \quad (9.41)$$

Из математики известна лемма об обращении матрицы

$$[A + BCD]^{-1} = A^{-1} - A^{-1}B[C^{-1} + DA^{-1}B]DA^{-1}, \quad (9.42)$$

справедливая для квадратных невырожденных матриц.

Тогда из (9.39) – (9.42) следует, что

$$[X_N'X_N + x_{N+1}'x_{N+1}]^{-1} = [X_N'X_N]^{-1} - [X_N'X_N]^{-1}x_{N+1}'[1 + x_{N+1}[X_N'X_N]^{-1}x_{N+1}]^{-1}x_{N+1}[X_N'X_N]^{-1}$$

Введем матрицу

$$P_N = [X_N'X_N]^{-1}.$$

Тогда в краткой форме можно представить

$$P_{N+1} = P_N - P_N x_{N+1}' [1 + x_{N+1} P_N x_{N+1}]^{-1} x_{N+1} P_N$$

Поскольку

$$[X_{N+1}'X_{N+1}]\hat{A}_{N+1} = X_{N+1}'y_{N+1}$$

то

$$\begin{aligned} P_{N+1}^{-1}\hat{A}_{N+1} &= X_{N+1}'y_{N+1} = X_N'y_N + x_{N+1}'y_{N+1} = \\ &= X_N'X_N\hat{A}_N + x_{N+1}'y_{N+1} = \\ &= [X_{N+1}'X_{N+1} - x_{N+1}'x_{N+1}]\hat{A}_N + x_{N+1}'y_{N+1} = \\ &= [X_{N+1}'X_{N+1}]\hat{A}_N - [x_{N+1}'x_{N+1}]\hat{A}_N + x_{N+1}'y_{N+1} = \\ &= P_{N+1}^{-1}\hat{A}_N + x_{N+1}'[y_{N+1} - x_{N+1}\hat{A}_N] \end{aligned}$$

Или

$$\hat{A}_{N+1} = \hat{A}_N + K_{N+1}[y_{N+1} - x_{N+1}\hat{A}_N],$$

где $K_{N+1} = P_{N+1}x_{N+1}'$

Так как

$$X_{N+1}'X_{N+1} = X_N'X_N + x_{N+1}'x_{N+1},$$

то

$$P_{N+1}^{-1} = P_N^{-1} + x_{N+1}'x_{N+1}.$$

Откуда

$$P_N = P_{N+1} + P_{N+1}x_{N+1}'x_{N+1}P_N \quad (9.43)$$

или

$$P_N x_{N+1}' = P_{N+1} x_{N+1}' [I + x_{N+1} P_N x_{N+1}].$$

Следовательно,

$$K_{N+1} = P_{N+1}x_{N+1}' = P_N x_{N+1}' [I + x_{N+1} P_N x_{N+1}]^{-1} \quad (9.44)$$

Полученные таким образом уравнения (9.43), (9.44) полностью определяют рекурсивный алгоритм наименьших квадратов, для инициализации которого можно выбрать, например, значения

$$\hat{A}_0 = \mathbf{0} \text{ и } P_0 = \alpha I,$$

где α – достаточно большое положительное число.

9.6.4. Оценивание параметров случайных сигналов

Напомним, что при рассмотрении параметрического представления случайных сигналов в форме временных рядов, были представлены три основные модели: АР-модель (модель авторегрессии), СС-модель (модель скользящего среднего) и АРСС-модель (модель авторегрессии скользящего среднего).

Представление случайного сигнала в бесконечной рекурсивной форме имеет вид

$$\begin{aligned} x(k) &= a_1 x(k-1) + a_2 x(k-2) + \dots + a_i x(k-i) + \dots + e(k) = \\ &= \sum_{i=1}^{\infty} a_i x(k-i) + e(k). \end{aligned} \quad (9.45)$$

Бесконечная нерекурсивная форма может быть записана как

$$\begin{aligned} x(k) &= b_0 e(k) + b_1 e(k-1) + \dots + b_i e(k-i) + \dots = \\ &= \sum_{i=0}^{\infty} b_i e(k-i). \end{aligned} \quad (9.46)$$

Использование конечного числа в суммах данных бесконечных рядов (9.45), (9.46) приводят к трем наиболее распространенным моделям случайных процессов в виде временных рядов.

Авторегрессионная модель (АР-модель) имеет вид

$$x(k) = \sum_{i=1}^p a_i x(k-i) + e(k), \quad (9.47)$$

где p – порядок модели,

a_i – параметры модели.

Модель скользящего среднего (СС-модель) имеет вид

$$x(k) = \sum_{i=0}^q b_i e(k-i), \quad (9.48)$$

где q – порядок модели,

b_i – параметры модели.

Объединение АР и СС моделей (9.47), (9.48) приводит к форме, получившей название **авторегрессии скользящего среднего** (АРСС), которая имеет следующий вид

$$x(k) = \sum_{i=1}^p a_i x(k-i) + \sum_{i=0}^q b_i e(k-i), \quad (9.49)$$

где p, q – порядки соответствующих частей модели,

a_i, b_i – параметры модели.

На примере модели авторегрессии (9.47) рассмотрим процесс оценивания параметров с использованием метода наименьших квадратов.

Для процесса, описываемого стационарной АР-моделью, запишем уравнение

$$x(k) = a_1 x(k-1) + a_2 x(k-2) + \dots + a_p x(k-p) + e(k),$$

где $a_k, k = 1, \dots, p$ – параметры модели, значения которых требуется оценить по имеющему набору из N наблюдений $x(0), x(1), \dots, x(N-1)$.

Умножим левую и правую часть этого уравнения на величину $x(k-m)$ и получим

$$\begin{aligned} x(k)x(k-m) &= a_1 x(k-1)x(k-m) + a_2 x(k-2)x(k-m) + \\ &+ \dots + a_p x(k-i)x(k-m) + \dots + e(k)x(k-m). \end{aligned}$$

Переходя в полученном уравнении к математическим ожиданиям имеем

$$R(m) = a_1 R(m-1) + a_2 R(m-2) + \dots + a_p R(m-p),$$

где $R(m)$ – автокорреляционная функция процесса $x(k)$

для $m > 0$.

и

$$R(0) = a_1 R(-1) + a_2 R(-2) + \dots + a_p R(-p) + \sigma^2,$$

где σ^2 – дисперсия процесса $e(k)$.

Таким образом, изменяя m в пределах от 0 до p получаем систему уравнений

$$\begin{bmatrix} R(0) & R(-1) & \dots & R(-p) \\ R(1) & R(0) & \dots & R(-p+1) \\ \vdots & \vdots & \ddots & \vdots \\ R(p) & R(p-1) & \dots & R(0) \end{bmatrix} \begin{bmatrix} 1 \\ -a_1 \\ \vdots \\ -a_p \end{bmatrix} = \begin{bmatrix} \sigma^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \quad (9.50)$$

Данную систему уравнений (9.50) называют уравнениями Юла-Уокера.

Эффективный способ решения данного уравнения, основанный на использовании свойства четности автокорреляционной функции ($R(n) = R(-n)$) и теплицевой структуре матрицы, входящей в уравнение Юла-Уокера, был разработан Левинсоном и модернизирован Дурбиным, поэтому лн носит название алгоритма Левинсона-Дурбина.

Этот алгоритм позволяет рекурсивно вычислять наборы параметров для моделей возрастающего порядка

$$\begin{aligned} & \{a_{11}, \sigma_1^2\}, \\ & \{a_{21}, a_{22}, \sigma_2^2\} \\ & \dots \dots \dots \dots \dots \dots \\ & \{a_{p1}, a_{p2}, \dots, a_{pp}, \sigma_p^2\}, \end{aligned}$$

последний из которых является искомым решением.

Данный алгоритм иницируется значениями параметров для модели первого порядка вида

$$a_{11} = \frac{R(1)}{R(0)}, \sigma_1^2 = (1 - a_{11}^2)R(0).$$

Далее рекурсии в алгоритме Левинсона-Дурбина выполняются в виде

$$\begin{aligned} a_{kk} &= \left[R(k) - \sum_{l=1}^{k-1} a_{k-1,l} R(k-l) \right] \sigma_{k-1}^2, \\ a_{ki} &= a_{k-1,i} - a_{kk} a_{k-1,k-i}, \\ \sigma_k^2 &= (1 - a_{kk}^2), \end{aligned}$$

для $k = 2, 3, \dots, p$.

Полученный алгоритм позволяет вычислить параметры **АР-модели** по известным значениям автокорреляционной функции случайного процесса. Однако, как отмечалось выше, в нашем распоряжении таких данных нет, а есть только набор из N наблюдений $\{x(0), x(1), \dots, x(N-1)\}$.

В этом случае вместо точных значений автокорреляционной функции обычно используется ее оценки, получаемые по выражению

$$\hat{R}(k) = \frac{1}{N-k} \sum_{n=0}^{N-k-1} x(k)x(k+n).$$

Далее полученные оценки автокорреляционной функции могут быть использованы в алгоритме Левинсона-Дурбина для получения оценок $\{\hat{a}_{p1}, \hat{a}_{p2}, \dots, \hat{a}_{pp}, \hat{\sigma}_p^2\}$. При этом понятно, искомые оценки могут быть получены только после совместной обработки всего набора наблюдений $\{x(0), x(1), \dots, x(N-1)\}$. Поэтому такой метод получения оценок относят к методам блочной обработки данных.

В процессе выполнения рекурсий Левинсона-Дурбина получаются параметры всех аппроксимирующих АР-моделей меньшего порядка. Это свойство крайне удобно с практической точки зрения, так как искомый порядок модели, как правило, заранее неизвестен. Тогда алгоритм можно продолжать выполнять для моделей возрастающего порядка, пока дисперсия σ_k^2 не уменьшится до приемлемой величины.

Поскольку порядок модели заранее неизвестен, то на практике приходится использовать несколько моделей различного порядка, а затем по какому-либо критерию выбирать лучшую из них.

Для выбора порядка АР-модели часто используются критерии финальной ошибки прогнозирования (ФОП)

(9.51)

и информационный критерий Акаике

$$\text{ИКА}(p) = \ln \sigma_p^2 + \frac{2(p+1)}{N}, \quad (9.52)$$

где N – объем выборки, используемый для получения оценок АР-модели.

Выбор оптимального значения порядка модели $p_{\text{опт}}$ осуществляется по минимуму одного из критериев (9.51), (9.52). Чаще всего оба критерия приводят к совпадающим значениям оптимального значения порядка модели. В случае расхождения результатов их применения требуется привлечение дополнительных условий для принятия соответствующего решения.

Помимо рассмотренного метода блочной обработки данных наблюдений можно построить рекурсивные алгоритмы оценки параметров, осуществляющие последовательную обработку получаемых наблюдений случайного процесса.

Общая теория таких методов достаточно сложна и не входит в задачи нашего курса, поэтому ограничимся рассмотрением лишь общих аспектов некоторых вариантов последовательной обработки данных наблюдения.

Перепишем исходную АР-модель в виде

$$x(k) - a_1 x(k-1) - a_2 x(k-2) - \dots - a_p x(k-p) = e(k)$$

или

$$(1 - a_1 z^{-1} - a_2 z^{-2} - \dots - a_p z^{-p})x(k) = e(k).$$

Структурная схема, соответствующая такому представлению АР-модели приведена на рис.9.13.

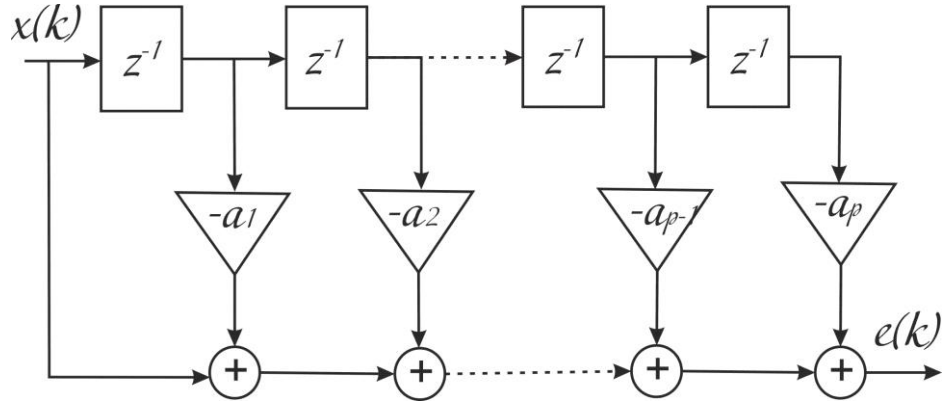


Рис.9.13. Структурная схема АР-модели

Запишем данную модель в виде

$$x(k) = \boldsymbol{\vartheta}(k)\mathbf{A} + e(k),$$

где

$$\mathbf{A} = [-a_1, -a_2, \dots, -a_p]'$$

$$\boldsymbol{\vartheta}(k) = [x(k-1), x(k-2), \dots, x(k-p)].$$

Тогда, применяя к этому уравнению, полученные ранее выражения для рекурсивного метода наименьших квадратов (9.43), (9.44), запишем

$$\begin{aligned} \hat{\mathbf{A}}_{N+1} &= \hat{\mathbf{A}}_N + \mathbf{K}_{N+1}[x(k+1) - \boldsymbol{\vartheta}_{N+1}\hat{\mathbf{A}}_N] \\ \mathbf{K}_{N+1} &= \mathbf{P}_N \boldsymbol{\vartheta}'_{N+1} [1 + \boldsymbol{\vartheta}_{N+1} \mathbf{P}_N \boldsymbol{\vartheta}'_{N+1}]^{-1} \end{aligned} \quad (9.53)$$

$$P_{N+1} = P_N - P_N \vartheta'_{N+1} [1 + \vartheta_{N+1} P_N \vartheta'_{N+1}] \vartheta_{N+1} P_N$$

с начальными условиями $\hat{A}_0 = 0$, $P_0 = 0$.

Заметим, что выражения в квадратных скобках являются скалярами, поэтому в данном случае нет необходимости обращать матрицы, что существенно упрощает вычислительную процедуру.

Рассмотренный алгоритм придает одинаковой вес ошибкам на всех тактах вычисления, что в случае изменяющихся характеристик сигнала может снижать эффективность работы алгоритма. Для отслеживания изменений характеристик оцениваемой величины используют так называемое экспоненциальное забывание, придающее различный вес ошибкам, полученным в различное время.

В этом случае основные формулы алгоритма принимают вид

$$\begin{aligned} K_{N+1} &= P_N \vartheta'_{N+1} [\lambda + \vartheta_{N+1} P_N \vartheta'_{N+1}]^{-1} \\ P_{N+1} &= \frac{1}{\lambda} [P_N - P_N \vartheta'_{N+1} [\lambda + \vartheta_{N+1} P_N \vartheta'_{N+1}] \vartheta_{N+1} P_N], \end{aligned} \quad (9.54)$$

где λ – коэффициент забывания.

Помимо рассмотренных рекурсивных методов оценки параметров АР-модели (9.53), (9.54) на практике часто используют алгоритмы, основанные на вычислении градиента функции квадрата ошибки.

В настоящее время предложено несколько алгоритмов данного типа, различающихся только способом оценки градиента.

В общем случае

$$\frac{\partial e^2(k)}{\partial A} = 2e(k) \frac{\partial e(k)}{\partial A} = -2e(k) \vartheta'(k).$$

Тогда формула для обновления оценки вектора параметров может быть записана в виде

$$\hat{A}(k) = \hat{A}(k-1) + \mu \left[-\frac{\partial e^2(k)}{\partial A} \right] = \hat{A}(k-1) + 2\mu e(k) \vartheta'(k), \quad (9.55)$$

где μ – шаг алгоритма.

Каждый коэффициент фильтра в этом случае корректируется по простой формуле

$$\begin{aligned} \hat{a}_i(k) &= \hat{a}_i(k-1) + 2\mu e(k) x(k), \\ e(k) &= x(k) - \sum_{n=1}^p \hat{a}_i(k-1) x(k-n). \end{aligned}$$

Для обеспечения сходимости алгоритма шаг выбирается из условия

$$0 < \mu < \lambda_{max},$$

где λ_{max} – максимальное собственное число корреляционной матрицы процесса $x(k)$.

Естественно, что для практического использования такой критерий выбора не очень удобен, поэтому для удобства часто вводится дополнительный параметр α такой, что

$$\mu = \frac{\alpha}{p\sigma_x^2}.$$

оценивание дисперсии процесса в этом случае проводится в виде

$$\sigma_x^2(k) = (1 - \alpha) \sigma_x^2(k-1) + \alpha x^2(k).$$

Применение в алгоритме (9.55) подобного экспоненциального забывания предыдущих данных приводит к снижению скорости сходимости, но его простота и низкая вычислительная сложность несомненны.

Модели скользящего среднего и авторегрессии скользящего среднего, к сожалению, приводят к более сложным с точки зрения организации эффективной вычислительной процедуры результатам.

Так для СС-модели (9.48)

$$x(k) = \sum_{i=0}^q b_i e(k-i)$$

справедливо

$$\begin{cases} R(k) = \sum_{j=0}^q b_{j-k} b_j & k = 1, 2, \dots, q \\ R(k) = 0, & k > q, \end{cases}$$

где $R(k)$ - автокорреляционная функция процесса.

Из этого уравнения видно, что для $k > q$ автокорреляционная функция обращается в ноль, а при всех других значениях k , связь между параметрами модели и значениями корреляционной функции является нелинейной.

Для АРСС-модели (9.49)

$$x(k) = \sum_{i=1}^p a_i x(k-i) + \sum_{i=0}^q b_i e(k-i)$$

аналогичным образом нетрудно получить уравнения связи параметров модели и значений автокорреляционной функции в виде

$$\begin{cases} R_{xx}(m) = \sum_{k=1}^p a_k R_{xx}(m-k) + \sum_{k=0}^q b_k R_{ex}(m-k), & m = 0, 1, \dots, q \\ R_{xx}(m) = \sum_{k=1}^p a_k R_{xx}(m-k), & m = q+1, q+2, \dots \end{cases}$$

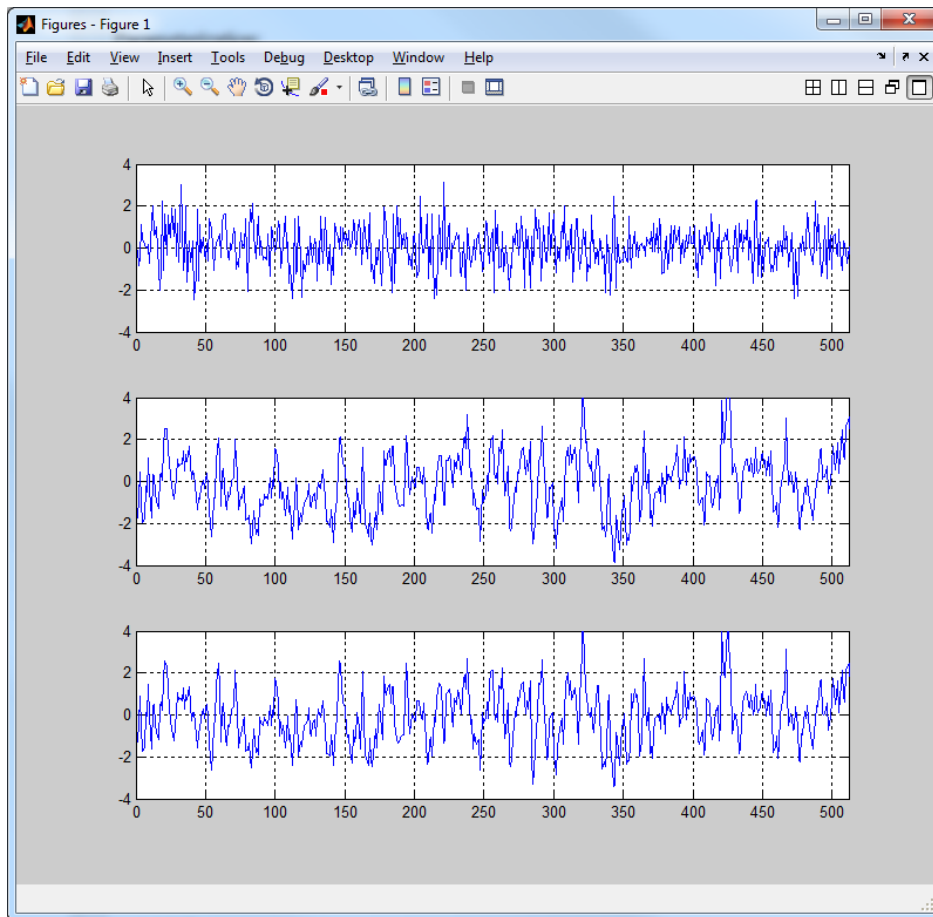
Приведенное уравнение показывает, что в случае АРСС-модели для расчета параметров необходимо учитывать и взаимную корреляцию процессов $x(t)$ и $e(t)$.

На сегодняшний день разработано достаточно большое количество алгоритмов оценивания параметров СС и АРСС-моделей. Вопросы их применения рассматриваются в специальных курсах по цифровой обработке сигналов. В рамках нашего курса подробного рассмотрения данных вопросов производиться не будет.

Пример.9.4.

Рассмотрим процедуру нахождения параметров АР-модели с применением программы MATLAB. Листинг рабочей сессии представлен на рис.9.14.

```
a1=-0.8           % Коэффициенты АР-модели
a2=0.1
n=randn(1024,1); % Формирование сигнала возбуждающего белого шума
Y1=filter(1,[1 a1],n); % Формирование выходного сигнала АР-фильтра 1-го порядка
Y2=filter(1,[1 a1 a2],n); % Формирование выходного сигнала АР-фильтра 2-го порядка
subplot(3,1,1); plot(n); % Вывод графика возбуждающего шума
subplot(3,1,2); plot(yy1); % Вывод графика выходного сигнала АР-фильтра 1-го порядка
subplot(3,1,3); plot(yy2); % Вывод графика выходного сигнала АР-фильтра 1-го порядка
A11=ar(Y1,1)      %расчет параметров АР-модели 1-го порядка для сигнала АР-фильтра 1-го
порядка
A12=ar(Y1,2)      %расчет параметров АР-модели 2-го порядка для сигнала АР-фильтра 1-го
порядка
A22=ar(Y2,2)      %расчет параметров АР-модели 2-го порядка для сигнала АР-фильтра 2-го
порядка
[aBurg,e1]=arburg(Y2,2) %расчет параметров АР-модели 2-го порядка для сигнала АР-фильтра 2-го
порядка (метод Бурга)
th=ar(Y2,2,'yw'); % расчет параметров АР-модели 2-го порядка для сигнала АР-фильтра 2-го
порядка
present(th)        % вывод параметров АР-модели 2-го порядка для сигнала АР-фильтра 2-го
порядка (метод Юла-Уокера)
```



a1 =

-0.8000

a2 =

0.1000

A11 =

Discrete-time AR model: $A(z)y(t) = e(t)$

$A(z) = 1 - 0.7923 z^{-1}$

Sample time: 1 seconds

Parameterization:

Polynomial orders: na=1

Number of free coefficients: 1

Use "polydata", "getpvec", "getcov" for parameters and their uncertainties.

Status:

Estimated using AR ('fb/now') on "Y1".

Fit to estimation data: 38.96% (prediction focus)

FPE: 0.9623, MSE: 0.9605

A12 =

Discrete-time AR model: $A(z)y(t) = e(t)$

$A(z) = 1 - 0.797 z^{-1} + 0.004744 z^{-2}$

Sample time: 1 seconds

Parameterization:

Polynomial orders: na=2

Number of free coefficients: 2

Use "polydata", "getpvec", "getcov" for parameters and their uncertainties.

Status:

Estimated using AR ('fb/now') on "Y1".

Fit to estimation data: 38.96% (prediction focus)

FPE: 0.9651, MSE: 0.9604

A22 =

Discrete-time AR model: $A(z)y(t) = e(t)$

$$A(z) = 1 - 0.7998 z^{-1} + 0.09952 z^{-2}$$

Sample time: 1 seconds

Parameterization:

Polynomial orders: na=2

Number of free coefficients: 2

Use "polydata", "getpvec", "getcov" for parameters and their uncertainties.

Status:

Estimated using AR ('fb/now') on "Y2".

Fit to estimation data: 31.65% (prediction focus)

FPE: 0.9652, MSE: 0.9605

aBurg = 1.0000 -0.7989 0.0995

e1 = 0.9613

th =

Discrete-time AR model: $A(z)y(t) = e(t)$

$$A(z) = 1 - 0.7942 (+/- 0.03101) z^{-1} + 0.09595 (+/- 0.03101) z^{-2}$$

Name: th

Sample time: 1 seconds

Parameterization:

Polynomial orders: na=2

Number of free coefficients: 2

Use "polydata", "getpvec", "getcov" for parameters and their uncertainties.

Status:

Estimated using AR ('yw/ppw') on "Y2".

Fit to estimation data: 31.65% (prediction focus)

FPE: 0.9652, MSE: 0.9605

More information in model's "Report" property.

Рис.9.14. Параметрическое оценивание АР-моделей 1-го и 2-го порядка.

Пример 9.5.

Рассмотрим пример рекурсивной оценки коэффициента АР-модели первого порядка с применением программы MATLAB. Для наглядности демонстрации алгоритмов оценивания специально выбран простейший скалярный случай с одним параметром, в котором все уравнения являются скалярными. Встроенные функции рекурсивного параметрического оценивания программы MATLAB специально не использованы, чтобы

наглядно продемонстрировать вычислительную простоту алгоритмов. Листинг рабочей сессии представлен на рис.9.15.

```
a1=-0.8;           % Задание коэффициента формирующего фильтра 1-го порядка
alf=0.002;         % Задание параметров адаптивных алгоритмов
mu=0.002;
lam1=0.99;
lam2=1;
N=1000;           % Задание длины выборки наблюдений
n=randn(N,1);      % Формирование возбуждающего сигнала
y=filter(1,[1 a1],n); % Формирование вызода формирующего фильтра
E(1)=1;           % задание начальных условий алгоритмов LMS
A1(1)=0;
A2(1)=0;
for k=2:N          E(k)=(1-alf)*E(k-1)+alf*y(k)*y(k); % Алгоритм LMS #1
                  A1(k)=A1(k-1)+2*(alf/E(k-1))*(y(k)-A1(k-1)*y(k-1))*y(k-1);
                  A2(k)=A2(k-1)+2*mu*(y(k)-A2(k-1)*y(k-1))*y(k-1); %Алгоритм LMS #2
end;
P(1)=10.;         % Задание начальных условий алгоритмам RLS
K(1)=0;
A3(1)=0;
P2(1)=10.;
K2(1)=0;
A4(1)=0;
for k=2:N          x(k)=y(k-1); % Алгоритм RLS с экспоненциальным забыванием
                  K(k)=P(k-1)*x(k)/(lam1+P(k-1)*x(k)*x(k));
                  A3(k)=A3(k-1)+K(k)*(y(k)-A3(k-1)*x(k));
                  P(k)=1/lam1*(P(k-1)-P(k-1)*x(k)*x(k)*P(k-1)/(lam1+P(k-1)*x(k)*x(k)));
end;
for k=2:N          x(k)=y(k-1); % Алгоритм RLS
                  K2(k)=P2(k-1)*x(k)/(lam2+P2(k-1)*x(k)*x(k));
                  A4(k)=A4(k-1)+K2(k)*(y(k)-A4(k-1)*x(k));
                  P2(k)=1/lam2*(P2(k-1)-P2(k-1)*x(k)*x(k)*P2(k-1)/(lam2+P2(k-1)*x(k)*x(k)));
end;
subplot(4,1,1);plot(A1); % Вывод графика оценки коэффициента AP-фильтра 1-го порядка по LMS #1
subplot(4,1,2);plot(A2); % Вывод графика оценки коэффициента AP-фильтра 1-го порядка по LMS #2
subplot(4,1,3);plot(A3); % Вывод графика оценки коэффициента AP-фильтра 1-го порядка по PLS с эксп.
забыванием
subplot(4,1,4);plot(A4); % Вывод графика оценки коэффициента AP-фильтра 1-го порядка по LMS
```

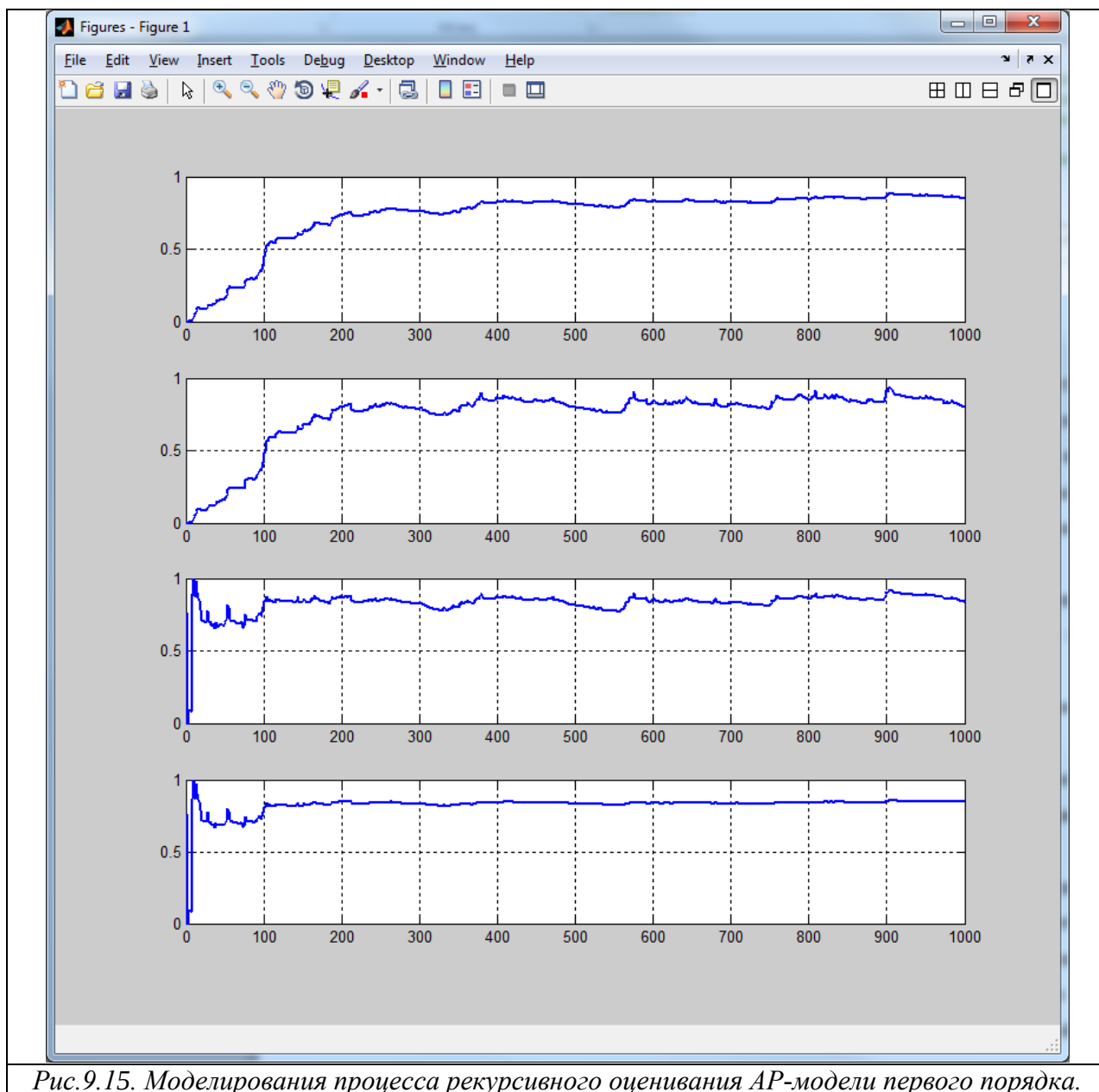


Рис.9.15. Моделирования процесса рекурсивного оценивания АР-модели первого порядка.

9.7. Основы адаптивной обработки сигналов

9.7.1. Общие сведения об адаптивных фильтрах

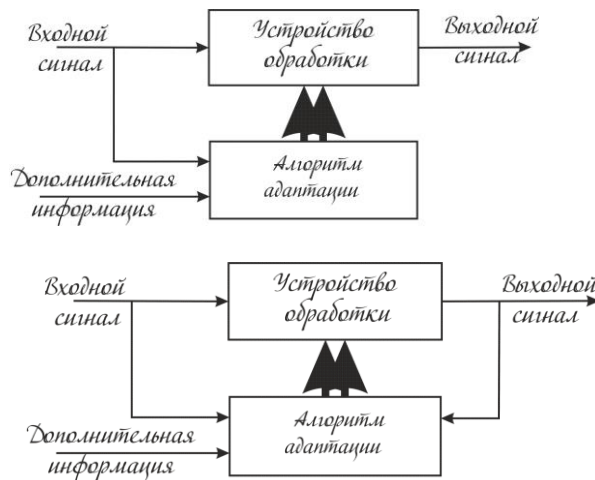
Все рассматриваемые до сих пор сигналы и цепи предполагались стационарными и имели неизменные характеристики и параметры. В реальной практике очень часто приходится иметь дело с процессами, меняющими свои характеристики во времени. При этом понятно, что устройства, обеспечивающие оптимальные результаты при одних значениях характеристик обрабатываемых сигналов, могут оказаться не столь эффективными при других характеристиках.

Для работы с процессами, изменяющими свои свойства во времени, используются адаптивные системы, способные к перенастройке своей структуры и параметров в зависимости от изменения условий функционирования.

Большинство адаптивных систем помимо параметрического изменения во времени являются также нелинейными с изменяющейся структурой.

Процесс адаптации принято разделять на два больших подкласса: адаптация с обратной связью и адаптация без обратной связи. Обобщенные структурные схемы обоих видов адаптивных систем приведены на рис.9.10.

Выбор варианта адаптивного алгоритма обработки сигналов определяется многими факторами. Достоинством алгоритмов адаптации с обратной связью является возможность их использования в условиях отсутствия или невозможности применения аналитических методов синтеза обрабатывающих устройств. Однако наличие петли обратной связи требует повышенного внимания к обеспечению устойчивости, как самой системы, так и используемых вычислительных алгоритмов. В задачах, не имеющих однозначного оптимума, такие решения могут приводить к возникновению колебательных или неопределенных процессов в системе.



Применительно к цифровым фильтрам схема адаптации может быть представлена в обобщенном виде, как показано на рис 9.16. В рамках этой схемы можно выделить два основных типа решаемых задач. В первом случае неизвестная система $W(z)$ подключается к адаптивному фильтру таким образом, чтобы она имела с ним общий входной сигнал, а задающим сигналом является выходной сигнал этой системы.

Рис.9.16. Адаптация без обратной связи и с обратной связью.

Такой способ соединения показан на рис.9.17. В этом случае идеальная импульсная характеристика фильтра является точной моделью системы $W(z)$. При другом способе соединения, показанном на том же рисунке неизвестная система $W(z)$ подключается к

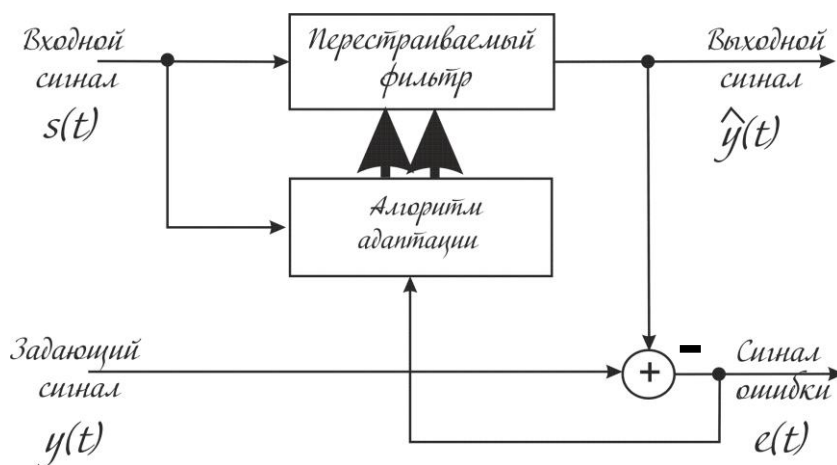


Рис.9.17. Обобщенная схема адаптивного фильтра

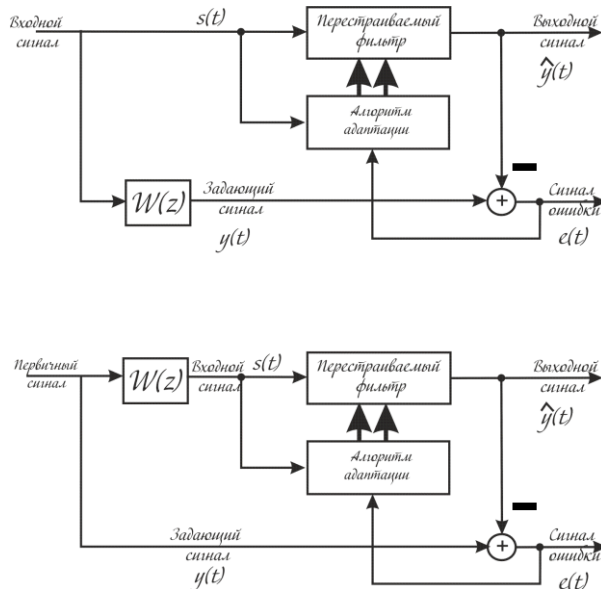


Рис.9.18. Схемы прямого и обратного моделирования

основному входу адаптивного фильтра, в результате чего формирующаяся импульсная характеристика адаптивного фильтра является функцией, обратной к импульсной характеристике неизвестной системы $W(z)$.

Практическим примером, соответствующим первому типу включения адаптивного фильтра является подавление отраженного сигнала в каналах связи. Второй тип включения, приведенный на рис.9.18. используется, например, при коррекции искажений канала связи.

В зависимости от соотношения скоростей изменения условий функционирования и свойств распространяющихся в системе сигналов можно выделить два принципиально различающихся случая.

Первый вариант соответствует относительно медленному характеру изменения условий функционирования по сравнению с характеристиками самого информационного сигнала. В этой ситуации подстройку параметров фильтра можно производить не на каждом новом наблюдении, а через некоторое количество наблюдений. Поэтому в задачах такого типа вполне уместно применение блочных методов оценки параметров сигналов. Данный подход во многом созвучен с идеей использования фильтра Винера с периодическим пересчетом его параметров. В промежутках между коррекциями параметров используемый фильтр имеет постоянные коэффициенты и является стационарным.

Если же скорости изменения условий функционирования и характеристик сигнала близки, то обновление подстройки желательно производить с каждым новым полученным отсчетом. В этой ситуации наиболее подходящими оказываются рекурсивные методы оценки параметров сигнала. Поскольку данный вариант является нестационарным, то, естественно, применение стационарных фильтров не может дать эффективного результата. Один из вариантов решения задачи может быть основан на использовании нестационарного фильтра Калмана.

9.7.2. Дискретный Фильтр Калмана

В разделе, посвященном вопросам оптимальной фильтрации непрерывных сигналов, мы кратко говорили о фильтре Калмана-Бьюси для непрерывного времени (8.23). Сейчас рассмотрим дискретный вариант этого нестационарного фильтра.

Для целей большей наглядности начнем рассмотрение с простейшего скалярного случая. Предположим, что интересующий нас случайный сигнал $x(k)$ может быть рассмотрен как результат прохождения через формирующий фильтр первого порядка белого шума $\xi(k)$, то есть

$$x(k) = a(k)x(k-1) + \xi(k), \quad (9.55)$$

где $a(k)$ – изменяющийся во времени коэффициент фильтра.

Наблюдаемые отсчеты сигнала $x(k)$ из (9.55) дополнительно искажены присутствием аддитивного шума $v(k)$, что отражает уравнение наблюдений вида

$$y(k) = cx(k) + v(k), \quad (9.56)$$

где c – некоторый постоянный коэффициент.

По аналогии с фильтром Калмана-Бьюси нетрудно предположить, что в варианте дискретного времени искомая оценка сигнала $x(k)$ должна иметь вид

$$\hat{x}(k) = b(k)\hat{x}(k-1) + K(k)y(k). \quad (9.57)$$

Для обеспечения оптимального (с точки зрения метода наименьших квадратов) результата соотношение между значениями $b(k)$ и $K(k)$ в (9.57) следует искать таким, чтобы минимизировать математическое ожидание квадрата ошибки

$$M\{e^2(k)\} = M\{(\hat{x}(k) - x(k))^2\} \rightarrow \min.$$

Введем обозначение

$$p(k) = M\{(\hat{x}(k) - x(k))^2\} = M\{(b(k)\hat{x}(k-1) + K(k)y(k) - x(k))^2\}.$$

Для нахождения оптимальных значений $a(k)$ и $K(k)$ продифференцируем выражение для $p(k)$ по этим переменным и приравняем результаты нулю.

$$\begin{aligned} \frac{\partial p(k)}{\partial b(k)} &= 2M\{[b(k)\hat{x}(k-1) + K(k)y(k) - x(k)]\hat{x}(k-1)\} = 0, \\ \frac{\partial p(k)}{\partial K(k)} &= 2M\{[b(k)\hat{x}(k-1) + K(k)y(k) - x(k)]y(k)\} = 0. \end{aligned}$$

Первое из полученных после дифференцирования уравнений позволяет записать

$$M\{[b(k)\hat{x}(k-1) - x(k-1)]\hat{x}(k-1)\} = M\{[x(k) - K(k)y(k)]\hat{x}(k-1)\}$$

Добавляя и вычитая слагаемое $x(k-1)$ в левую часть данного равенства, запишем

$$M\{[b(k)[\hat{x}(k-1) - x(k-1) + x(k-1)]\hat{x}(k-1)\} = M\{[x(k) - K(k)y(k)]\hat{x}(k-1)\}$$

или

$$\begin{aligned} M\{[b(k)[\hat{x}(k-1) - x(k-1)]\hat{x}(k-1) + b(k)[x(k-1)]\hat{x}(k-1)\} = \\ = M\{[x(k) - K(k)y(k)]\hat{x}(k-1)\} \end{aligned}$$

Принимая во внимание связь между $x(k)$ и $y(k)$ получаем

$$\begin{aligned} b(k)M\{[e(k)\hat{x}(k-1) + x(k-1)]\hat{x}(k-1)\} = \\ = M\{[x(k)[1 - cK(k)] + K(k)v(k)]\hat{x}(k-1)\}. \end{aligned}$$

Если проектируемый фильтр является оптимальным, то для него должен выполняться принцип ортогональности, в соответствии с которым

$$M\{e(k)\hat{x}(k-1)\} = 0, \quad \text{и} \quad M\{v(k)\hat{x}(k-1)\} = 0.$$

Эти соотношения позволяют упростить предыдущую формулу и привести ее к виду

$$b(k)M\{x(k-1)\hat{x}(k-1)\} = [1 - cK(k)]M\{x(k)\hat{x}(k-1)\},$$

что с учетом модели формирования сигнала $x(k)$ дает

$$\begin{aligned} b(k)M\{x(k-1)\hat{x}(k-1)\} = \\ = [1 - cK(k)]M\{b(k)x(k-1)\hat{x}(k-1) - \xi(k-1)\hat{x}(k-1)\} \end{aligned}$$

Из уравнения фильтрации следует, что

$$\hat{x}(k) = b(k)\hat{x}(k-1) + K(k)[cx(k) + v(k)]$$

или

$$\hat{x}(k) = b(k)\hat{x}(k-1) + K(k)b(k)cx(k-1) + K(k)\xi(k-1) + K(k)v(k)$$

откуда

$$\begin{aligned} \hat{x}(k-1) &= b(k-1)\hat{x}(k-2) + K(k-1)b(k-1)cx(k-2) + \\ &+ K(k-1)\xi(k-2) + K(k-1)v(k) \end{aligned}$$

Если данное уравнение умножить на $\xi(k-1)$, то очевидно

$$M\{\xi(k-1)\hat{x}(k-1)\} = 0.$$

Следовательно,

$$\begin{aligned} b(k)M\{x(k-1)\hat{x}(k-1)\} &= \\ &= [1 - cK(k)]M\{b(k)x(k-1)\hat{x}(k-1)\} \end{aligned}$$

и искомое соотношение между параметрами фильтра определяется как

$$b(k) = a[1 - cK(k)]$$

Окончательное уравнение для фильтра принимает вид

$$\hat{x}(k) = a\hat{x}(k-1) + K(k)[y(k) - ac\hat{x}(k-1)].$$

Полученное уравнение является определением оптимального рекурсивного устройства оценки первого порядка (или скалярный случай дискретного фильтра Калмана), которое оценивает текущее значение интересующей величины как взвешенную сумму предыдущей оценки и разницы между текущим значением сигнала и его прогноза на основе предыдущей оценки.

После определения структуры скалярного фильтра Калмана требуется найти выражения для изменяющегося во времени коэффициента усиления данного фильтра $K(k)$.

Квадрат ошибки фильтрации можно представить в виде

$$p(k) = M\{e^2(k)\} = M\{e(k)[b(k)\hat{x}(k-1) + K(k)y(k) - x(k)]\} = -M\{e(k)x(k)\}.$$

Поскольку

$$M\{e(k)x(k)\} = \frac{1}{c}M\{e(k)[y(k) - v(k)]\},$$

то

$$p(k) = \frac{1}{c}M\{e(k)v(k)\}.$$

Далее понятно, что

$$p(k) = \frac{1}{c}K(k)M\{y(k)v(k)\} = \frac{1}{c}K(k)\sigma_v^2,$$

где $\sigma_v^2 = M\{v(k)v(k)\}$.

Теперь, используя основное уравнение фильтра запишем

$$\begin{aligned} p(k) &= M\left\{[a\hat{x}(k-1) + K(k)(y(k) - ac\hat{x}(k-1)) - x(k)]^2\right\} = \\ &= M\left\{[a(1 - cK(k))e(k-1) - (1 - cK(k))\xi(k-1) + K(k)v(k)]^2\right\} = \\ &= a^2(1 - cK(k))^2 p(k) + (1 - cK(k))^2 \sigma_\xi^2 + K^2(k)\sigma_v^2. \end{aligned}$$

Откуда для коэффициента $K(k)$ можно получить

$$K(k) = c(a^2 p(k-1) + \sigma_\xi^2)[\sigma_v^2 + c^2 \sigma_\xi^2 + c^2 a^2 p(k-1)]^{-1} \quad (9.58)$$

$$\text{и } p(k) = \frac{\sigma_v^2}{c} K(k). \quad (9.59)$$

Два последних уравнения (9.58), (9.59) совместно с основным структурным уравнением фильтра

$$\hat{x}(k) = a\hat{x}(k-1) + K(k)[y(k) - ac\hat{x}(k-1)] \quad (9.60)$$

позволяют сформировать рекурсивный алгоритм соответствующий фильтру Калмана первого порядка.

Структурная схема формирования сигнала и скалярного фильтра (9.60) для его оценивания приведена на рис.9.19.

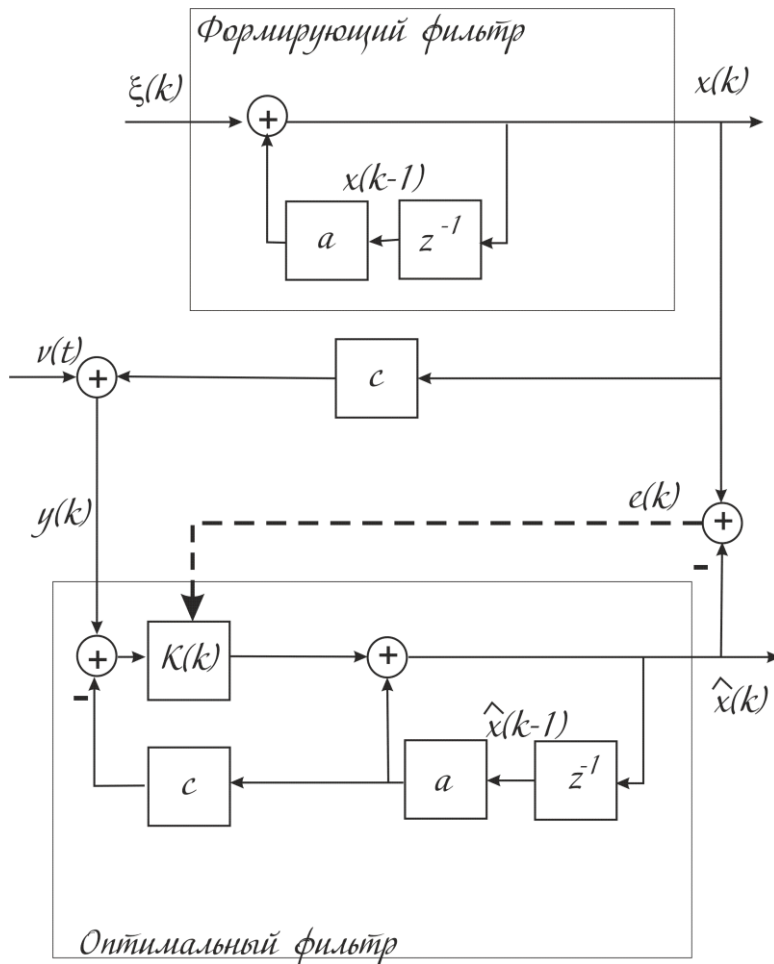


Рис.9.19. Скалярный вариант дискретного фильтра Калмана

На практике приходится иметь дело с более сложными сигналами, чем те, описываемые скалярными моделями первого порядка. Кроме того шумы, входящие в модель сигнала, тоже могут иметь меняющиеся во времени параметры.

В таких случаях используют векторный вариант модели формирования сигнала

$$\begin{aligned} X(k) &= AX(k-1) + \xi(k-1) \\ Y(k) &= CX(k) + v(k), \end{aligned} \quad (9.61)$$

где $M\{\xi(k)\xi'(k)\} = \Xi(k)$, $M\{v(k)v'(k)\} = Q(k)$,
и соответствующий ему векторный вариант фильтра Калмана

$$\begin{aligned} X(k) &= AX(k-1) + K(k)(Y(k) - CAX(k-1)) \\ K(k) &= P_1(k)C'(CP_1(k)C' + Q(k))^{-1} \\ P_1(k) &= AP(k-1)A' + \Xi(k-1) \\ P(k) &= P_1(k) - K(k)CP_1(k) \end{aligned} \quad (9.62)$$

Заметим, что эти уравнения (9.61), (9.62) чрезвычайно близки тем, которые мы получали, рассматривая рекурсивный вариант оценки параметров АР-модели. По своей сути это один и тот же алгоритм, основанный на обеспечении минимума квадрата ошибки, но примененный не к параметрам модели, а к сигналу. В случаях, когда требуется одновременно оценивать и параметры модели и значение сигнала может применяться расширенный фильтр Калмана. Более подробно данная тема изучается в курсе цифровой обработки сигналов.

9.8. Области применения адаптивных фильтров

В традиционных методах обработки сигналов для их обработки используются линейные системы с постоянными параметрами. Эти системы могут иметь как конечную, так и бесконечную импульсную характеристику, но их передаточная функция не изменяется во времени.

Адаптивные устройства обработки сигналов характеризуются наличием определенной связи между параметрами передаточной функции и характеристиками обрабатываемых сигналов и другой дополнительной информации, обеспечивающей системе самонастройку на режим оптимальной обработки сигналов.

Одно из основных направлений адаптивной фильтрации связано с подавлением нестационарных помех.

Традиционный метод оценки искаженного влиянием помех сигнала состоит в стремлении подавить искажающую помеху, оставляя, по возможности, не тронутым полезный сигнал. Однако, если характеристики сигнала или помехи не являются неизменными, то применение фильтров с постоянными параметрами, оптимизированными по априорным данным, окажется не эффективным.

Рассмотрим основную идею фильтрации помех с помощью адаптивных фильтров.

Обобщенная схема адаптивной фильтрации помех приведена на рис.9.20.

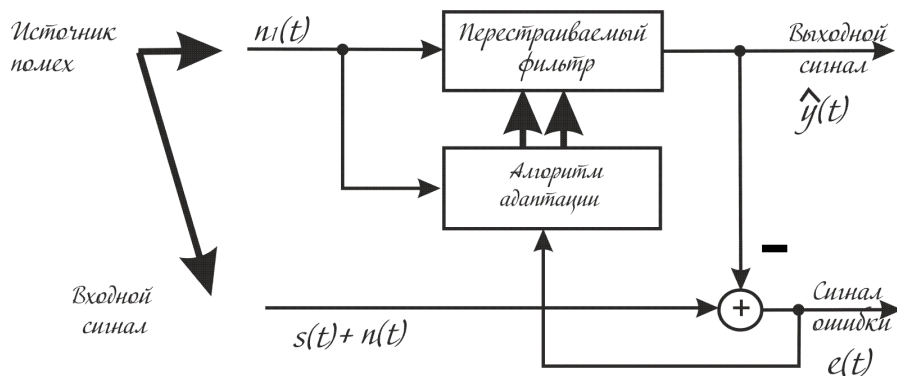


Рис.9.20. Адаптивное подавление помех

В соответствии с приведенной схемой сигнал передается по некоторому каналу связи на приемное устройство, которое принимает смесь полезного сигнала и некоррелированной с ним помехи $s(t) + n(t)$.

Другое приемное устройство принимает

помеху $n_1(t)$, которая не коррелирована с полезным сигналом $s(t)$, но некоторым неизвестным образом коррелирована с помехой $n(t)$.

В данном приемном устройстве формируется сигнал $\hat{y}(t)$, который вычитается из принимаемого первым приемником сигнала. В результате формируется сигнал ошибки в виде

$$e(t) = s(t) + n(t) - \hat{y}(t).$$

При известных и неизменных характеристиках каналов, по которым помеха поступает на оба приемника, можно синтезировать оптимальный фильтр с постоянными параметрами, преобразующий $n_1(t)$ в $\hat{y}(t)$, таким образом, что $\hat{y}(t) \rightarrow n(t)$. Тогда вычитая выходной сигнал фильтра из входного сигнала можно получить очищенный от помехи полезный сигнал.

Однако в большинстве случаев характеристика каналов связи неизвестны или известны приблизительно, а кроме того они могут меняться во времени. Поэтому применение оптимальных фильтров с постоянными параметрами может оказаться малоэффективным.

В соответствии со схемой представленного фильтра обработка сигнала проводится перенастраиваемым фильтром, который изменяет свои параметры по какому-либо среднеквадратическому алгоритму, функционирующему по сигналу ошибки $e(t)$. Таким

образом, при выборе правильного алгоритма и обеспечении условий его сходимости представленный адаптивный фильтр может работать в широком диапазоне изменения характеристик каналов связи.

Второе направление использования адаптивных фильтров базируется на так называемом методе обратного моделирования. Обобщенная схема обратного моделирования с помощью адаптивных фильтров представлена на рис.9.21.

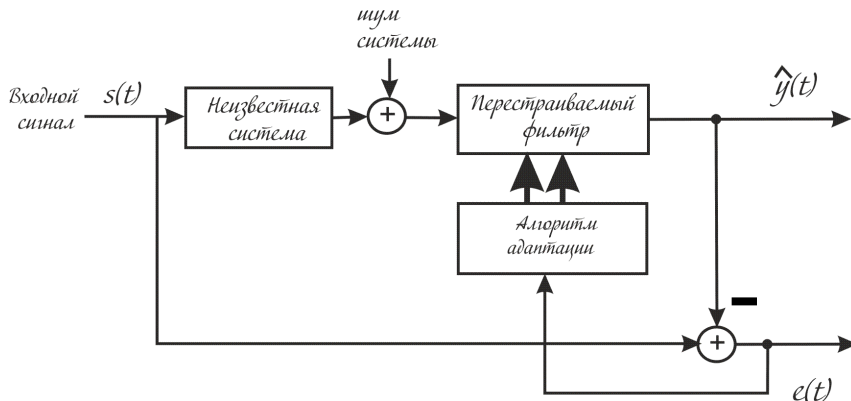


Рис.9.21. Схема обратного моделирования

Оценка передаточной функции неизвестной системы в данном случае проводится с использованием известного входного сигнала $s(t)$. Выходной сигнал оцениваемой системы зашумлен некоторым шумом $n(t)$. Алгоритм настройки перестраиваемого фильтра работает таким

образом, чтобы формируемый им сигнал $\hat{y}(t)$ был максимально близок к входному сигналу системы. Таким образом, сигнал ошибки, подлежащей минимизации определяется как

$$e(t) = s(t) + n(t) - \hat{y}(t).$$

Параметры фильтра выбираются исходя из критерия минимума среднего квадрата ошибки, и для этих целей может быть использован любой из рассмотренных ранее среднеквадратичных алгоритмов.

Подобная схема широко применяется, например, при настройке телефонных линий связи. Перед началом работы на вход линии подается последовательная серия известных тестовых сигналов, в соответствии с которыми перестраивается адаптивный фильтр, имитируя обратную передаточную функцию канала связи. Последовательное соединение канала связи и перенастроенного на обратную характеристику канала фильтра позволяет получить общую передаточную функцию близкую к единице и таким образом откорректировать искажения канала.

На практике подобные схемы реализуются как в виде, представленном на рис.9.15, так и с использованием дополнительной временной задержки для формирования опорного сигнала алгоритма, используемого алгоритмом адаптации фильтра.

Рассмотренные области применения адаптивной фильтрации относились к временной области. Два других приложения относятся к частотной и пространственной областям применения. Частотная область применения по своей сути очень близка к рассмотренному материалу, поэтому специально останавливаться на ней мы не будем. Область пространственного применения адаптивных фильтров обладает рядом специфических особенностей, поэтому данный вопрос рассмотрим несколько подробнее. В наиболее общем виде данная область применения может быть обозначена как создание адаптивных антенных решеток.

Для достижения пространственной избирательности необходимо осуществлять прием сигналов на антенную решетку, имеющую два и более независимых и разнесенных в пространстве элемента. В зависимости от специфики конкретной задачи элементами антенны могут быть и диполи для приема электромагнитных колебаний, и акустические приемники, и гидрофоны, и датчики сейсмической активности и другие датчик, предназначенные для приема распространяющихся в пространстве сигналов различной физической природы.

Адаптивная антенна (или адаптивная система формирования лучей) состоит из множества разнесенных в пространстве элементов. Эти элементы далее соединяются с одноканальными или многоканальными устройствами обработки сигналов.

Области применения адаптивных антенных решеток довольно разнообразны и широки. К числу основных областей применения можно отнести следующие

- подавление направленных помех за счет адаптивного формирования нулей скомпенсированной диаграммы направленности,
 - самонастройка по направлению сигнала,
 - определение направления прихода сигнала,
 - выделение слабых сигналов на фоне сильных, при их приходе под разными углами,
 - обеспечение избирательной чувствительности по дальности источника сигналов и форме принимаемого сигнала
- и ряд других.

В задачи нашего курса не входит подробное изучение антенных систем, поэтому ограничимся кратким демонстрационным примером простейшего вида адаптивной антенны с подавлением боковых лепестков.

Схема такого устройства представлена на рис.9.22.

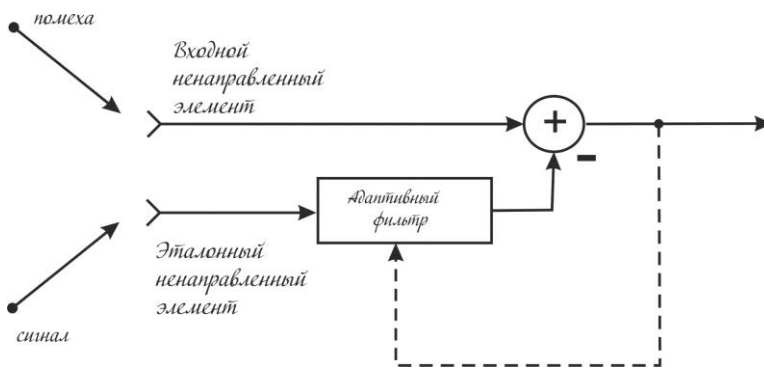


Рис.9.22. Адаптивное подавление боковых лепестков в двухэлементной антенне

Данная схема включает в себя два ненаправленных (то есть с одинаковой чувствительностью по всем направлениям приема сигнала) элемента. Один из элементов условно назовем входным, а второй эталонным. Допустим, что в рассматриваемый момент времени на антенну воздействует один полезный сигнал и одна помеха. Оба элемента

антенны принимают как полезный сигнал, так и помеху. Поскольку элементы антенны разнесены в пространстве, то в силу конечности скорости распространения сигнала и помехи, входные сигналы на обоих элементах будут отличаться друг от друга. С другой стороны, так как приходящие сигналы имеют общие источники возникновения, то они должны быть связаны некоторой временной зависимостью.

Практически важным случаем является ситуация при которой помеха по мощности значительно превышает полезный сигнал. Тогда из схемы включения адаптивного фильтра видно, что его коэффициента будут почти полностью определяться свойствами помехи. Поэтому (по аналогии с рассмотренным ранее адаптивным подавлением помех во временной области) на выходе фильтра по окончании процесса адаптации формируется сигнал близкий к сигналу помехи, приходящей на входной ненаправленный элемент антенны. Поэтому выходной сигнал всей системы практически не будет содержать составляющей помехи. При этом выходной сигнал системы примерно равен разности полезного сигнала на входном элементе и преобразованного фильтром полезного сигнала, полученного с эталонного входа.

Поскольку без априорной информации невозможно предсказать, какими будут характеристики каналов распространения сигнала и помехи до каждого антенного элемента, то невозможно заранее определить и конечный вид импульсной характеристики фильтра после окончания адаптации. Следовательно, в общем случае получаемый

результат оказывается неопределенным. Но при значительной мощности помехи такая схема дает увеличение отношения сигнал-шум.

На практике сигнал с каждого ненаправленного элемента антенны передается для усиления, фильтрации и выделения сигнала на отдельное приемное устройство. При этом и сами элементы антенн, и приемные устройства вносят в обрабатываемые сигналы дополнительные независимые шумы, называемые шумами приемников. Как отмечалось выше, чтобы рассматриваемое адаптивное устройство подавляло направленную помеху, но не подавляло полезный сигнал, мощность помехи должны превышать и мощность полезного сигнала, и мощность шумов приемника.

Пример 9.6.

Рассмотрим пример применения адаптивного фильтра для обратного моделирования неизвестной системы в условия присутствия шумов наблюдения. Предполагается, что во время наблюдения в системе произошел скачок параметров. (Скачок параметров происходит в момент отсчета номер 2500). Моделирование проведем с помощью программы MATLAB. Листинг рабочей сессии приведен на рис.9.23.

```
a1=0.8;           %Задание параметров исследуемой системы до скачка
a2=0.1;
a12=-0.1;        %Задание параметров исследуемой системы после скачка
a22=0.1;
N=5000;          % Задание длины выборки наблюдений
u=randn(N,1);    % Формирование возбуждающего сигнала
n=0.01*randn(N,1); %Формирование шума наблюдений
d(1:N/2)=filter(1,[1 a1 a2],u(1:N/2))+n(1:N/2); % Формирование выхода системы с шумом до скачка
d(N/2+1:N)=filter(1,[1 a12 a22],u(N/2+1:N))+n(N/2+1:N); % Формирование выхода системы с шумом после скачка
p=7;             % Задание порядка адаптивного КИХ фильтра для обратного моделирования системы
mu=0.008;        % Шаг алгоритма LMS
ha1=adaptfilt.lms(p,mu); % Задание адаптивного фильтра
[y,e]=filter(ha1,d,u); % Адаптивная фильтрация
plot(e);          % Построение графика ошибки обратного фильтра.
```

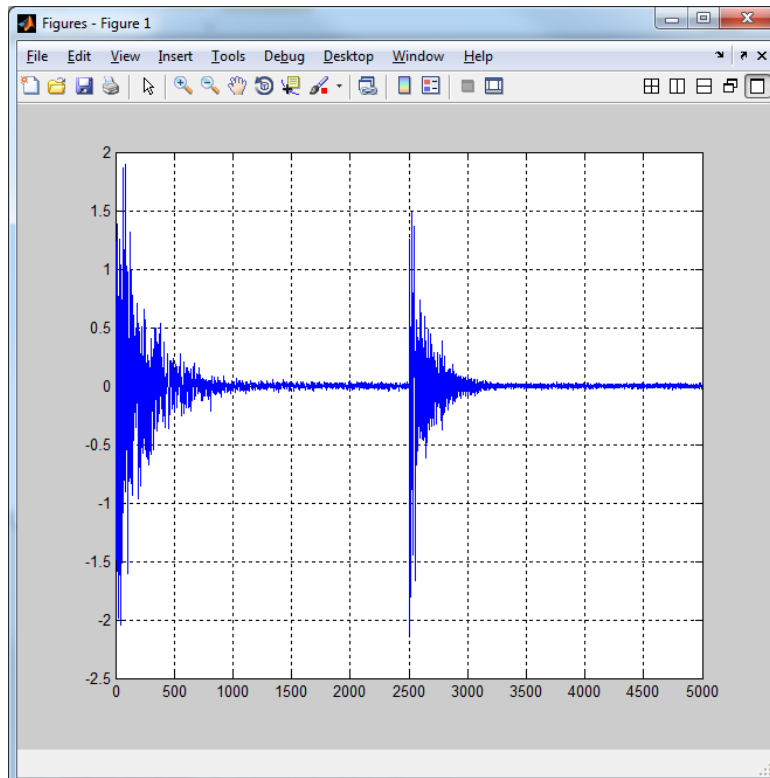


Рис.9.23. Использование адаптивного фильтра для аппроксимации неизвестной системы

Пример 9.7.

На рис.9.24. представлены результаты моделирования в программе Matlab адаптивного подавления помехи в соответствии со схемой рис.9.20. В вашем распоряжении имеются дискретные записи полезного сигнала $s(t)$, аддитивно искаженного неким окрашенным шумом $n(t)$. (сигнал $s(t)+n(t)$: Файл SignalPlusNoise). Кроме того имеется дискретная запись случайного сигнала $n1(t)$, коррелированного с шумом $n(t)$, искажившим полезный сигнал.(сигнал $n1(t)$: Файл Noise2).

```
subplot(3,1,1);
plot(SignalPlusNoise); %Визуализируем запись полезного сигнала, искаженного шумом
title('Сигнал,искаженный шумом');
xlabel('Отсчеты времени'); ylabel('Величина сигнала');
soundsc(SignalPlusNoise,44100);% Проводим звуковое воспроизведение искаженного сигнала
pause;
subplot(3,1,2);
plot(Noise2)% Визуализируем запись сигнала, коррелированного с искажающим шумом
title('Сигнал, коррелированный с шумом');
xlabel('Отсчеты времени'); ylabel('Величина сигнала');
pause;
Nfiltr=32; % Задаем порядок адаптивного RLS фильтра
P0 = 10*eye(Nfiltr); % Задаем начальное значение обратной корреляционной матрицы
lam = 0.99; % Задаем RLS коэффициент забывания
FIL1 = adaptfilt.rls(Nfiltr,lam,P0); % Создаем адаптивный RLS-фильтр с заданными параметрами
[y,e]=filter(FIL1,Noise2,SignalPlusNoise); % Применяем встроенную функцию адаптивной фильтрации
subplot(3,1,3); plot(e);
title('Сигнал после адаптивной фильтрации RLS-фильтром');
xlabel('Отсчеты времени'); ylabel('Величина сигнала');
soundsc(e,44100);% Проводим воспроизведение отфильтрованного сигнала
```

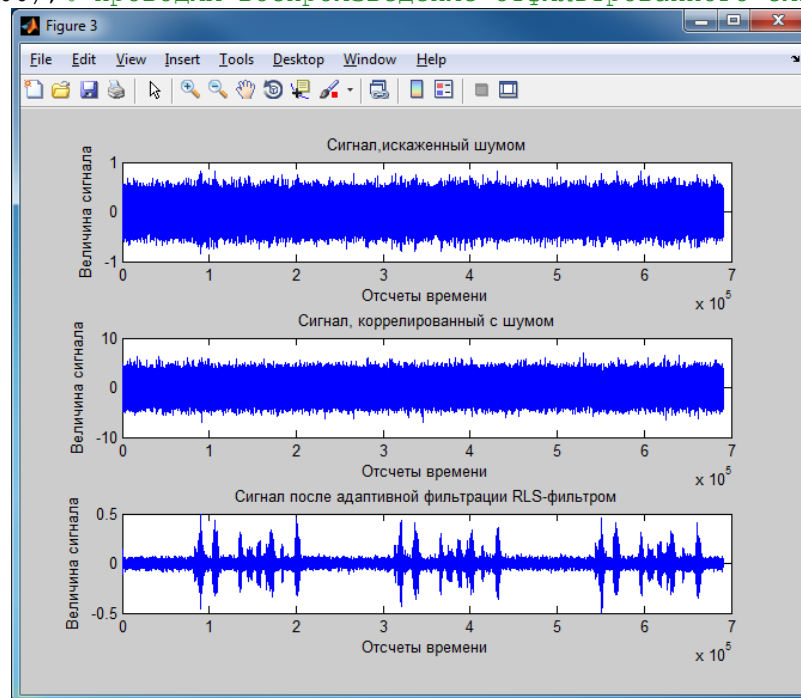


Рис.9.24. Демонстрация применения адаптивного RLS фильтра к речевому сигналу

9.9. Общие сведения о цифровом спектральном оценивании

В общем случае спектральная плотность мощности дискретного сигнала $x(k)$ может быть определена по формуле

$$S_{Px}(\omega) = T_0 \sum_{m=-\infty}^{\infty} R_x(m) e^{-j\omega m T_0}, \quad (9.63)$$

где T_0 – интервал временного квантования непрерывного сигнала.

Если требуется оценить спектральную плотность мощности (9.63) сигнала $x(k)$ по конечной выборке наблюдений $x(0), x(1), \dots, x(N-1)$, то классический подход к решению вопроса базируется на двух основных методах: косвенный подход, связанный с вычислением автокорреляционной функции, и прямой метод, базирующийся на прямом вычислении спектра с помощью алгоритма БПФ.

При наличии конечного количества отсчетов сигнала можно определить только конечное количество значений автокорреляционной функции, остальная часть которых будет просто недоступна. Спектральная оценка, получаемая таким методом, была предложена Блэкманом и Тьюки и соответствует формуле

$$\widehat{S}_{Px}(\omega) = T_0 \sum_{m=-M}^M \widehat{R}_x(m) e^{-j\omega m T_0}, \quad (9.64)$$

или

$$\widehat{S}_{Px}(f) = T_0 \sum_{m=-M}^M \widehat{R}_x(m) e^{-j2\pi f m T_0},$$

где $-\frac{1}{2T_0} \leq f \leq \frac{1}{2T_0}$.

Выражения (9.64) по своей сути являются дискретным аналогом уравнения Винера-Хинчина с ограниченным количеством слагаемых.

Вычисление оценки автокорреляционной функции в данном случае производится по формуле

$$\widehat{R}_x(m) = \frac{1}{N-m} \sum_{n=0}^{N-m-1} x(n+m) x^*(n),$$

где $m = 0, 1, \dots, M$, $M \leq N-1$.

Для отрицательных значений аргумента оценка автокорреляционной функции получается из соотношения

$$\widehat{R}_x(-m) = \widehat{R}_x^*(m).$$

В случае вещественного сигнала операцию комплексной сопряженности можно опустить

$$\widehat{R}_x(m) = \frac{1}{N-m} \sum_{n=0}^{N-m-1} x(n+m) x(n).$$

Дженкинс и Ватс для целей спектрального оценивания предложили использовать смещенную оценку автокорреляционной функции в виде

$$\widehat{R}_x'(m) = \frac{1}{N} \sum_{n=0}^{N-m-1} x(n+m)x(n),$$

которая для многих конечных наборов данных дает меньшую среднеквадратическую ошибку, чем использование $\widehat{R}_x(m)$.

Оценка $\widehat{R}_x'(m)$ является смещенной, так как

$$M\{\widehat{R}_x'(m)\} = \frac{N-m}{N} R_x(m).$$

Среднее значение этой оценки равно истинной автокорреляционной функции, взвешенной треугольным окном (иногда его называют окном Бартлета).

Прямой метод спектрального оценивания представляет собой современный вариант периодограммы Шустера и проводится по выражению

$$\widehat{S}_{Px}(f) = \frac{1}{NT_0} \left| T_0 \sum_{k=0}^{N-1} x(k) e^{-j2\pi f k T_0} \right|^2, \quad (9.65)$$

где $-\frac{1}{2T_0} \leq f \leq \frac{1}{2T_0}$.

Операция математического ожидания в уравнении (9.65) опущена, поэтому оно оценивает мгновенный спектр мощности, который строго говоря, зависит от времени.

Использование БПФ

$$\widehat{S}_{Px}(f) = \frac{1}{NT_0} |X_n|^2,$$

где X_n – БПФ от $x(k)$,

позволяет вычислить данную оценку на дискретном множестве эквидистантных частот

$$f_n = n \frac{1}{NT_0},$$

где $n = 0, 1, 2, \dots, N-1$.

Во многих практических задачах исследуемые сигналы содержат детерминированные составляющие (чаще всего периодические), поэтому удобно, когда пик графика СПМ соответствует мощности предполагаемого сигнала. Для этой цели уравнение оценки модернизирую в виде

$$\widehat{S}_{Px}'(f) = \widehat{S}_{Px}(f) \frac{1}{NT_0} = \frac{1}{N} \left| \sum_{k=0}^{N-1} x(k) e^{-\frac{j2\pi f k n}{N}} \right|^2. \quad (9.66)$$

Непосредственное применение оценки (9.66) по методу периодограммы (то есть без учета операции математического ожидания) может приводить к несостоятельным оценкам. Поэтому к настоящему времени предложено несколько его модификаций для усреднения по ансамблю.

Использование конечной выборки из процесса $x(k)$ требует введения каких-либо предположений о его значениях за пределами интервала наблюдений. Если эти значения ненулевые то последовательность анализируемых данных можно рассматривать как полученную из бесконечной с помощью прямоугольного окна. Применение окна приводит к тому, что получаемая оценка спектральной плотности мощности содержит информацию не только о самом сигнале, но и о примененном окне, поскольку итоговое преобразование будет представлять собой свертку требуемого преобразования с преобразованием функции примененного окна.

Если истинный спектр сигнала сосредоточен в относительно узкой частотной полосе, то операция его свертки с функцией окна приводит к "просачиванию" (или

"утечке" мощности в соседние частотные области. Для минимизации данного эффекта предложено большое количество разнообразных форм окна.

Интерполяцию значений спектра для придания ему более гладкой формы можно проводить с применением процедуры добавления нулей, которая была подробно рассмотрена нами ранее.

Классический метод спектрального оценивания на основе алгоритма Блэкмана-Тьюки имеет ряд достоинств к числу которых можно отнести вычислительную эффективность, удобство для практических приложений, а также то, что оценка СПМ прямо пропорциональна мощности синусоидальных процессов. При этом имеются и явные недостатки, наиболее значимым из которых являются подавление главных лепестков слабых сигналов боковыми лепестками сильных сигналов. Кроме того частотное разрешение такой оценки ограничено длиной имеющейся выборки наблюдений и не зависит от характеристик сигнала.

Во многих практических случаях информация об исследуемом процессе дает возможность сделать более разумные допущения о характере его изменений за пределами окна, чем просто предположить, что вне интервала наблюдения значения процесса являются нулевыми.

Использование такой априорной информации позволяет более точно выбрать модель аппроксимирующую реальный процесс, а в этом случае на основе более точной модели часто удается получить лучшие оценки свойств процесса, включая оценку спектра.

При таком подходе процедура спектрального оценивания включает в себя три основных этапа: выбор типа модели, оценка параметров принятой модели, расчет спектра на основе полученной параметрической модели процесса.

При рассмотрении вопросов прохождения случайного сигнала через линейные цепи мы выяснили, что спектральные плотности мощности входного и выходного сигналов связаны между собой уравнением

$$S_{Px}(\omega) = W(j\omega)W(-j\omega)S_{Pu}(\omega) = |W(j\omega)|^2 S_{Pu}(\omega).$$

Аналогичное выражение для дискретных систем имеет вид

$$S_{Px}(z) = W_z(z)W_z^*(z^{-1})S_{Pu}(z).$$

Полагая входной процесс дискретным белым шумом с нулевым средним и дисперсией σ^2 , получим

$$S_{Px}(z) = W_z(z)W_z^*(z^{-1})\sigma^2 T_0,$$

$$\text{где } W_z(z) = \frac{\sum_{n=0}^M b_n z^{-n}}{1 - \sum_{n=1}^L a_n z^{-n}}.$$

Таким образом, определение параметров передаточной функции формирующего фильтра и дисперсии возбуждающего белого шума позволяют провести оценку спектральных характеристик анализируемого процесса.

Для указанных целей используются все три основных представления передаточной функции формирующего фильтра АР, СС и АРСС.

Мы остановимся только на рассмотрении применения моделей авторегрессии для целей спектрального оценивания сигналов.

В данном случае для исследуемого сигнала $x(k)$ ищется аппроксимирующая модель вида

$$x(k) = a_1 x(k-1) + a_2 x(k-2) + \dots + a_p x(k-p) + e(k),$$

где p – порядок АР-модели.

Оценка спектральной плотности мощности далее производится по уравнению

$$\widehat{S}_{Px}(z) = \frac{\widehat{\sigma^2} T_0}{\widehat{A}(z)\widehat{A}^*(z^{-1})}$$

или

$$\widehat{S}_{Px}(f) = \frac{\widehat{\sigma}^2 T_0}{\left| 1 - \sum_{k=1}^p \widehat{a}_k e^{-j2\pi f k T_0} \right|}, \quad (9.67)$$

где $\widehat{\sigma}^2, \widehat{a}_k, k = 1, 2, \dots, p$ - оценки параметров модели авторегрессии.

Рассмотрим применение АР-модели на примере гармонического сигнала. Ранее для процесса

$$x(t) = A_0 \sin(\omega_0 t + \theta_0)$$

при условии его дискретизации с интервалом времени T_0 мы получали разностное уравнение вида

$$x(k) = 2 \cos \omega_0 T_0 x(k-1) - x(k-2)$$

или

$$x(k) = a_1 x(k-1) - x(k-2),$$

где $a_1 = 2 \cos \omega_0 T_0 = \text{const.}$

Таким образом, для сигнала, представляющего собой единственную синусоиду, разностное уравнение принимает вид авторегрессии второго порядка с нулевым входным сигналом. Если исследуемый сигнал состоит из m синусоид, то он будет описываться разностным уравнением $2m$ порядка.

Теперь предположим, что для наблюдения доступен не сам сигнал $x(k)$, а его аддитивная смесь с шумовым воздействием, то есть

$$y(k) = x(k) + n(k),$$

где $n(k)$ – шумовой сигнал.

Для простоты будем считать шум белым и запишем следующее

$$x(k) = a_1 x(k-1) - x(k-2) + n(k) - a_1 n(k-1) + n(k-2).$$

Следовательно, в случае присутствия шумов наблюдения итоговый процесс становится процессом АРСС с особым видом симметрии коэффициентов.

Поэтому, строго говоря, для его описания следует использовать не АР, а АРСС модель. При этом порядок модели должен выбираться исходя из удвоенного количества присутствующих в сигнале синусоид.

На практике часто используют другой путь – вместо АРСС модели применяется АР модель более высокого порядка. Простота ее использования и снижение погрешности аппроксимации за счет увеличения порядка модели во многих случаях позволяют получить хорошие результаты оценивания спектра на основании (9.67).

Пример 9.8.

Рассмотрим процедуру проведения спектрального оценивания сигнала состоящего из смеси двух синусоид и белого шума с использованием АР-моделей разного порядка. (Для оценки параметров модели воспользуемся методом Берга). Моделирование проведем с помощью программы MATLAB. Листинг рабочей сессии приведен на рис.9.24.

```

NN=1024; % Количество используемых отсчетов сигнала
k=1:1:NN; % номера отсчетов сигнала
A1=1; % Амплитуда первой синусоиды
A2=1; % Амплитуда второй синусоиды
F1=0.05; % относительная частота первой синусоиды
F2=0.1; % относительная частота второй синусоиды
X=A1*sin(2*pi*F1*k)+A2*sin(2*pi*F2*k) %Формирование полезного сигнала
subplot(4,1,1);plot(X); % Вывод графика полезного сигнала
SN=randn(NN,1); % Формирование шумового сигнала
Y=X+SN; % Формирование наблюдаемого сигнала
subplot(4,1,2);plot(k,Y); % Вывод графика наблюдаемого сигнала
    
```

```
[a,e]=arburg(Y,5)      % расчет параметров АР-модели 5-го порядка методом Бурга
[a1,e]=arburg(Y,15)    % расчет параметров АР-модели 15-го порядка методом Бурга
subplot(3,2,5); pburg(Y,5,1024,1); % Построение графика оценки СПМ для АР-модели 5-го порядка
subplot(3,2,6); pburg(Y,15,1024,1); % Построение графика оценки СПМ для АР-модели 15-го порядка
a = 1.0000 -0.3525 -0.1524 -0.0174 0.1133 0.1746
```

e = 1.4090

a1 =

Columns 1 through 10

1.0000 -0.2162 -0.0929 -0.0106 0.0945 0.1680 0.1397 0.0785 0.0878 0.0017

Columns 11 through 16

0.0129 0.0721 0.0873 0.1406 0.1627 0.0911

e = 1.2141

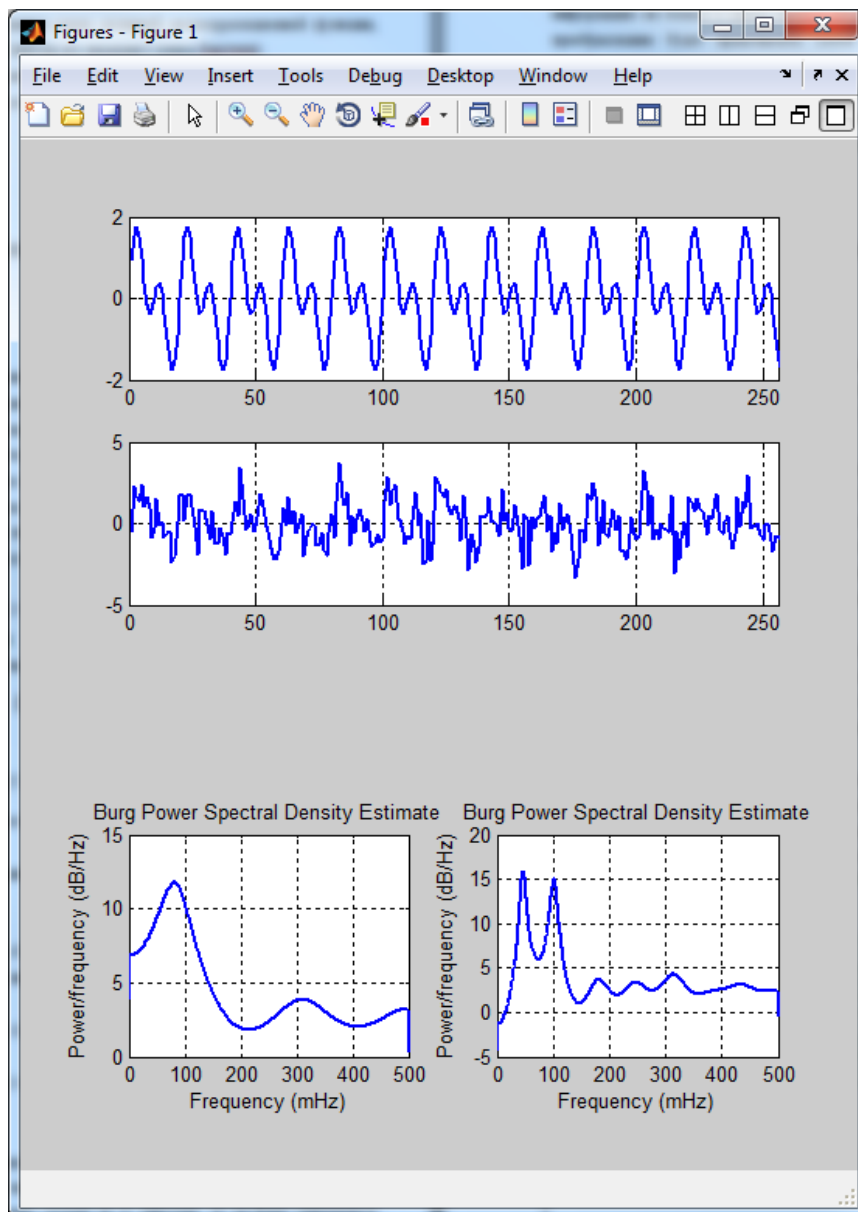


Рис.9.24. Спектральное оценивание сигнала на основе параметрических АР-моделей

Полученный результат наглядно показывает, что для решения данной задачи выбор порядка АР-модели равным 5 оказался недостаточным, так как спектральные

составляющие двух синусоид полезного сигнала слились в один пик. Повышение порядка модели позволило обнаружить обе спектральные составляющие.

9.10. Вопросы для самоконтроля и повторения

1. Поясните разницу между процессами дискретизации и квантования аналогового сигнала.
2. Поясните разницу между дискретными и цифровыми сигналами.
3. Что такое разностное уравнение цифрового фильтра.
4. Приведите структурную схему БИХ фильтра.
5. Приведите схему КИХ фильтра.
6. Дайте определение дискретной импульсной характеристики.
7. Что такое аналоговый фильтр-прототип для цифрового фильтра.
8. Что представляет собой Z-преобразование.
9. Перечислите основные свойства Z-преобразования.
10. Как определяется передаточная функция цифрового фильтра.
11. Назовите основные формы представления передаточных функций цифровых фильтров.
12. Назовите основные методы проектирования дискретных фильтров.
13. Поясните сущность метода билинейного Z-преобразования.
14. Поясните сущность метода инвариантной импульсной характеристики.
15. В чем состоит метод преобразования АЧХ фильтров нижних частот.
16. Поясните сущность метода максимального правдоподобия.
17. Поясните сущность метода наименьших квадратов.
18. Поясните разницу между блочным и рекурсивным методами наименьших квадратов.
19. Перечислите известные вам модели, используемых для оценки параметров случайных процессов.
20. Из каких этапов состоит процедура оценивания модели случайного процесса?
21. Поясните сущность адаптивной обработки сигналов.
22. Что представляет собой адаптивный фильтр?
23. Приведите две основные схемы использования адаптивного фильтра в задачах прямого и обратного моделирования.
24. Приведите обобщенную схему дискретного фильтра Калмана.
25. Каковы области применения адаптивных фильтров?
26. Перечислите известные вам методы спектрального оценивания сигналов.
27. Поясните разницу между параметрическими и непараметрическими методами спектрального оценивания.

Рекомендуемая литература

1. Баскаков С.И. Радиотехнические цепи и сигналы: Учебное пособие для вузов по спец. "Радиотехника"- "Радиотехника".-М: Высш. шк., 2000 – 462 с.
2. Васильев В.П., Муро Э.Л., Смольский С.М. Основы теории и расчетов цифровых фильтров.-Мю: Академия,2007,- 272 с.
3. Гоноровский И.С. Радиотехнические цепи и сигналы.: Учебное пособие для вузов – М.Дрофа, 2006, - 608 с.
4. Карташкин А.С. Авиационные радиосистемы.- 2-е изд. М. ИП Радиософт,2012, 304 с.
5. Основы цифровой обработки сигналов : учеб. пособие для вузов / А.И. Солонина, Д.А. Улахович, С.М. Арбузов, Е.Б. Соловьева . - 2-е изд., [испр. и перераб.]. - СПб. : БХВ-Петербург, 2005. - 753 с.
6. Рудой В.М. Системы передачи информации.-М.: МГОУ, 2004, - 172 с.
7. Сергиенко А.Б. Цифровая обработка сигналов: Учебное пособие для вузов- 3-е изд. СПб.: БХВ-Петербург, 2011.-751 с.
8. Теоретические основы радиотехники :Учеб.пособие/ М.Т.Иванов, А.Б. Сергиенко, В.Н. Ушаков; под.ред. В.Н. Ушакова.- М.: Высш.шк.,2002.- 306 с.
9. Радиотехнические цепи и сигналы: Учебник для вузов. Стандарт третьего поколения/ Под ред В.Н. Ушакова- СПб.:Питер, 2014.-336 с.