В.И. Нефедов, А.С. Сигов

# Основы радиоэлектроники и связи

*Под редакцией профессора В.И. Нефедова*

*Допущено*
*Министерством образования и науки*
*Российской Федерации*
*в качестве учебного пособия*
*для студентов высших учебных заведений,*
*обучающихся по специальности «Проектирование*
*и технология радиоэлектронных средств»*
*направления подготовки «Проектирование*
*и технология электронных средств»*




Москва «Высшая школа» 2009

УДК 621.37
ББК 32.842
Н 58

Рецензенты:

д-р техн. наук, проф. *А.В. Прохоров*
(Московский государственный технический университет гражданской авиации);
кафедра радиотехнических устройств и систем
Московского государственного открытого университета
(зав. кафедрой д-р техн. наук, проф. *В.А. Каплун*)

**Нефедов В.И.**

Н 58 Основы радиоэлектроники и связи: Учеб. пособие/В.И. Нефедов, А.С. Сигов; Под ред. В.И. Нефедова. — М.: Высш. шк., 2009. — 735 с.: ил.

ISBN 978-5-06-006161-1

Рассмотрены вопросы передачи информации при помощи электромагнитных колебаний и принципы построения основных радиотехнических устройств и систем. Изложены способы математического представления сигналов и помех, современные методы формирования, преобразования и обработки сигналов в системах связи. Проанализированы различные классы радиотехнических цепей и процессы, протекающие в них. Широко представлены основные электронные устройства и узлы, выполненные на интегральных аналоговых и цифровых микросхемах: усилители, модуляторы, детекторы, генераторы, счетчики, делители частоты и т.д. Приведены общие сведения о пропускной способности каналов связи, методах помехоустойчивого кодирования, оптимального приема сообщений. Описаны элементы современной теории вейвлетной и фрактальной обработки информации. Изложение теоретического материала проиллюстрировано практическими расчетами и примерами построения отдельных узлов.

*Для студентов вузов, обучающихся по специальности «Проектирование и технология радиоэлектронных средств» (направление подготовки дипломированных специалистов «Проектирование и технология электронных средств»), аспирантов, а также для инженерно-технических работников различных специальностей, нуждающихся в приобретении знаний в области радиоэлектроники и техники связи. Учебное пособие будет полезно студентам вузов, обучающимся по направлению подготовки «Радиотехника», и студентам других технических специальностей.*

УДК 621.37
ББК 32.842

ISBN 978-5-06-006161-1

© ОАО «Издательство «Высшая школа», 2009

Оригинал-макет данного издания является собственностью издательства «Высшая школа», и его репродуцирование (воспроизведение) любым способом без согласия издательства запрещается.

# ПРЕДИСЛОВИЕ

Новейший этап научно-технического прогресса связан с революционными изменениями в передаче, приеме и обработке информации, что оказывает существенное влияние на все стороны жизни общества. Все это стало возможным благодаря успехам в развитии радиоэлектроники и теории связи. В последние годы радиоэлектроника во многом определяет развитие большинства областей науки и техники. Теория связи тоже относится к числу фундаментальных дисциплин, и ее достижения позволяют повысить как скорость и объем передаваемой информации, так и качество ее передачи. Две эти дисциплины имеют много общего и поэтому объединены в общий курс «Основы радиоэлектроники и связи».

Основной целью курса «Основы радиоэлектроники и связи» является обучение студентов теории, методам и основам схемотехнического построения устройств формирования, передачи, приема и обработки сигналов, развитие навыков системного подхода к разработке и проектированию радиоаппаратуры и средств связи в используемых в настоящее время диапазонах волн. Дисциплина отличается разнообразием содержания, обилием понятий и методов, причем большую роль играют математические приемы исследования. Прочное овладение ими совершенно обязательно, поскольку они служат логическим фундаментом построения последующих радиотехнических курсов.

Книга написана в предположении, что студенты, приступая к изучению курса, имеют хорошую теоретическую и практическую подготовку по следующим дисциплинам: «Высшая математика», «Физика», «Информатика», «Физические основы микроэлектроники», «Общая электротехника и электроника», «Схемотехника электронных средств», «Интегральные устройства радиоэлектроники» и др. Курс «Основы радиоэлектроники и связи» является, в свою очередь, базовым для специальных дисциплин, изучаемых позже. В учебном пособии избран скорее физико-математический, нежели лишь математический (но с должным применением математического аспекта) характер изложения сведений. Вместе с тем авторы использовали достаточно математики, чтобы развить базисное понимание теории и потом проиллюстрировать ее примерами. Работая над текстом, авторы стремились как можно теснее сблизить излагаемый материал с практикой учебной работы в вузе. Это в первую очередь определило принцип отбора материала и степень детальности освещения: на страницы книги вынесено лишь то, что, как показывает практический опыт преподавания, может быть полностью усвоено большинством студентов за отведенное на это время. Изучение конкретных схемотехнических решений, их сравнительный анализ — все это относится уже к специальным дисциплинам, изучаемым позднее. Отметим, что теоретический аппарат ряда новейших положений радиотехники и теории связи, изложенных в учебном издании, несколько упрощен: это касается основ теории кодирования, вейвлетной и фрактальной обработки информации.

Авторы старались учитывать фундаментальные и широко известные работы, последние публикации и материалы, которые имеют оригинальный характер или содержат подробное изложение некоего круга вопросов (что по-английски называют *expository papers*). Среди наиболее значимых работ отметим учебник С.И. Баскакова (некоторые вопросы авторами описаны на основе его прекрасной методологии), книги И.С. Гоноровского, Д.Д. Кловского и др. Из последних работ упомянем книгу А.Б. Сергиенко «Цифровая обработка сигналов», отличающуюся оригинальностью и полнотой изложения, хотя, на наш взгляд, перенасыщенную программными продуктами, иногда усложняющими изложение материала, а также не всегда четко поясненными. Авторы экономили на ссылках на первоисточники, но надеются, что даже простое упоминание некоторых результатов и далеко не полный список литературы окажутся полезными как исходная информация о том, что делается сейчас в радиотехнике и связи.

Терминология и условные буквенные обозначения в книге в основном соответствуют современным стандартам. Однако есть и устоявшиеся в учебной литературе обозначения, отличные от рекомендованных, что связано с удобством их восприятия. Это касается и международной терминологии. Наука и техника интернациональны и бурное развитие информационных технологий сделало это еще более очевидным. Поэтому тем, кто занимается радиотехникой и связью, приходится иметь дело с зарубежными, прежде всего англоязычными, источниками. Все это требует знания терминологии, которая не всегда соответствует принятой в нашей литературе. На русский язык эти термины часто не переводят, а в текстах используют «кальку» с английского. Чтобы облегчить студентам работу с англоязычной литературой, при введении новых понятий в книге приводятся соответствующие английские термины.

Авторы в большом долгу перед преподавателями, аспирантами и студентами МИРЭА за поддержку, оказанную при написании многих разделов книги. Мы многим обязаны профессору В.М. Свистову, который провел много часов за просмотром рукописи, и его рекомендации были бесценными. С нескрываемым удовольствием отмечаем работу профессоров А.В. Прохорова и В.А. Каплуна, взявших на себя нелегкий труд по рецензированию учебного пособия. В результате они высказали ряд замечаний, несомненно, улучшивших книгу. Хотим поблагодарить за полезные советы профессоров В.К. Битюкова и В.И. Каганова.

При подготовке книги к изданию авторы приложили много усилий к тому, чтобы устранить замеченные погрешности. Если внимательные читатели обнаружат неточности в книге, то мы полностью берем на себя моральную ответственность за их появление. Вместе с тем заранее просим прощения у авторов, с материалами учебников, монографий и статей которых возможно совпадают фрагменты некоторых вопросов, примеров, алгоритмов и рисунков.

С глубокой признательностью примем любые полезные замечания, комментарии, предложения и пожелания, которые просим также присылать по адресу: 127994, Москва, Неглинная ул., 29/14, стр. 1, издательство «Высшая школа».

*Авторы*

Каждое десятилетие в радиоэлектронике имеет свое ключевое понятие:

50-е XX в. — транзистор;
60-е XX в. — космическая электроника;
70-е XX в. — микросхема;
80-е XX в. — персональный компьютер;
90-е XX в. — Интернет;
2000-е — мобильная связь;
2010-е — ГЛОНАСС и GPS;
2020-е — полимерные микросхемы и нанотехнологии.

# ВВЕДЕНИЕ

*Радиоэлектроника* — собирательное название обширного комплекса областей науки и техники, связанного с проблемами передачи, приема и преобразования информации с помощью электромагнитных колебаний радиочастотного диапазона. Радиоэлектроника образовалась в результате синтеза радиотехники и электроники. Понятие «радиоэлектроника» появилось в 50-х гг. XX в. и в некоторой степени условное. Радиоэлектроника охватывает радиотехнику, радиофизику и электронику, а также ряд новых областей, выделившихся в результате их развития и дифференциации. В основном радиоэлектроника «обязана» успехам развития радиотехники.

*Радиотехника* (от лат. *radio* — испускаю лучи; от греч. *téchnē* — искусство, мастерство) является основным фундаментом радиоэлектроники, и поэтому часто под термином «радиоэлектроника» понимают радиотехнику. Бурное внедрение в последние годы цифровых технологий во многом изменило как смысл самого понятия «радиотехника», так и требования, предъявляемые к подготовке специалистов в этой области, сделав необходимыми новые знания и умения. Что же представляет собой современная радиотехника?

Главная задача радиотехники состоит в передаче информации на расстояние с помощью электромагнитных колебаний (радиоволн). Вместе с тем радиотехника — многоплановая научно-техническая дисциплина.

В *научном* аспекте радиотехника занимается проблемами, связанными с передачей и приемом информации по радиотехническому каналу, т. е. исследованием распространения «информационных» электромагнитных колебаний в свободном пространстве, анализом, расчетом и оптимизацией радиотехнических схем, устройств и систем, исследованием протекающих в них процессов.

В *техническом* аспекте радиотехника связана с разработкой разнообразных систем, предназначенных для передачи и приема информации с помощью электромагнитных колебаний (в том числе и оптических).

К их числу относятся:

• системы звукового и телевизионного радиовещания;

• глобальные космические (спутниковые) системы радиосвязи, телевизионного вещания и радионавигации;

• системы подвижной радиосвязи с помощью наземных средств — сотовая, профессиональная (транкинговая), пейджинговая и беспроводная связь;

• системы связи с воздушными, подвижными наземными объектами, морскими надводными и подводными судами и другие виды радиосвязи;

• системы радиоуправления, биотелеметрии и радиотелеметрического контроля разнообразных объектов;

• радиотехнические системы комплексов радиолокационной, противовоздушной и противоракетной обороны;

• метеорологические и информационно-измерительные системы и системы различного мониторинга, в том числе космического;

• мультимедийные и прочие системы.

К радиотехнике относятся также радиоастрономия, радиография, радиовидение, радиоразведка и радиопротиводействие, промышленная электроника и радиотехника, медицинская радиотехника и пр.

По мере развития радиотехники из нее выделился ряд совершенно новых областей науки и техники: квантовая электроника, уже упоминавшиеся твердотельная электроника и микроэлектроника, оптоэлектроника, инфракрасная техника, криогенная электроника, акусто- и магнитоэлектроника, хемотроника, голография, биоголография и многое другое. По существу и вычислительная техника строится полностью на элементной базе радиотехники.

Радиотехника занимает некоторое промежуточное положение между науками фундаментального профиля (физикой, радиофизикой, электроникой, оптикой, квантовой физикой) и техническими науками (электротехникой, автоматикой, вычислительной техникой, технической кибернетикой). Разумеется, отмеченное расположение различных отраслей науки является в достаточной степени условным вследствие того, что они во многих своих направлениях и компонентах пересекаются, взаимно проникают друг в друга, обогащая соседние области и идеями, и их техническими решениями. Действительно, в радиотехнической аппаратуре давно и широко используют различные системы автоматического управления и регулирования, например автоматическую регулировку усиления, автоматическую подстройку частоты и фазы, системы, обеспечивающие неизменность выходных параметров в зависимости от действительной на данный момент ситуации в месте приема информации (например, изменение напряжения в сети питания), изменение режима работы радиоаппаратуры (скажем, автоматический переход на другую рабочую частоту при нарушении или резком ухудшении условий прохождения радиоволн на трассе), управление летательными аппаратами, слежение за целью и т. п. С другой стороны, радиотехническая аппаратура является непременной составной частью систем автоматического управления и регулирования (например, полетами самолетов, ракет, космических кораблей, режимом работы единой энергетической системы и др.).

В *математическом* аспекте радиотехника использует такие разделы математики, как линейные и нелинейные интегро-дифференциальные уравнения, математический анализ, матричная алгебра, дискретная математика, численные методы, комбинаторика, теория функций комплексного переменного, нелинейное программирование, теория случайных процессов и др.

В *конструкторско-технологическом* аспекте радиотехника опирается на автоматизированное проектирование радиотехнических устройств и систем. При этом современные радиотехнические устройства представляют собой узлы и сборки из микросхем, транзисторов, диодов, резисторов, конденсаторов, катушек индуктивностей и других элементов, соединенных между собой согласно определенной электрической схеме. Большинство современных конструкций полностью состоят из гибридных и интегральных микросхем.

Какие же задачи решает современная радиотехника? При разработке большинства радиотехнических устройств и систем приходится решать ряд задач, которые формально реализуются следующей логической цепочкой: объект — математическая модель — расчет — анализ — оптимизация — синтез.

Работа *объекта* может быть определена в виде его отклика (реакции) на входное воздействие для установившегося или переходного состояний.

Для исследования какой-либо электрической цепи, каскада, радиотехнического устройства или системы создают *математическую модель* объекта, в которой отображены наиболее существенные его свойства и признаки. Модель представляет собой отражение системы, его образ, используемый для решения задач анализа и синтеза реальной системы. В зависимости от задач и целей моделирования оно может производиться на различных уровнях абстракции. Модель используется для последующих теоретических и экспериментальных исследований объекта. В процессе этих исследований модель может совершенствоваться для более полного отражения свойств реальной системы.

Форма аналитической записи модели может быть отражена формулами, интегро-дифференциальными уравнениями, матрицей, графиками или таблицами. Можно использовать передаточные функции, переходные, импульсные, частотные, спектральные, корреляционные и прочие характеристики. Такая аналитическая модель должна, с одной стороны, с требуемой точностью отражать физические процессы в исследуемом объекте, а с другой — быть удобной для использования на компьютере. В одних случаях модель формируют на основе *аналитического анализа* физической модели объекта, в других — путем проведения *экспериментов*. Возможно получение математической модели и по результатам *экспериментально-аналитических исследований*. Как правило, практически во всех видах обработки имеющегося массива данных, характеризующего работу цепи, каскада, устройства или системы, используются компьютеры.

*Расчет* проводится на компьютере и включает в себя три процедуры: разработку алгоритма, составление программы и собственно численный счет на компьютере. По существу расчет заключается в определении параметров или характеристик цепи, каскада, устройства или всей системы с использованием их математических моделей. Алгоритм разрабатывают на основе определенных

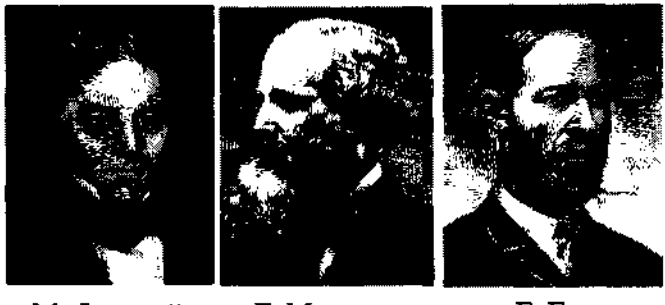
зависимостей, и он должен позволять находить отклик объекта на внешнее воздействие при неизменной структуре и фиксированных внутренних параметрах. В результате составляются методики расчета параметров и характеристик самых разнообразных радиотехнических цепей, каскадов, устройств или систем.

*Анализ* заключается в решении различных вопросов: исследования переходного процесса и установившегося режима работы цепи; определения отклика устройства на изменение его внутренних параметров или внешнего воздействия; исследования условий устойчивости или прохождения сложных сигналов и т. д.

*Оптимизация* состоит в нахождении такой комбинации значений внутренних параметров или характеристик объекта, при которой один или несколько его внешних параметров или характеристик имеют наилучшее значение согласно выбранному критерию. При этом составляется функция цели, в концентрированной форме отражающая конечный смысл решаемой задачи, а именно поиск оптимальной характеристики объекта с учетом определенных ограничений.

*Синтез* заключается в определении структуры разрабатываемого объекта и значений параметров его элементов, при которых, согласно выбранному критерию, она наилучшим образом отвечает необходимым требованиям. Нетрудно заметить, что оптимизацию структуры объекта можно рассматривать как частный случай синтеза. Более того, при оптимизации с перебором нескольких, наиболее подходящих структур объекта она практически сливается с синтезом. Поэтому оптимизацию часто называют *параметрическим синтезом*.

История радиотехники чрезвычайно интересна. Радио возникло на основе фундаментальных открытий в области физики и электротехники, сделанных в XIX в. Его создание явилось заключительным этапом длинной цепи теоретических исследований, опытов и технических разработок в области электричества и магнетизма. В начале XIX в. была создана наука об электромагнитных явлениях, которая и стала фундаментом радиотехники. У ее истоков лежит величайшее открытие электромагнитного поля, связанное с именами трех выдающихся ученых: Майкла Фарадея (*Michael Faraday*; 1791–1867; английский физик), установившего явление электромагнитной индукции (1831 г.); Джеймса Максвелла (*James Maxwell*; 1831–1879; английский физик), создавшего теорию электромагнитного поля (1865 г.); Генриха Герца (*Heinrich Hertz*; 1857–1894; немецкий физик), впервые экспериментально получившего вызываемые колебательным раз-



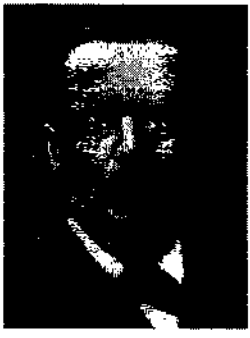
М. Фарадей Д. Максвелл Г. Герц

рядом электромагнитные волны (1887), описываемые теорией Максвелла. Эти открытия были подготовлены множеством других ученых и изобретателей.

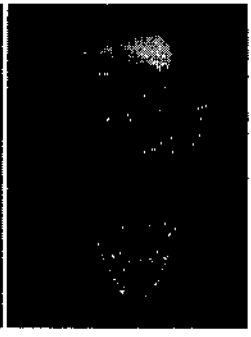
Майкл Фарадей в 1821 г. обнаружил вращение магнита вокруг проводника с током и вращение проводника с током вокруг магнита, на основании чего открыл в 1831 г. явление электромагнитной индукции, ставший основой электротехники. В 1873 г. Джеймс Максвелл опубликовал «Трактат по электричеству и магнетизму». Из составленных им уравнений электродинамики следовал вывод о возможности распространения электромагнитных волн в свободном пространстве со скоростью света. Однако полученному теоретическим путем открытию электромагнитных волн мало кто поверил, даже известные в ту пору физики. Спустя 15 лет Генрих Герц экспериментальным путем подтвердил теорию Максвелла. Суть опытов Герца состояла в следующем. К двум латунным стержням с малым зазором между ними подключалась индукционная катушка, создающая высокое напряжение (Герц назвал это устройство осциллятором). Когда это напряжение превышало напряжение пробоя, в зазоре проскакивала искра и происходило возбуждение и излучение электромагнитных колебаний. Излученные колебания регистрировались на расстоянии в несколько десятков метров, что неопровержимо доказывало распространение электромагнитных волн. Герцем была получена минимальная длина волны $\lambda = 60$ см. В современном представлении осциллятор Герца есть открытый колебательный контур, в котором при возбуждении его искровым способом возникают затухающие колебания, излучаемые в пространство. Все это и предопределило появление радио.

Более века минуло, с тех пор как было изобретено радио, и все это время ведутся споры по установлению его авторства. В России твердо уверены, что радио изобрел известный российский ученый Александр Степанович Попов (1859–1905), на Западе — что итальянец Гульельмо Маркони (*Guglielmo Marconi*; 1874–1937). Сейчас трудно установить историческую истину. Важно то, что мир получил уникальнейшее средство коммуникации. И все же мы можем гордиться, что именно наш соотечественник Александр Степанович Попов 7 мая 1895 г. впервые в мире продемонстрировал беспроводную связь.

Вместе с тем, пытаясь ответить на сложный вопрос, кто все-таки изобрел радио, А.С. Попов или Г. Маркони, следует обратиться к историческим фактам.



А.С. Попов



Г. Маркони



К. Браун



О. Лодж Н. Тесла Т. Эдисон

И тут ради справедливости необходимо отметить, что впервые передал электрический сигнал на небольшое расстояние без проводов английский физик Оливер Лодж (*Oliver Lodge*; 1851–1940). Опираясь на труды М. Фарадея, Д. Максвелла и Г. Герца, он летом 1894 г. продемонстрировал публике эксперимент по трансляции электрического сигнала на расстояние в 150 ярдов (около 137 м) без проводов. Почему О. Лодж не изобрел радио? Сам он так объяснил этот факт: «Я был слишком занят работой, чтобы браться за развитие телеграфа или любого другого направления техники. У меня не было достаточного понимания того, чтобы почувствовать, насколько это окажется экстраординарно важным для флота, торговли, гражданской и военной связи».

Наряду с О. Лоджем пионером беспроводной передачи электрического сигнала является также Никола Тесла (*Nikola Tesla*; 1856–1943). Н. Тесла в 1888 г. продемонстрировал нью-йоркской публике радиоуправляемую модель маленькой лодки на озере в «Мэдисон Сквер Гарден». Он также предлагал на основе тех же принципов модулирования радиоволн передавать речь, музыку и даже изображения. Стоит отметить, что проект первой системы беспроводной связи с использованием незатухающих колебаний низкой частоты предложен в 1885 г. Томасом Эдисоном (*Thomas Edison*; 1847–1931; США).

Анализ различных специальных источников показывает, что А.С. Попов первым публично продемонстрировал действующее радио и сделал сообщение о своем изобретении. А вот Г. Маркони преуспел в деле патентования и продвижения в промышленное производство созданных им радиотелеграфных аппаратов. Установлено, что работы по радио А.С. Попова опубликованы в январе 1896 г., а работы Г. Маркони — лишь в июне 1897 г. 7 мая 1895 г. А.С. Попов продемонстрировал на заседании физического отделения Русского физико-химического общества первый в мире чувствительный радиоприемник (названный грозоотметчиком), принимавший колебания, излучаемые видоизмененным осциллятором Герца. Этот день в нашей стране отмечается как День радио. Отчет о знаменательном заседании с описанием доклада и эксперимента А.С. Попова был опубликован в журнале Русского Физико-химического общества в августе 1895 г. и январе 1896 г. 24 марта 1896 г. на заседании того же общества А.С. Попов, помимо радиоприемника, демонстрирует разработанный им искро-

вой радиопередатчик, передав из одного здания в другое на расстояние 250 м азбукой Морзе (привычная «морзянка») первую в мире радиотелеграмму, которая содержала всего два слова: имя и фамилию первооткрывателя электромагнитных волн — «Генрих Герц». Этой телеграммой А.С. Попов продемонстрировал дань уважения своему предшественнику. В этом же году Александр Степанович установил радиосвязь между кораблем «Россия» и берегом на расстоянии 5 км. В 1897 г. при испытаниях на кораблях дальность связи с помощью аппаратов Попова достигла 5 км. В 1899 г. Попов осуществил радиосвязь на расстоянии 52 км, обеспечившую работы по спасению броненосца «Генерал-адмирал Апраксин», севшего на камни. Было передано 440 радиограмм (6 300 слов). В 1900 г. на IV Всемирном электротехническом конгрессе в Париже А.С. Попову за изобретение радио были присуждены почетный диплом и золотая медаль.

Несколько в ином ключе действовал другой изобретатель радио — Г. Маркони, применивший электромагнитные волны для радиосвязи после опубликования работ А.С. Попова. Принято считать, будто изобретение совершается, если в нем есть потребность. И вот в смысле создания потребности бесспорным изобретателем радио следует признать Г. Маркони. 2 июня 1986 г. Г. Маркони подал в Великобритании предварительную заявку под названием «Усовершенствования в передаче электрических импульсов и сигналов и в аппаратуре для этого» (научные публикации Попова на этот счет появились в том же месяце, но адресовались совершенно иной аудитории). Британский патент выдан 2 июля 1897 г. Тогда же он создает крупное акционерное общество «Маркони и К$^0$», сумев привлечь к своему изобретению значительные финансовые средства. Изобретением Г. Маркони заинтересовались почтовое ведомство и адмиралтейство Великобритании. В Великобритании Г. Маркони зарекомендовал себя не только великим изобретателем, но и крупным предпринимателем, быстро и эффективно внедрившим в промышленное производство изобретенные им радиотелеграфные аппараты. В 1901 г. с помощью аппаратов Г. Маркони была установлена радиосвязь через Атлантический океан с Америкой.

Заметим, что все это было *бессловесное* радио: первые радиопередатчики не умели передавать нормальный звук — оператор мог лишь включать и выключать электрический сигнал телеграфным ключом, т. е. передавать текст азбукой Морзе. Начальный период развития радио характеризуется применением искровых радиопередатчиков, формирующих затухающие колебания и служивших для передачи телеграфных сигналов. Эти генераторы создавали большие взаимные помехи. По сути, как и радиотелеграф А.С. Попова, это была первая дискретная (цифровая) модуляция электромагнитных волн, достоинства которой (малая полоса частот в эфире и высокая помехоустойчивость) позволили ей дожить до третьего тысячелетия.

Эпоха радио для слушателей началась лишь после изобретения известным американским инженером Луи де Форестом (*Lee de Forest*; 1873–1961) трехэлектродной электронной лампы — триода (от греч. *tri* — три и электрод). Триод позволил создавать ламповые генераторы для получения мощных незатухающих электромагнитных колебаний в передатчиках, усиливать их, модулиро-


Л. Форест


Р. Фессенден

вать передаваемый сигнал по амплитуде и выполнять различные преобразования при передаче сообщений любой формы и сложности. 24 декабря 1906 г. американский инженер Реджинальд Фессенден (*Reginald Fessenden*; 1866–1932) методом амплитудной модуляции с помощью машин высокой частоты осуществил первую в истории радиопередачу из своего дома в Брэнт Рок, штат Массачусетс, США, чем несказанно удивил радистов на морских судах. По случаю Рождества он сам сыграл на скрипке произведение Гуно «О, святая ночь», а потом прочел отрывки из Евангелия от Луки. Так слово стало Богом. В конце года со станции «Брэнт Рок» (шт. Массачусетс, США) была осуществлена передача первой радиовещательной программы, в которой можно было услышать речь, музыку и пение.

За изобретение и развитие радио Г. Маркони в 35 лет (совместно с профессором из г. Страсбурга Карлом Брауном; *Karl Braun*; 1850–1918) в 1909 г. была присуждена Нобелевская премия по физике (за работы по развитию беспроволочного телеграфа). За три года до этого события скончался А.С. Попов. Поскольку по уставу Нобелевская премия присуждается только при жизни изобретателя или ученого, то кандидатура А.С. Попова не выдвигалась и не рассматривалась. Кстати, вклад К. Брауна в радиотехнику достаточно весом. В 1898 г. он предложил заменить прямой метод возбуждения электромагнитных колебаний в антенне косвенным. С этой целью искровой передатчик помещался в отдельном колебательном контуре и индуктивно связывался с антенной. Такая конструкция передатчика позволила вынести разрядник из антенной цепи и тем самым уменьшить потери в антенне.

*Радиофизика* — раздел физики, в котором изучаются физические основы радиотехники. Важнейшими проблемами радиофизики являются исследование возбуждения и преобразования электрических сигналов и помех, а также излучения и распространения электромагнитных колебаний.

Развитие радиотехники непосредственно связано с созданием элементной базы, в частности, с разработкой электронных приборов для систем передачи информации на расстояние с помощью электромагнитных колебаний. Дальнейшее развитие радиотехники непрерывно ставило задачи по созданию и внедрению новых электронных элементов и узлов, что привело к появлению самостоятельной отрасли науки — электроники.

*Электроника* — наука о взаимодействии заряженных частиц (электронов, ионов) с электромагнитными полями и методах создания электронных приборов и устройств, используемых в основном для передачи, хранения и обработки информации, возникла в начале XX в. Первоначально развивалась вакуумная электроника; на ее основе были созданы электровакуумные приборы.


Д. Килби


Ж.И. Алферов


Г. Кремер

Разработка в начале 50-х гг. полупроводниковых приборов и в 60-х гг. XX в. интегральных микросхем позволила резко снизить массу и габаритные размеры радиотехнической аппаратуры, при одновременном повышении ее надежности и значительном уменьшении энергопотребления. При этом электроника четко разделилась на энергетическую или силовую электронику (мощные выпрямители, инверторы и т. д.) и микроэлектронику. После создания молекулярного генератора (1955 г.) началось развитие квантовой электроники.

*Микроэлектроника* — раздел электроники, связанный с созданием интегральных схем — неделимых изделий, выполняющих определенные функции по преобразованию и обработке сигналов и имеющих высокую плотность упаковки электрически соединенных элементов. Начало современной микроэлектроники в 1958 г. изобретением микросхемы положил Джэк Килби (*Jack Kilby*; р. 1923; Нобелевская премия совместно с Жоресом Ивановичем Алферовым (р. 1930) и Гербертом Кремером (*Herbert Kroemer*; р. 1928)).

Основной тенденцией микроминиатюризации является «интеграция» электронных схем, т. е. стремление к одновременному изготовлению большого количества элементов и узлов электронных схем, неразрывно связанных между собой. Поэтому из различных областей микроэлектроники наиболее эффективной оказалась интегральная микроэлектроника. Современные сверхбольшие аналоговые интегральные микросхемы и цифровые микросхемы и микропроцессоры на одном кристалле заменяют целые блоки и устройства радиоаппаратуры предшествующего поколения. Огромный вклад в развитие современной микроэлектроники внес Ж.И. Алферов (открыл явление сверхинжекции в гетероструктурах, предложил принципы использования гетероструктур в микроэлектронике).

В последние годы большое развитие получили квантовая электроника и оптоэлектроника.

*Квантовая электроника* (*квантовая радиофизика*) — область науки и техники, охватывающая исследования принципов действия, конструирование и применение генераторов, усилителей, преобразователей частоты электромагнитного излучения, действие которых основано на вынужденном излучении или на нелинейном взаимодействии излучения с веществом. Основоположниками кван-


А.М. Прохоров


Н.Г. Басов


Ч. Таунс

товой электроники считают лауреатов Нобелевской премии по физике (1964 г.; «За фундаментальные работы в области квантовой электроники, приведшие к созданию генераторов и усилителей на основе принципа мазера – лазера») русских ученых Александра Михайловича Прохорова (1916–2002), Николая Геннадиевича Басова (1922–2001) и американского физика Чарльза Таунса (*Charlz Tauns*; р. 1915). К устройствам квантовой электроники относятся прежде всего квантовые усилители сверхвысокой частоты и лазеры (*LASER* — light amplification by stimulated emission of radiation).

*Акустоэлектроника* — раздел акустики, на стыке акустики твердого тела, физики полупроводников и радиоэлектроники, занимающийся исследованием принципов построения ультразвуковых устройств для обработки сигналов.

Новая отрасль радиоэлектроники, получившая название акустоэлектроники, сулит заманчивые перспективы создания миниатюрных и надежных частотно-избирательных систем. Большую роль в развитии оптоэлектроники (акустоэлектроники, акустооптики) сыграли русские академики Юрий Васильевич Гуляев (р. 1935) и Владислав Иванович Пустовойт (р. 1936). Ю.В. Гуляев (совместно с В.И. Пустовойтом) впервые высказал идею использовать *поверхностные акустические волны* (ПАВ) для обработки сигналов и предложил слоистую структуру пьезоэлектрик–полупроводник в качестве базовой конструкции. Юрий Васильевич Гуляев (независимо и одновременно с американским физиком Дж. Блюстейном; *G. Blaustein*) открыл новый фундаментальный тип ПАВ в пьезоэлектрических материалах, которые известны в мировой литературе как волны Блюстейна–Гуляева.


Ю.В. Гуляев


В.И. Пустовойт

*Теория связи* — в настоящее время наиболее динамично развивающаяся отрасль науки. Прогресс в развитии средств связи (и вычислительной техники) привел к переходу в промышленно развитых странах от общества индустриального к обществу информационному. Электросвязь относится к сфере

высоких технологий, и ее развитие в значительной степени обусловлено достижениями в области теории передачи информации. Причем новейшие системы связи — *модемные* технологии, реализованные под конкретные скорости передачи. Можно выделить несколько разделов теории связи, оказавших существенное влияние на создание систем радиосвязи в XX в.: теория сигналов, к которой относятся спектральный и корреляционный анализ, теория аналитического сигнала и теория модуляции, теорема отсчетов и цифровая обработка сигналов; статистическая радиотехника; теория информации и теория потенциальной помехоустойчивости, теория кодирования. Эти разделы отражают фундаментальные законы передачи и приема аналоговых и цифровых сигналов в каналах связи и дают адекватный математический аппарат для описания этих законов.

Системы связи играют все большую роль в жизни людей, объединяя и сближая отдельные страны, континенты и объекты космоса. Последние годы отмечены интенсивным развитием как проводных, оптико-волоконных систем связи, спутниковых, так и подвижных систем радиосвязи. Новейшие разработки систем связи используют не только возможности современных технологий, но и достижения современной теории связи, позволяющие повысить не только объемы передаваемой информации, но и качество передачи сообщений.

Бог наделил человека зрением, благодаря которому он получает около 90 % информации об окружающем мире, используя при этом только электромагнитные волны оптического диапазона (свет). Вместе с тем во многих случаях применение такого рода электромагнитных волн для передачи информации оказалось недостаточным, и люди разработали и создали искусственные (технические) системы связи. Из истории известно, что еще в XII в. до нашей эры известие о победе над Троей пришло в Древнюю Грецию с помощью электромагнитных (световых) волн — разработанной эллинами системы зажженных костров, расположенных на расстоянии прямой видимости. Об этом упоминается в пьесе Эсхила «Агамемнон», где дозорный, лежа на плоской крыше царского дворца в Аргосе, говорит: «А я все жду условленного светоча, столба огня, который возвестит, взыграв, что пала Троя». После победы такой столб огня появился. Поэтому можно констатировать, что древние греки реализовали первый вариант «радиорелейной» (радиорелейные линии — это цепь приемопередающих станций, удаленных друг от друга на 40 … 50 км) линии связи. Во времена Французской революции (XVIII в.) для передачи информации применяли гелиограф, состоящий из системы зеркал, с помощью которых свет Солнца можно было направить в место приема и таким образом передать некоторое сообщение. По существу гелиограф являлся вторым вариантом радиорелейной линии связи.

Разработки систем связи последнего времени используют не только возможности современных технологий, но и новейшие достижения теории связи, позволяющие повысить как объемы передаваемой информации, так и качество передачи сообщений (верность связи).

В настоящее время важнейшее значение в жизни человеческого общества имеет *электросвязь* — передача сообщений (информации) с помощью электромагнитных колебаний. Совокупность оборудования для передачи/приема сигна-

лов электросвязи и среды распространения этих сигналов называется *системой электросвязи* или *телекоммуникационной системой* (от греч. *tele* — вдаль и лат. *communication* — связь; здесь — связь на расстоянии). Совокупность телекоммуникационных систем, объединенных по принципам территориального расположения, функционирования и подчиненности, называется *сетью электросвязи* (*сетью связи*) или *телекоммуникационной сетью*.

Отдельные виды и соответствующие им службы электросвязи относятся к передаче сообщений определенного типа: телефонных, телеграфных, цифровых данных, факсимильных (неподвижные изображения), телевизионных, видеотелефонных, газетных полос для децентрализованной печати, звукового вещания и др. Основной объем передаваемых сообщений — основной *трафик* (информационная нагрузка, передаваемые потоки информации, количество информации, поступающей через сеть связи) — пока составляет телефонная связь.

В истории развития систем электрической связи можно выделить несколько характерных этапов.

*Первый этап* характеризуется появлением безголосовой электрической связи (т. е. передачи письменных сообщений электрическим способом) — созданием первого электромагнитного телеграфа, изобретенного в 1832 г. русским ученым Павлом Львовичем Шиллингом. Прибор, созданный Шиллингом, имел стрелочную индикацию передаваемых по электрическим проводам сигналов, которые легко расшифровывались в буквы оператором приемного телеграфного аппарата, согласно разработанной Шиллингом специальной таблице кодов.

В 1837 г. американский изобретатель Сэмуэл Морзе (*Samuel Morse*; 1791–1872; немного позже предложил телеграфный код — *азбуку Морзе*) создал электромагнитный телеграфный аппарат с записью сигналов на бумажной ленте. В этом же году Морзе отправил по телеграфу первую в мире телеграмму: «Что же написал Господь Бог». Задача увеличения дальности связи была решена русским академиком Борисом Семеновичем Якоби (Мориц Герман; 1801–1874), предложившим в 1858 г. телеграфную трансляцию. Тогда же изобретением дуплексного телеграфирования было положено начало повышению эффективности использования линейных сооружений. В 1869 г. им предложен способ одновременного телеграфирования по общей электрической цепи токами разных частот.



П.Л. Шиллинг С. Морзе Б.С. Якоби А. Белл

Изобретение телефона в 1876 г. явилось отправной точкой в развитии телефонной связи, а также методов и технологий передачи голоса. Изобретателем телефонной связи является американский инженер Александр Белл (*Alexander Bell*; 1847–1922), который предложил применять для передачи речи на расстояние электромагнитный прибор, названный телефоном. В 1878 г. был разработан телефонный аппарат с угольным микрофоном, а также предложено использовать в передающей части схемы трансформатор, что обеспечивало двустороннюю передачу и большую дальность. В 1880 г. была создана схема одновременной передачи сигналов телеграфирования и телефонирования, основанная на частотном разделении каналов с помощью электрических фильтров.

Разработка и создание практических систем связи стало возможным в результате развития методов радиотехники, изобретения электронных ламп и применения их для усиления, генерации переменных токов, модуляции и демодуляции, разработки теории и методов проектирования электрических фильтров, выравнивателей и других элементов.

*Второй этап* начинается с изобретения радио 7 мая 1895 г. Это позволило в 20-х гг. XX в. построить радиовещательные станции для организации радиовещания. Развитие многоканальной связи начинается с создания дуплексных усилителей в 1915 г. и использования в них идеи двустороннего действия с дифференциальной системой соединения, которая до сих пор остается основой построения дуплексных усилителей каналов тональной частоты.

*Третий этап* характеризуется решением проблемы многоканальности. В конце 20-х гг. XX в. был реализован полосовой фильтр, позволяющий выделять одну боковую полосу частот (о фильтрах см. далее).

Основы теории передачи информации были заложены в фундаментальных трудах академика Владимира Александровича Котельникова (1908–2002) по теории потенциальной помехоустойчивости (1947 г.) и американского ученого Клода Элвуда Шеннона (*Claude Elwood Shannon*; 1916–2001) по теории информации (1948 г.). В 1933 г. Дмитрий Васильевич Агеев (1911–1997) опубликовал работу «Основы теории линейной селекции», в которой заложил основы общей теории линейной селекции сигналов и исследовал возможности применения в качестве переносчиков информации совершенно новых, в том числе и неортогональных, сигналов. Его работа предвосхитила некоторые более поздние идеи



К. Шеннон В.А. Котельников Д.В. Агеев А.Н. Колмогоров


Д. Миддлтон


Ф. Вудворд


В.И. Тихонов А.Я. Хинчин

применения широкополосных сигналов в многоканальных системах связи, выдвинутые в 50-х гг. и продвинутые до практического внедрения. Фактически Д.В. Агеев первым разработал системы подвижной связи.

Отдельные вопросы теории информации рассматривались в ранних трудах Х. Найквиста (1928 г.), Р. Хартли (1928 г.), Андрея Николаевича Колмогорова (1941 г.; русский математик; 1903–1987; интересно, что в 2003 г. германские исследователи советских архивов сообщили о том, что теория информации Шеннона была создана А.Н. Колмогоровым еще в 1938 г., но была засекречена, так как использовалась в военных разработках), Александра Яковлевича Хинчина (1894–1959), Н. Винера (1943 г.), А. Вальда (1950 г.). Дальнейшее развитие теория связи получила в трудах С. Райса (S. Rice), Р. Галлагера, Х. Хелстрема, Р. Фано, Л.М. Финка, Д. Витерби, В.И. Сифорова и многих других отечественных и зарубежных ученых. Большой вклад в становление и развитие радиотехники внесли русские ученые Л.И. Мандельштам, В.П. Вологдин, М.В. Шулейкин, О.В. Лосев, Н.М. Крылов, Н.Д. Папалекси, А.И. Берг, Ю.Б. Кобзарев, В.И. Тихонов, Л.Л. Расплетин, Н.М. Боголюбов, Р.Л. Стратонович и многие др.

Возникновение теории передачи информации связывают с появлением фундаментальной работы К. Шеннона «Математическая теория связи» (1948 г.). Однако элементы этой теории рассматривались в работах Р. Хартли по измерению количества информации, В.А. Котельникова «О пропускной способности «эфира» и проволоки в электросвязи» (1933 г.) и по оптимальным методам приема сигналов на фоне помех (1946 г.), Д.В. Агеева по основам теории линейной селекции сигналов (1935 г.). Развитию теории передачи информации способствовало и появление теории случайных функций и статистических решений. Они позволили установить строгие количественные соотношения в теории передачи информации и сделать ее точной наукой.

К 50-м годам XX в. стали классическими работы А.Я. Хинчина по теории корреляции стационарных случайных функций (1938 г.), Н.Н. Колмогорова и Н. Винера по интерполированию и экстраполированию стационарных случайных последовательностей (1941 г., 1949 г.). Значительный вклад в развитие отдельных разделов теории передачи информации внесли Д. Миддлтон (D. Middleton; р. 1920), Ф. Вудворд (P. Woodward), В.И. Бунимович (1904–1981); Р. Хемминг (1915–1998), А.А. Харкевич, Л.С. Гуткин, Л.Ф. Бородин и др.

# ЧАСТЬ I

# РАДИОТЕХНИЧЕСКИЕ СИСТЕМЫ И СИГНАЛЫ

## Глава 1. ОСНОВНЫЕ ПРИНЦИПЫ ПЕРЕДАЧИ И ПРИЕМА ИНФОРМАЦИИ

Радиоэлектроника сегодняшних дней и новейшие системы связи во многом определяют технический прогресс в различных областях науки, техники, производства и бизнеса. Это обусловлено тем, что с помощью различных радиотехнических систем, и в том числе систем связи, решаются такие проблемы, как передача, извлечение, обработка, хранение и отображение информации, передача команд на управляемые объекты, контроль и обеспечение функционирования измерительных и производственных комплексов. В радиоэлектронике и технике связи перенос информации в пространстве осуществляется с помощью электромагнитных колебаний (волн). Чтобы создать электромагнитные колебания и заложить в них передаваемую информацию, требуется ряд специфических радиотехнических операций. Для получения представления о наиболее типичных операциях, необходимых при передаче информации на расстояние, рассмотрим общие вопросы построения радиоэлектронных (радиотехнических) систем. Почти все проблемы, решаемые при этом, характерны для цепей и устройств, используемых в самых различных радиоэлектронных системах, включая системы радиосвязи.

### 1.1. Общие сведения об информации, сообщениях и сигналах

Радиотехника, радиоэлектроника и радиосвязь тесно связаны с теорией информации и передачей информации на расстояние. В свою очередь мы часто говорим о передаче получателю какого-либо сообщения, в принципе имея в виду при этом передачу информации. В настоящее время понятия «информация» и «сообщение» в повседневной жизни люди употребляют очень часто.

Вместе с тем эти понятия сложны, хотя и близки по смыслу, поэтому дать их точные определения через более простые термины достаточно трудно.

Обратимся к понятию «информация» и выясним, что оно собой представляет. Термином «информация» (от лат. *informatio* — разъяснение, изложение, ознакомление, осведомленность) с древних времен люди обозначали процесс разъяснения, изложения, толкования различных сведений. Позднее так называли и сами сведения, и их передачу пользователю в любом конкретном виде. Интересно, что С.И. Ожегов в своем труде «Словарь русского языка» термин «информация» трактовал как сообщение, осведомляющее о положении дел, состоянии чего-нибудь.

Начало пониманию сущности информации как всеобщего свойства материи было положено Норбертом Винером (*Norbert Wiener*; 1894–1964; великий американский ученый, создатель науки «кибернетика») в его знаменитой монографии «Кибернетика, или управления и связь в животном и машине», изданной в 1948 г. Современная информационная наука находит применение в самых разнообразных областях. Поэтому до сих пор еще нет всеобщего для всех наук классического определения понятия «информация». Фактически в каждом направлении ученые и специалисты используют определение отдельных составляющих информации, наиболее важных для данной науки. Однако под информацией понимают не все получаемые сведения, а только те, которые еще не известны и являются новыми для получателя. Для специалистов по радиотехнике и теории связи важно то, что информация — это сведения, являющиеся, прежде всего, объектом передачи, обработки и использования.

Применяемое в радиотехнике и связи современное понятие «информация» ввели в начале XX в. Ральф Хартли (*Ralph Hartley*; американский инженер; 1887–1974) и К. Шеннон. В широком смысле сейчас под *информацией* подразумевают *совокупность сведений о каких-либо событиях, явлениях или предметах, предназначенных для передачи, приема, обработки, преобразования, хранения или непосредственного использования*. Заметим, что в более узком смысле *информация — содержание сообщения, рассматриваемое в процессе его передачи, восприятия и использования* (Физическая энциклопедия, т. 1, с. 176).

Жизнь современного общества немыслима без обмена информацией. Человеческое общество живет в информационном мире, который постоянно изменяется и пополняется. То, что человек видит, слышит, помнит, знает, переживает, — все это различные формы информации. Как образно отметил К. Шеннон: «Информация — послание, которое уменьшает неопределенность» (есть и такое похожее определение «информация — есть устраненная неопределенность»). Но любое использование информации возможно лишь при условии ее передачи на расстояние. А для радио и телевидения как средств массовой информации возможность передачи является условием их существования.

Информация — нематериальное свойство материи и подчиняется определенным законам. Важнейший из них закон сохранения информации: «Информация сохраняет свое значение в неизменном виде, пока остается в неизменном виде носитель информации — память». А вот носитель информации в общем случае всегда материален и его существование зависит от времени.

*Зависимость информации от времени существования ее носителя — важнейшее информационное свойство материи, которое и является памятью.* Весьма важно то, что в отличие от материального и энергетического ресурсов, информационный ресурс не уменьшается при потреблении, а существенно пополняется и накапливается со временем (ученые считают, что объем человеческих знаний удваивается каждые 10 лет, а объем информации — каждые 5 лет; для сравнения можно отметить, что производительность и быстродействие компьютеров увеличивается вдвое менее чем через 1,5 года). Более того, с помощью специальных технических радиоэлектронных и вычислительных средств информационный ресурс сравнительно просто обрабатывается, преобразуется, хранится и передается на значительные расстояния. Появилась даже новая наука о преобразовании информации — *информатика*.

Часто наряду с информацией, особенно при описании действия цифровых устройств и вычислительных систем, употребляют такое понятие как *данные*. Рассматривают данные как признаки или зарегистрированные наблюдения, которые по каким-то причинам не используются, а лишь хранятся в памяти или на определенном носителе. Когда же появляется возможность использования этих данных для уменьшения неопределенности о чем-либо, они каким-либо определенным образом превращаются в информацию. Поэтому используемые данные являются уже информацией.

Специфической особенностью информации является возможность ее многократного применения для различных целей. В частности, при извлечении информации из памяти человека или компьютера, информация, запомненная или записанная в памяти, не исчезает, а начинает свое самостоятельное существование и может быть использована различным образом. Как правило, с течением времени память ухудшается в результате роста *энтропии* {от греч. *entropia* — круговорот, поворот, превращение; понятие «энтропия» введено в 1865 г. Рудольфом Клаузиусом (*Rudolf Clausius*, немецкий физик; 1822–1888) в термодинамике для определения меры необратимого рассеяния энергии; см. гл. 9} системы памяти, и записанная информация может постепенно стираться. Другая принципиальная особенность информации состоит в том, что обычно ее получают в одном месте, а используют в другом, и поэтому требуется ее передать на какое-то расстояние. *Передача информации — процесс переноса каких-либо сведений из одной точки пространства в другую.*

Прогресс в теории и технике передачи информации тесно связан с развитием теории передачи сигналов, поскольку информация непосредственно не передается, а предварительно преобразуется в сигналы, которые и передаются. То есть информация передается с помощью (и в форме) сигналов. Когда говорят о передаче информации, то имеют в виду лишь форму сообщения, в которую облечена информация, способ преобразования сообщения в сигнал и его передачу по каналам связи. Когда создание информации связано с деятельностью человека, поток информации можно разбить на части и представить их в разной форме. Например, можно изменить порядок слов во фразе или использовать синонимы, получив тем самым разные формы представления одной и той же информации.





Н. Винер Р. Хартли Г. Найквист

Для передачи или хранения любой нужной информации используют различные знаки (символы; от греч. *symbol* — знак — *character*, опознавательная примета), позволяющие представить ее в некоторой форме. Такими знаками могут быть цифры, буквы и элементы алфавита, отдельные слова и фразы человеческой речи, жесты и рисунки, формы электрических или световых колебаний, математические знаки и т. д.

Первой серьезной работой по теории передачи информации считают статью американского связиста Ральфа Хартли (*Ralph Hartley*; 1888–1970) «Передача информации» («*Transmission of Information*»), опубликованную в 1928 г. Хартли сделал открытие, состоящее в том, что информация допускает количественную оценку. Он предложил «количество информации», передаваемое по каналу связи относительно *двух равновероятных исходов* и снимающее неопределенность, оценивать путем принятия одного из них за единицу информации, потом получившую название *бит*. Однако логарифмическая формула Хартли позволяла определять количество информации только для случая, когда появление символов равновероятно и они статистически независимы. Однако эти условия выполняются редко. При определении количества информации необходимо учитывать не только число разнообразных сообщений, которые можно получить от источника, но и вероятность их получения.

Немаловажное значение для теории передачи дискретной информации (по телеграфным линиям) имела работа Гарри Найквиста (*Harry Nyquist*; американский физик; 1889–1976) «Некоторые факторы, воздействующие на скорость телеграфирования» (1924 г.). Но наиболее значимым шагом в становлении теории передачи информации явились работы К. Шеннона «Математическая теория связи» (1948 г.) и «Связь при наличии шума» (1949 г.), развившие идеи Хартли. Шеннон впервые стал исследовать статистическую структуру передаваемых сообщений и действующих в канале связи шумов и, кроме того, анализировал не только конечные, но и непрерывные множества передаваемых сообщений. Шеннон рассматривал информацию как сообщение об исходе случайных событий, о реализации случайных сигналов. Поэтому количество информации ставилось им в зависимость от вероятности наступления этих событий: если сообщение несет сведения о часто встречающихся событиях, вероятность появления которых стремится к единице, то такое сообщение малоинформативно.

К. Шеннон ввел в теорию информации понятие «энтропия источника сообщений», как метрику измерения объема информации. Теория информации Шеннона позволяла ставить и решать задачи об оптимальном кодировании (и модуляции) передаваемых сигналов с целью повышения пропускной способности (*Bandwidth*) каналов связи, подсказывала пути борьбы с помехами на линиях связи и т. д. Введение Шенноном способа измерения количества информации привело к формированию самостоятельного научного направления в электросвязи под названием «Теория информации». По существу теория информации стала основываться на вероятностных, статистических закономерностях явлений и решать задачу максимизации средней скорости передачи информации. Кстати, Шеннон впервые предложил абстрактную структурную схему системы связи, состоящую из пяти элементов: источника информации, передатчика, линии связи, приемника и адресата, и сформулировал теоремы о пропускной способности, помехоустойчивости, кодировании.

Параллельно на основе работ В.А. Котельникова развивалось другое научное направление — *теория потенциальной помехоустойчивости*. В.А. Котельниковым в 1946 и 1956 гг. были опубликованы работы (в частности, докторская диссертация, 1947 г.) по оптимальным методам приема информации и потенциальной помехоустойчивости. Теория потенциальной помехоустойчивости определяет предельные возможности приема сигналов при наличии шумов. Главной задачей теории помехоустойчивости является отыскание таких способов передачи и приема, при которых обеспечивалась бы наивысшая достоверность принятого сообщения. Использование результатов работ В.А. Котельникова дало возможность судить о том, насколько конкретная система передачи информации близка к идеальной по своей структуре и способности выделять сигнал из смеси его с помехами и шумами.

Теория информации решала задачу максимизации средней скорости ее передачи. Главная задача теории помехоустойчивости — отыскание таких способов передачи и приема информации, при которых обеспечивается наивысшая достоверность принятого сообщения. Обе задачи являются, по сути, различными сторонами одного и того же процесса обработки информации при ее передаче и приеме. Фундаментальные работы В.А. Котельникова и К. Шеннона создали основу теории передачи информации (сигналов), которая получила дальнейшее развитие благодаря работам многих ученых по отдельным ее разделам.

Теперь рассмотрим, что же собой представляет сообщение. *Совокупность знаков* (*символов*), *отображающая* (*несущая*) *информацию*, называется *сообщением*. Иногда специалисты сообщение трактуют как информацию, выраженную в определенной форме и подлежащую передаче. Чтобы передать информацию, надо передать содержащие эту информацию сообщение.

Сообщение может быть представлено в виде текста телеграммы, сведений, передаваемых по телефону, радио, телевидению и другим видам радиосвязи, совокупности электронных данных, хранящихся на магнитных носителях — дисках, флэш-памяти (от англ. *Flash* — «вспышка»; перепрограммируемая постоянная энергонезависимая память, допускающая многократную перезапись),

используемых в компьютерах. Последний вид информации получил название *электронной*. Ее роль в повседневной жизни человека и во всех сферах его деятельности — научной, производственной, военной, торговой, финансовой и др. — с каждым годом все более возрастает, что позволяет утверждать, что XXI в. будет столетием информационных технологий. *Internet* (Интернет, см. далее) — глобальная мировая сеть, объединившая миллионы компьютеров и позволяющая обмениваться информацией миллиардам людей, является наглядным тому подтверждением. В телевизионных (*телевидение* — от греч. *τήλε* — далеко и лат. *video* — вижу; в данном случае — *далеко видеть*) системах при передаче движущихся изображений сообщение представляет собой изменение во времени яркости элементов изображения. Сообщения чаще всего являются функциями времени, например речь в телефонных переговорах, температура, влажность, давление при передаче телеметрических (*телеметрия* — измерение каких-либо параметров объекта на расстоянии) данных, телевизионный спектакль и т. п. Однако сообщение может и не быть функцией времени (например, телеграмма, неподвижное изображение и пр.).

Передают сообщение с помощью материального носителя. Например, при передаче сообщения по почте носителем служит бумага. В радиотехнике и радиосвязи носителями являются различные сигналы. Причем для передачи информации используются специфические сигналы — физические процессы, значения параметров которых отражают передаваемые сообщения. В качестве сигнала можно использовать любой физический процесс, изменяющийся в соответствии с переносимым сообщением.

Итак, что же такое сигнал (от лат. *signum* — знак)? В наиболее упрощенной и известной формулировке сигнал — зависимость одной физической величины от другой (т. е. с математической точки зрения сигнал является *функцией* какого-либо аргумента).

*Сигнал* — физический процесс (или явление), несущий информацию о состоянии какого-либо объекта наблюдения. По существу значения параметров несущего сигнала отражают передаваемое сообщение. Сигнал переносит информацию в пространстве и во времени и это материально-энергетическая форма представления информации.

Цепь информация – сообщение – сигнал — это пример процесса радиотехнической обработки, необходимой там, где находится источник информации. На стороне потребителя информации осуществляется обработка в обратном порядке: сигнал – сообщение – информация.

По своей физической природе радиотехнические сигналы бывают электрическими, электромагнитными, оптическими, акустическими, магнитостатическими и др. В радиотехнике, радиоэлектронике и системах связи в основном используют электрические (в последние годы и оптические) сигналы. Физической величиной, характеризующей электрический сигнал, является напряжение, несколько реже ток (иногда мощность). Для сигнала возможны и ряд других определяющих физических величин, например, зависимость давления воздуха в точке от времени можно характеризовать как звуковой сигнал. Зависимость яркости

от положения точки на плоскости можно рассматривать как черно-белое изображение. В системах оптической обработки информации сигналом может являться зависимость интенсивности света от пространственных координат изображения. Чаще рассматривают зависимости от времени, хотя это не обязательно. При временно́м подходе аналитическим описанием конкретного сигнала может быть некоторая функция времени $u(t)$. Определив каким-либо образом эту функцию, можно определить и сигнал. Однако на практике полное описание сигнала не всегда требуется. Часто достаточно более общего описания в виде нескольких параметров сигнала, характеризующих основные его свойства. В книге далее везде подразумевается (если иное не оговорено специально), что электрический сигнал $u(t)$ представляет зависимость *напряжения от времени.*

Сигналы, отражающие информацию, могут воздействовать на преобразователи и усилители сигналов. Преобразователи сигналов делятся на два класса. На преобразователи одного класса воздействует физический процесс одной природы (например, звуковой сигнал), а на выходе получается сигнал другой природы (в частности, электрический сигнал на выходе микрофона, телевизионной камеры и пр.). В преобразователях (и усилителях) другого класса осуществляется, как правило, преобразование (и усиление) электрических сигналов без изменений их физической природы.

Если необходимо получить информацию о состоянии физических или технических процессов (объектов), то для этой цели используют воздействие на датчик некоторого физического процесса, характеризующего исследуемый процесс (объект). Например, для получения информации о числе изготовленных изделий, которые двигаются на ленте конвейера, применяют датчик в виде фотоэлемента. На выходе фотоэлемента появляются импульсы тока, число которых подсчитывает счетчик, т. е. идет информация о числе готовых деталей.

Сигнал передает (иногда говорят — развертывает) сообщение во времени. Следовательно, он всегда — функция времени, даже если сообщение таковым не является. Передаваемые (далее часто, полезные) сигналы формируют путем изменения тех или иных параметров физического носителя в соответствии с передаваемым сообщением. Этот процесс изменения параметров носителя сообщений в радиотехнике и связи называют *модуляцией.*

Существует несколько физических характеристик, общих для любого сигнала. Физическая характеристика сигнала — это описание любым способом его свойств. Полезный электрический сигнал является объектом передачи, а техника его передачи — по существу техникой транспортировки сигналов по каналам и линиям связи. Поэтому с этой точки зрения целесообразно ввести параметры передаваемого сигнала, которые являются основными с точки зрения его передачи. Такими параметрами являются *длительность сигнала* $T_c$, его *ширина спектра* $F_c$ и *динамический диапазон* $D_c$.

Практически каждый электрический сигнал, рассматриваемый как временно́й процесс, имеет начало и конец. Поэтому длительность сигнала $T_c$ является естественным его параметром, определяющим интервал времени, в пределах которого данный сигнал существует.

Ширина спектра передаваемого сигнала $F_c$ дает представление о скорости изменения этого сигнала внутри интервала его существования. Спектр передаваемого сигнала в принципе может быть неограниченным. Однако для любого сигнала можно указать диапазон частот, в пределах которого сосредоточена его основная (до 90 %) энергия. Этим диапазоном и определяется ширина спектра полезного сигнала.

В радиоэлектронике и теории связи реальную ширину спектра передаваемого сигнала часто сознательно сужают. Это связано с тем, что радиоэлектронная аппаратура и линии связи систем передачи информации имеют ограниченную полосу пропускаемых частот. При этом сужают спектр исходя из допустимых искажений (*Distortion*) сигнала.

При *радиотелефонной связи* (как и при любой другой телефонии) элементы речи (звуки, слоги, слова и т. д.) произносятся фактически слитно и не имеют четких границ. Типичные для конкретного языка звуки речи называют *фонемами*. В русской речи различают 41 фонему. *Сигналы телефонирования* представляют собой последовательности речевых импульсов, отделенных друг от друга паузами. Импульсы соответствуют звукам речи, произносимым слитно, и весьма разнообразны по форме и амплитуде. Длительности отдельных речевых импульсов также отличаются друг от друга, но обычно они близки к 100...150 мс. Паузы между импульсами изменяются в значительно большем диапазоне: от нескольких *миллисекунд* (межслоговые паузы) до *нескольких минут* или даже *десятков минут* — паузы при выслушивании ответа собеседника. Частотный спектр речевого сигнала очень широк, однако экспериментально установлено, что для передачи речи с достаточно высоким качеством (узнаваемостью голоса абонента, воспроизведением тембра, удовлетворительной натуральностью и разборчивостью слогов 90 % и фраз 99 %) можно ограничиться полосой частот 0,3...3,4 кГц (300...3 400 Гц). Передача более широкого спектра речи в этом случае нецелесообразна, поскольку ведет к техническим усложнениям приемопередающей радиоаппаратуры и увеличению материальных и прочих затрат.

В *системах радиовещания* для высококачественной передачи музыки требуется полоса частот 30 ... 20 000 Гц. Это связано с тем, что звуковых колебаний с более высокими частотами человек практически не слышит. Однако передача такого широкого спектра без взаимных помех большого числа радиостанций в диапазонах километровых, гектометровых и декаметровых волн технически весьма затруднительно. Поэтому при радиовещании на этих волнах ограничиваются передачей спектральных составляющих в полосе 50 ... 4 500 Гц. На метровых и дециметровых волнах (устаревшее название — УКВ-диапазон) и, в частности, звуковое сопровождение телевизионного изображения передают более широкие спектры (30...10 000 Гц), чем достигается более высокое качество передачи музыки. Точно так же необходимая ширина спектра телевизионного сигнала определяется требуемой четкостью передаваемого изображения.

В *телевидении* для качественного воспроизведения изображения необходимо разложить его на большое количество элементов и предать информацию об яркости каждого элемента. Кроме того, для слитного восприятия глазом движу-

щегося изображения частота смены кадров на экране должна быть достаточно высокой. Если применяется телевизионный стандарт с 625 строками, то верхняя частота спектра телевизионного сигнала составляет 6 МГц. Спектр сигнала изображения всегда много шире спектра сигнала звукового сопровождения (сравним 6 МГц и 10 кГц), что в достаточной мере усложняет построение телевизионных систем по сравнению с системами звукового радиовещания. У нижней границы спектра телевизионного сигнала располагаются спектральные составляющие звукового сопровождения и сигналы синхронизации (*Synchronization*).

*Телеграфный сигнал* в общем случае представляется колебанием с дискретной модуляцией. Ширина спектра *телеграфного сигнала* зависит от скорости передачи (соответственно и от длительности самих импульсных посылок $T$) и обычно принимается равной $F \approx 1{,}5v$, где $v$ — *скорость телеграфирования*, или *скорость модуляции* (часто *скорость передачи импульсных посылок*, или *техническая скорость*) в Бодах (*baud*; Бод). Один Бод (введен в телеграфию Жан Бодо; *Jean Baudot*; 1845–1903; изобретатель первого телеграфного аппарата) — это скорость, при которой за 1 с передается одна посылка. Тогда при телетайпной передаче сообщений и скорости передачи $v = 50$ Бод ширина спектра телеграфного сигнала составляет $F \approx 75$ Гц. Если длительность посылки $T$ выражена в секундах, то техническая скорость (скорость модуляции) $v = 1/T$ [Бод]. Ограничение скорости модуляции (скорости телеграфирования) обусловлено техническими характеристиками любой реальной системы передачи информации.

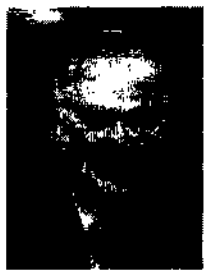
По скорости передачи телеграфного сигнала системы передачи делят на:

- низкоскоростные (НС) — до 200 Бод;
- среднескоростные (СС) — 600 … 1 200 Бод;
- высокоскоростные (ВС) — 2 400 … 96 000 Бод.

Частота следования двоичных (*binary*) посылок (т. е. «1» и «0») называется *тактовой* (*clock*) *частотой* $F_т$. Численно $F_т$ соответствует скорости передачи информации в Бодах.

При передаче цифрового сообщения в виде комбинации символов «1» и «0», т. е. двоичным кодом, элементарную посылку (двоичный символ, двоичную цифру «1» или «0») называют *битом* (от англ. *binary digit* — «*bit*» — двоичная единица; слово предложено американским компьютерщиком Джон Тьюки; *John Tukey*; 1915–2000).

Итак, бит — один двоичный разряд — символ, принимающий лишь одно из двух значений: «1» или «0» (следует иметь в виду, что слово «бит» в теории информации имеет два различных назначения; одно используется в качестве синонима двоичного символа, а второе — обозначает единицу количества информации, необходимое для различения двух равновероятных сообщений; например «орел» – «решка», и т. д.). Так, представление символов в виде 101 есть 3-битовое число. При этом количество передаваемой за секунду информации (символьная скорость — *symbol rate*) измеряется в битах в секунду (бит/с = BPS — *Bits Per Second*).



Д. Тьюки



Ж. Бодо



Г. Лейбниц

Для избежания неясности (см. гл. 9) заметим, что при синхронной передаче цифровых данных обозначение «бит/с» аналогично обозначению «Бод», т. е. скорость 1 бит/с = 1 Бод), поэтому при совпадении технической скорости системы и требуемой символьной скорости передачи информации, скорость можно определять как в бит/с, так и в Бодах.

Первое опубликованное представление (1670 г.) двоичной системы счисления (*система счисления* — правило записи чисел с помощью заданного набора специальных знаков — цифр) принадлежит испанскому священнику Х.С. Лобковицу (*J.C. Lobkowitz*; 1606–1682). Всеобщее же внимание к этой системе привлекла статья немецкого математика Г.В. Лейбница (*G.W. Leibniz*; 1646–1716), опубликованная в 1703 г. В этой статье пояснялись двоичные операции сложения, вычитания, умножения и деления. Со временем двоичная система счисления (основание системы $R = 2$) становится хорошо известной.

В двоичной системе целое число представляют как сумму степеней числа 2 с соответствующими коэффициентами от 0 до 1. Работа большинства компьютеров основана именно на этой системе чисел. С помощью двоичной системы кодирования можно зафиксировать любые данные. Это легко понять, если вспомнить принцип кодирования и передачи информации с помощью азбуки (кода) Морзе (*Morse code*).

Восемь бит называется *байтом* (от англ. *bite* — слог; ввел американский компьютерщик Роберт Бемер — *Robert Bemer*; 1920–2005). При использовании двоичного представления кодовая комбинация может выражать целое число, равное соответствующему уровню непрерывного сигнала в момент его дискретного отсчета. Байт (байт = $2^3$ = 8 бит) служит в качестве единицы представления информации: букв, слогов и специальных символов (занимающих обычно все 8 бит) или десятичных цифр (по две цифры в 1 байте). С помощью 1 байта можно получить 256 разных двоичных кодовых комбинаций и отобразить с их помощью 256 различных символов. Кодирование заключается в том, что каждому символу ставится в соответствие уникальный десятичный код от 0 до 255 или соответствующий ему двоичный код от 00000000 до 11111111.

В импульсных линиях систем радиотелеметрии, радиотелеуправления, радиолокации, радионавигации, управления воздушным движением, комплексов противовоздушной и противоракетной обороны и дальнего обнаружения для достижения необходимой разрешающей способности и точности измерений параметров используют стабильные кратковременные импульсы напряжения длительностью 0,1…1 мкс. Ширина спектра такого сигнала составляет 1…10 МГц. Следовательно, в зависимости от требований к системе передачи информации первичные (см. далее) сигналы могут иметь ширину спектра от сотен герц до нескольких мегагерц.

Обычно спектр модулированного сигнала шире спектра сигнала, отражаемого передаваемое сообщение, и зависит от вида модуляции. Поэтому в радиотехнике и теории сигналов используют такой параметр, как *база сигнала*

$$B_c = 2F_cT_c. \quad (1.1)$$

В теории информации вводят более общую характеристику — *объем сигнала*

$$V_c = T_cF_cD_c. \quad (1.2)$$

Чем выше частота излучаемых сигналов, тем большего объема и с более высокой скоростью можно передавать сообщения. Объем сигнала дает более общее представление о возможностях данного множества сигналов как переносчиков сообщений. Чем больше объем сигнала, тем больше информации можно «заложить» в него и тем труднее передать такой сигнал с требуемым качеством.

Динамический диапазон (*dynamic range*) сигнала $D_c$ — отношение наибольшей мгновенной мощности передаваемого сигнала к той наименьшей мощности, которую необходимо отличать от нулевой мощности при заданном качестве передачи информации. Обычно его удобнее выражать в децибелах. Напомним, что в случае сравнения мощностей $P_2$ и $P_1$:

$$1 \text{ дБ} = 10 \lg (P_2/P_1),$$

при этом

$$P_2/P_1 = 10^{1/10} = 1{,}259.$$

Примечание. В последнее время специалисты по системам связи находят удобным измерять уровень мощности непрерывных сигналов по отношению к некоторому заданному значению абсолютной мощности. В этом случае говорят об уровне абсолютной мощности в ваттах, используя в то же время преимущества логарифмического масштаба. Обычно используется опорный уровень в 1 мВт.

Например, если $P_1$ представляет собой опорный уровень в 1 мВт, то

$$\text{абсолютная мощность } P_2 = 10 \lg (P_2/P_1) = 10 \lg (P_2 \text{ в Вт}/1 \text{ мВт}) \text{ дБм}.$$

Единица измерения дБм в последней формуле читается как «децибел относительно одного милливатта». Значит, если для непрерывного сигнала известно, что его мощность равна 3 дБм, то абсолютная мощность этого сигнала в 2 раза превышает 1 мВт, или равна 2 мВт. Аналогично, сигнал уровня –10 дБм имеет абсолютную мощность 0,1 мВт.

При сравнении напряжений (иногда токов) $U_2$ и $U_1$:

$$1 \text{ дБ} = 20 \lg (U_2/U_1).$$

В этом случае

$$U_2/U_1 = 10^{1/20} = 1{,}122.$$

Динамический диапазон речи телевизионного диктора составляет 25…35 дБ, художественного чтения достигает 50 дБ, музыкальных и хоровых ансамблей до 55 дБ; симфонического оркестра 65…90 дБ. Во избежание перегрузок передающего канала в вещании динамический диапазон часто сокращают до 35…45 дБ.

Размерные характеристики сигнала (длительность, объем, динамический диапазон, ширина о спектра) не всегда достаточны для его описания. Например, для анализа помехоустойчивости систем связи надо исследовать не только временну́ю и спектральную структуры сигнала, но и вводить более тонкие его характеристики — корреляцию, энтропию, количество информации и т. д.

При проектировании и создании систем передачи информации обычно оказывается, что спектр передаваемого сигнала сосредоточен не на тех частотах, которые эффективно пропускает имеющийся канал связи. Часто необходимо в одном канале передавать несколько независимых сигналов одновременно.

Для передачи информации на большие расстояния используют электромагнитные волны (акустические волны пригодны лишь для ограниченных расстояний). При этом передачу можно осуществлять по медным проводам, оптоволоконному кабелю или непосредственно, по схеме передатчик–приемник. В последнем случае используются антенны. Для того чтобы антенна была эффективна, ее размеры должны быть сравнимы с длиной передаваемой волны. Однако электрические сигналы, отражающие передаваемые сообщения, как правило, маломощны и низкочастотны. Из курса физики известно, что электрические сигналы с низкими частотами не могут эффективно излучаться в свободное пространство, поскольку размеры антенн должны быть соизмеримы с длинами излучаемых радиоволн. Передавать их непосредственно можно только по проводным или кабельным линиям (телефонная, телеграфная связь и т. д.).

Теперь ответим на вопрос: зачем для передачи низкочастотных сигналов нужна несущая? Дело в том, что передачу электромагнитного колебания на расстояние выполняют с помощью антенн, размер которых зависит от длины волны $\lambda$. Для мобильных телефонов размер антенны обычно равен $\lambda/4$, а длина волны

$$\lambda = c/f, \tag{1.3}$$

где $c \approx 3 \cdot 10^8$ м/с — скорость света в свободном пространстве; $f$ — циклическая частота, Гц (герц — частота, при которой происходит одно колебание в секунду). Для частоты, определяемой в килогерцах, мегагерцах, гигагерцах, соответственно получают следующие соотношения:

$$f\,[\text{кГц}] = 300/\lambda\,[\text{км}]; \quad f\,[\text{МГц}] = 300/\lambda\,[\text{м}]; \quad f\,[\text{ГГц}] = 30/\lambda\,[\text{см}].$$

Рассмотрим передачу низкочастотного сигнала (например, с средней частотой $f = 1\,500$ Гц), поступающего в антенну без использования несущей. Какая антенна будет нужна для мобильного телефона при его размерах $l = \lambda/4$. Получаем для сигнала 1 500 Гц $l = \lambda/4 = c/(4f) = 3 \cdot 10^8/(4 \cdot 1\,500)$ м $= 50\,000$ м $= 50$ км. Итак, для передачи сигнала с частотой 1 500 Гц без несущей требуется антенна размером 50 км. При этом, если низкочастотный сигнал передается с помощью несущей, например 1500 МГц, размер антенны составит порядка 5 см. Именно поэтому для передачи используются частоты, начиная с многих сотен килогерц и выше (длина волн сотни метров и меньше) — «быстрые» радиоволны.

При использовании оптических носителей передача сигнала непосредственно по лучу лазера ограничена расстояниями 100…3 000 м и становится неус-

тойчивой при наличии осадков даже для инфракрасных длин волн. Между тем человек воспринимает акустические колебания в диапазоне 20...12 000 Гц и для целей передачи звука (например, телефония) требуется именно этот диапазон частот. Динамический диапазон частот (отношение максимальной частоты к минимальной) в этом случае равен 600, а для высококачественного воспроизведения звука он в два раза шире. При решении проблемы используют преобразование частот и различные методы модуляции, что позволяет сделать компактную антенну и упростить частотное выделение сигнала. Перечисленные причины приводят к необходимости такой трансформации исходного сигнала, чтобы требования, предъявляемые к занимаемой им полосе частот, были выполнены, а сам передаваемый сигнал можно было с достаточной верностью восстановить на приемном конце. Чтобы уменьшить воздействие помех, необходимо представить сигнал в помехоустойчивой форме, подвергнув его дополнительным преобразованиям. Такими преобразованиями являются кодирование и модуляция.

Суть *модуляции* (от англ. — *modulation*; от лат. *modulatio* — размеренность) заключается в следующем. Для передачи информации формируют электрические сигналы (*переносчики сообщений*), которыми обычно являются хорошо излучающиеся и распространяющиеся (с низким коэффициентом затухания) в свободном пространстве мощные высокочастотные гармонические электромагнитные колебания — *несущие колебания*, или просто *несущие* (*carrier*). Несущие колебания не содержат информации, а только ее переносят. Передаваемая по каналам связи информация путем модуляции закладывается в один или ряд параметров несущего колебания. Исходный сигнал называют *модулирующим* (*modulating signal*), а результирующее колебание с изменяющимися во времени параметрами — *модулированным сигналом* (*modulated signal*). Обратный процесс — выделение модулирующего сигнала из модулированного колебания — называется *демодуляцией* (*demodulation*) или *детектированием* (*detection*).

### Диапазоны электромагнитных колебаний и их использование в радиотехнике и связи

В современной радиотехнике используют практически весь спектр электромагнитных колебаний, расположенных в частотном диапазоне (*frequency range*) 10 ... $10^{13}$ Гц. Электромагнитные колебания с такими частотами называют *радиоволнами* (*волнами*). Учитывая некоторые особенности отдельных участков частотного спектра, его условно можно разделить на диапазоны, внутри которых волны обладают некими общими свойствами, отличающими их от характеристик соседних и тем более далеких диапазонов. Разумеется, такое деление условно и резких границ между соседними диапазонами нет: особенности одного диапазона плавно трансформируются в особенности соседнего.

Деление радиоволн на диапазоны установлено Международным регламентом радиосвязи. Международная классификация диапазонов волн и соответствующих им диапазонов частот приведена в табл. 1.1. Эта классификация основана на особенностях распространения радиоволн в канале связи (свободном пространстве, ионосфере, оптическом волокне, морской воде, земной коре и т. д.).

Таблица 1.1. **Классификация диапазонов радиоволн (частот)**

| Наименование | Диапазон | | Область применения |
|---|---|---|---|
| | волн | частот | |
| Декамегаметровые (крайне-низкие частоты – КНЧ; ELF) | $10^5...10^4$ км | 3 ... 30 Гц | Подводная и служебная связь, звукозапись, звуковоспроизведение |
| Мегаметровые (сверхнизкие частоты – СНЧ; SLF) | $10^4...10^3$ км | 30 ... 300 Гц | Подводная и служебная связь, звукозапись, звуковоспроизведение |
| Гектокилометровые (инфра-низкие частоты – ИНЧ; ULF) | $10^3...10^2$ км | 300...3 000 Гц | Подводная и служебная связь, звукозапись, звуковоспроизведение |
| Мириаметровые (очень низкие частоты – ОНЧ; VLF) | 100...10 км | 3 ... 30 кГц | Дальняя радионавигация, подводная, подземная и служебная связь |
| Километровые (низкие частоты – НЧ; LF) | 10...1 км | 30 ... 300 кГц | Радиовещание, радионавигация, радиомаяки |
| Гектометровые (средние частоты – СЧ; MF) | 1000...100 м | 300...3000 кГц | Радиовещание, радионавигация, радиомаяки |
| Декаметровые (высокие частоты – ВЧ; HF) | 100...10 м | 3 ... 30 МГц | Радиовещание, подвижная связь, радиотелеграфия, загоризонтная связь (локация) |
| Метровые (очень высокие частоты – ОВЧ; VHF) | 10 ... 1 м | 30 ... 300 МГц | Радио- и телевещание, подвижная и самолетная связь, радиолокация, радиоастрономия |
| Дециметровые (ультравысокие частоты – УВЧ; UHF); L-, S-диапазоны | 100...10 см | 300 ... 3 000 МГц | Телевещание, локация, радиовысотомеры, сотовая, радиорелейная и космическая радиосвязь |
| Сантиметровые (сверхвысокие частоты – СВЧ; SHF); C-, X-, K-диапазоны | 10 ... 1 см | 3 ... 30 ГГц | Радиолокация, радионавигация, космическая связь, телевещание, радиоастрономия |
| Миллиметровые (крайне-высокие частоты – КВЧ; EHF) | 10 ... 1 мм | 30 ... 300 ГГц | Радиолокация, космическая связь, радиоастрономия |
| Децимиллиметровые и субмиллиметровые (гипер-высокие частоты – ГВЧ) | 1 ... 0,1 мм | 300 ... 3 000 ГГц | Инфракрасная (ИК) локация, космическая связь |

Из табл. 1.1 видно, что Международным регламентом радиосвязи не установлено обозначение для частот ниже 3 Гц.

Для обеспечения устойчивой и надежной передачи информации по радиоканалу важна длина волны несущего колебания. На выбор диапазона волн для конкретной системы передачи информации влияет ряд специфических факторов, связанных с особенностью формирования, излучения и распространения волн (прежде всего несущих колебаний), характером и параметрами имеющихся в заданном диапазоне шумов и помех, параметрами сообщения, характеристиками и габаритными размерами передающих и приемных антенн.

## 1.2. Элементы теории распространения радиоволн

Для радиотехники и связи громадное значение имеет способность некоторых радиотехнических устройств эффективно создавать и излучать электромагнитные волны («возбуждать» канал связи). Для эффективного излучения волны необходимо, чтобы минимальный размер антенны (например, провода) был не менее половины длины волны $\lambda$. Очевидно, что по мере укорочения длины волны $\lambda$ выполнение этого условия становится все более легко осуществимым.

Механизм возникновения и распространения высокочастотных электромагнитных колебаний (электромагнитных волн) весьма сложен. Как известно из курса физики, постоянные электрические и магнитные поля не могут существовать без породивших их источников, т. е. зарядов или токов. Возбужденное же каким-либо источником переменного тока электромагнитное поле может существовать само по себе в отрыве от источника и, после излучения в виде электромагнитных волн в свободное пространство, будет распространяться в нем по некоторому пути практически со скоростью света.

Простейшим *излучателем* электромагнитных колебаний может служить короткий отрезок электрического проводника, физическая длина которого $L$ много меньше длины излучаемой волны $\lambda$. На практике такой излучатель электромагнитных колебаний был реализован с помощью *диполя Герца*, представляющего собой два небольших металлических шара, соединенных отрезком проводника, в разрез которого включен источник переменного тока (рис. 1.1, *а*). Благодаря наличию шаров на концах отрезка проводника длина диполя $L$ может быть существенно уменьшена, поскольку диполь — это электромагнитный резонатор (контур), а металлические шары эквивалентны емкости контура, которая, снижая значение индуктивности, уменьшает длину диполя. В этом случае модуль линейной плотности электрического тока $I$ будет распределен по длине излучателя равномерно, т. е. $I_x = I_0$ (рис. 1.1, *б*). Ширина разреза проводника $\Delta$ должна быть намного меньше длины диполя $L$.

Упрощенно возникновение электромагнитного поля можно трактовать следующим образом. Согласно законам электромагнитной индукции, в пространстве, окружающем диполь с протекающим в нем переменным током, будет возникать переменное магнитное поле. Изменяющееся магнитное поле создает, в свою очередь, в окружающем пространстве переменное электрическое поле. Совместно изменяющиеся электрическое и магнитное поля порождают в соседних точках пространства соответственно магнитные и электрические поля, поэтому эти взаимосвязанные явления создают электромагнитную волну.

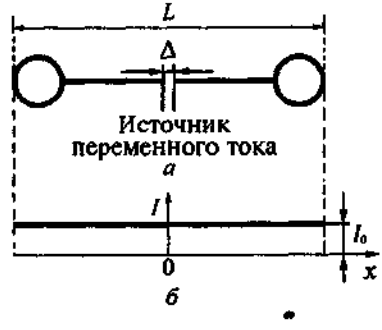



**Рис. 1.1.** Элементарный излучатель:
*а* — диполь Герца;
*б* — распределение тока по излучателю

*Электромагнитным излучением* (ЭМИ) называют процесс возбуждения распространяющихся в пространстве электромагнитных волн в пространстве, окружающем источник колебаний. Эти сложные физические явления теоретически обосновал Д. Максвелл и описал аналитически в виде двух фундаментальных положений: *всякое изменение во времени магнитного поля вызывает появление вихревого (т. е. замкнутого) электрического поля; всякое изменение во времени электрического поля приводит к появлению вихревого магнитного поля.* Основные положения теории электромагнитного поля Д. Максвелл выразил четырьмя связанными между собой уравнениями электродинамики. Они подтверждают существование излученных источником переменного тока электромагнитных колебаний, которые распространяются в свободном пространстве прямолинейно со скоростью света.

При своем распространении в пространстве электромагнитные волны переносят энергию, выделяемую в месте излучения источником переменного тока. Соответствующая этой энергии мощность $P_и$ (энергия, переносимая полем за единицу времени) называется *мощностью излучения.* Предположим теоретически, что источник излучения является точечным. В этом случае электромагнитная волна излучается источником равномерно во всех направлениях, создавая воображаемую поверхность сферы, и во всех точках пространства, равноудаленных от источника, т. е. лежащих на поверхности воображаемой сферы, фаза колебаний одинакова. *Плотность потока мощности,* или *интенсивность излучения* (мощность, проходящая через единицу площади поверхности сферы, Вт/м$^2$), равна:

$$П = \frac{P_и}{S_с} = \frac{P_и}{4\pi r^2}, \qquad (1.4)$$

где $r$ — радиус сферы, м; $S_с$ — поверхность сферы, м$^2$.

Из курса физики известно, что плотность потока мощности электромагнитного поля пропорциональна квадрату напряженности электрического поля $E_и$, создаваемого точечным источником излучения. В теории излучения и распространения электромагнитных колебаний доказано, что эта зависимость определяется формулой

$$П = \frac{E_и^2}{120\pi}. \qquad (1.5)$$

Приравняв выражения (1.4) и (1.5), находим напряженность электрического поля, создаваемого точечным источником излучения

$$E_и = \frac{\sqrt{30P_и}}{r}. \qquad (1.6)$$

**Пример 1.1.** Определить необходимую мощность излучения точечного источника тока, если плотность потока мощности электромагнитного поля на расстоянии 1 000 м от него должна составлять не менее 0,002 Вт/м$^2$.

Р е ш е н и е. Из соотношения (1.4) находим

$$P_и = 4 \cdot 3{,}14 \cdot (1\,000)^2 \cdot 0{,}002 = 25\,120 \text{ Вт} = 25{,}12 \text{ кВт}.$$

**Пример 1.2.** Рассчитать, на каком расстоянии от точечного излучателя электромагнитного поля мощностью 270 Вт будет создано электрическое поле напряженностью 0,01 В/м.

Р е ш е н и е. Воспользовавшись формулой (1.6), находим

$$r = \frac{\sqrt{30P_и}}{E_и} = \frac{\sqrt{30 \cdot 270}}{0{,}01} = 9\,000 \text{ м} = 9 \text{ км}.$$

Итак, напряженность электрического поля электромагнитной волны убывает обратно пропорционально расстоянию от источника излучения (напомним, что напряженность электростатического поля убывает обратно пропорционально квадрату расстояния). Именно такая линейная зависимость напряженности и позволяет передавать информацию на очень большие расстояния с помощью электромагнитных колебаний при реально достижимой мощности излучения системы передачи информации.

Принятое ранее предположение о равномерном излучении энергии во всех направлениях и прямолинейном распространении электромагнитных волн весьма условно. В большинстве реальных случаев передающие и приемные пункты радиотехнических систем передачи информации располагаются либо на земной поверхности, либо в непосредственной близости от нее. Исключением являются системы космической связи и связи Земля – Космос. Поэтому при исследовании условий распространения радиоволн следует учитывать влияние на них Земли и окружающей ее атмосферы.

С физической точки зрения земная поверхность и окружающее ее пространство представляют собой среду с различными электрическими параметрами (электропроводностью, диэлектрической проницаемостью и т. д.), в которой распространяются радиоволны, испытывая преломление, отражение, затухание (*attenuation*) и частичное поглощение.

### Поляризация волн

Весьма важное значение для распространения и приема излучаемых электромагнитных волн имеет их поляризация (франц. — *polarisation*; от греч. *pólos* — ось, полюс). Как физическое явление *поляризация* — ориентация в пространстве электрического и магнитного векторов электромагнитной волны. Обычно поляризацию принято характеризовать ориентацией вектора напряженности электрического поля. Плоскость, в которой лежит этот вектор, называют *плоскостью поляризации*. Различают линейную (вертикальную и горизонтальную), эллиптическую и круговую поляризацию. В пределах поля тяжести Земли в кана-

лах радиотехнических систем обычно используют вертикальную (вектор поля параллелен силе тяжести) или горизонтальную линейную поляризацию.

Можно достаточно условно объяснить процесс поляризации радиоволны по аналогии с механическими колебаниями натянутой и возбужденной музыкальной струны. Натянутая струна соответствующим возбуждением (отводом ее, например, пальцами) на одном конце приводится в колебательное состояние, представляющим собой механическую волну. Созданная волна распространяется вдоль струны и может быть зарегистрирована на другом ее конце. Такую механическую волну можно упрощенно представить как модель радиоволны (или световой волны), которая движется по пространству от источника электромагнитных колебаний к приемнику (регистратору). Натянутая струна может быть возбуждена по-разному: отклонение струны можно произвести или в вертикальной, или в горизонтальной плоскости. Когда речь идет о радиоволне, излучаемой антенной передающего устройства, то в первом случае можно говорить об аналоге *вертикальной*, а во втором случае — об аналоге *горизонтальной поляризации* волны.

Для обычного радиоприемника (стоящего после антенны) на другом конце линии связи это тонкое различие в свойствах потока электромагнитных колебаний не существенно. Приемник не реагирует на плоскость поляризации радиоволны, а регистрирует только ее мощность (в модели это степень отклонения струны); он не различает горизонтальную и вертикальную поляризацию. Однако имеются элементы, которые реагируют на поляризацию электромагнитных колебаний. Их называют *поляризационными фильтрами*. В случае со струной таким упрощенным аналогом фильтра может служить горизонтальная (для горизонтально колеблющейся струны) или вертикальная (для вертикально колеблющейся струны) прорезь в картоне, поставленном перед человеческим глазом перпендикулярно направлению распространения механической волны. Поляризационный фильтр, поставленный в заданном положении относительно направления распространения электромагнитной волны, становится радиопрозрачным лишь для определенного вида поляризации.

Если горизонтальная и вертикальная компоненты возбуждения волны появляются в определенной временно́й последовательности, то это приводит к *круговой поляризации* электромагнитных (или световых) волн. Излучение волн с круговой поляризацией может быть представлено в виде суммы двух волн с линейными поляризациями, ориентированными перпендикулярно друг к другу.

Использование волн с различной поляризацией позволяет передавать сигналы на близких частотах или одной частоте и эффективно разделять их при приеме. Управляя поляризатором дистанционно, можно выбирать сигналы, передаваемые с определенной поляризацией. В основном применяют поляризаторы двух типов: электромагнитные (ферритовые) и механические. Их электрические характеристики приблизительно одинаковы. Отсутствие в электромагнитном поляризаторе движущихся частей предполагает его большую надежность. В то же время механические поляризаторы обладают несколько более низким коэффициентом шума (см. гл. 9).

При распространении радиоволн вдоль земной поверхности возможно также их отражение от Земли. Аналогично тому, как световые волны отражаются от окружающих предметов, так и излученные радиоволны отражаются от поверхности Земли. Радиоволны, распространяющиеся в непосредственной близости от поверхности Земли, в радиотехнике называют *земными* или *поверхностными волнами* (*поверхностными лучами*). Существенное влияние на распространение радиоволн различных диапазонов оказывает также земная атмосфера (от греч. *atmos* — пар и *spháira* — шар). Окружающую Землю атмосферу принято делить на три характерных слоя: тропосферу, стратосферу и ионосферу.

*Тропосфера* (от греч. *tropos* — поворот) представляет собой нижний слой атмосферы, простирающийся до высот 10...20 км. Тропосфера неоднородна по своим электрическим свойствам, которые определяются атмосферным давлением, температурой и влажностью и меняются при изменении метеоусловий. Кроме того, воздушные течения интенсивно перемешивают газы тропосферы, что приводит к созданию локальных неоднородностей. Все это существенно влияет на распространение радиоволн в тропосфере.

Слой атмосферы, лежащий выше тропосферы и располагающийся на высотах до 50 км, называют *стратосферой* (от лат. *stratum* — слой). Плотность газов в ней значительно меньше, чем в тропосфере. С точки зрения электрических свойств стратосфера является практически однородной средой, и волны распространяются в ней прямолинейно со скоростью света и без существенных потерь.

Над стратосферой (высота до 20 000 км) располагается *ионосфера* (от *греч. ión* — идущий) — верхние, ионизированные слои атмосферы, которые образуются под воздействием космического излучения и ультрафиолетовых лучей Солнца. В результате ионизации молекул воздуха возникают положительные ионы газа и свободные электроны. Чем больше концентрация свободных электронов, тем сильнее влияют они на распространение радиоволн.

Число свободных электронов, содержащихся в 1 м$^3$ воздуха атмосферы, называют *концентрацией* и обозначают $N_э$, эл/м$^3$. Концентрация электронов в различных слоях ионосферы меняется по высоте. На малых высотах от поверхности Земли она невелика, так как недостаточна энергия ионизации. На больших высотах концентрация свободных электронов невелика вследствие малой плотности газа в атмосфере. На высотах 300...400 км концентрация свободных электронов в ионосфере максимальна. Изменение плотности атмосферы с увеличением расстояния от Земли и сложная зависимость ее температуры от высоты приводят к тому, что в результате ионизации в ионосфере образуются четыре ярко выраженных слоя: $D$, $E$, $F_1$ и $F_2$.

Ионосферный *слой D* расположен на высотах 60...90 км над поверхностью Земли. Слой представляет собой нерегулярное образование ионосферы и практически существует только в дневные часы, когда велика интенсивность солнечного ионизирующего излучения. На высотах 100...120 км от поверхности Земли находится ионосферный *слой E* (слой Кеннелли–Хевисайда). В зависимости от времени года и суток изменяется лишь концентрация свободных электронов в этом слое. Днем слой $E$ располагается несколько ниже, а ночью — значи-

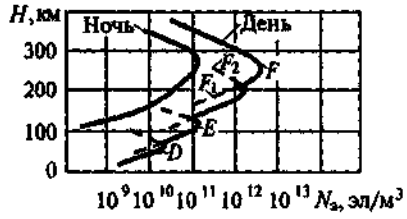



**Рис. 1.2.** Суточная концентрация электронов в слоях ионосферы

тельно выше, что связано с изменениями уровня солнечного потока. *Слои* $F_1$ и $F_2$ ионосферы занимают области на высотах от 120 до 450 км от поверхности Земли. Обычно их рассматривают как один слой $F$, имеющий наибольшую концентрацию свободных электронов в ионосфере.

На рис. 1.2 представлена примерная зависимость распределения концентрации свободных электронов от высоты в характерных слоях ионосферы для конкретного случая.

Характеристики всех слоев ионосферы сильно меняются в зависимости от времени года и времени суток и циклов солнечной активности. Летом ионизация газов интенсивнее, чем зимой; в дневное время она больше, чем ночью. Наличие в атмосфере слоя ионосферы существенно влияет на распространение некоторых типов радиоволн.

Как и световые волны, радиоволны могут преломляться при переходе из одной среды в другую. Исследования ионосферы показали, что ее коэффициент преломления для радиоволн различной частоты уменьшается с ростом высоты. В результате такого влияния направление лучей в ионосфере постепенно искривляется в сторону стратосферы. В некоторых условиях преломление в ионосфере столь велико, что излученные радиоволны могут полностью от нее отразиться и вернуться к Земле. Степень преломления лучей электромагнитных колебаний в слоях ионосферы зависит как от угла падения радиоволны $\gamma$ с поверхности Земли на ионизированный слой, так и от ее частоты $f$. Возможные траектории распространения радиоволны (луча) при различных углах падения показаны на рис. 1.3.

Исследования распространения радиоволн в различных условиях выявили определенную закономерность между углом падения $\gamma$, концентрацией свободных электронов в ионосфере $N_э$ и частотой радиоволны $f$.

Угол падения, при котором радиоволна не проходит через ионосферу и распространяется вдоль нее (луч *1*), называется *критическим*:

$$\gamma_{кр} = \arcsin\sqrt{1 - 80{,}8N_э/f^2}. \qquad (1.7)$$

Если углы падения меньше критического, то преломление лучей невелико, и радиоволны проникают через ионосферу в космос (луч *2*). При углах падения, больших критического, преломление настолько велико, что радиоволны полностью отразятся и вернутся на Землю (луч *3*).

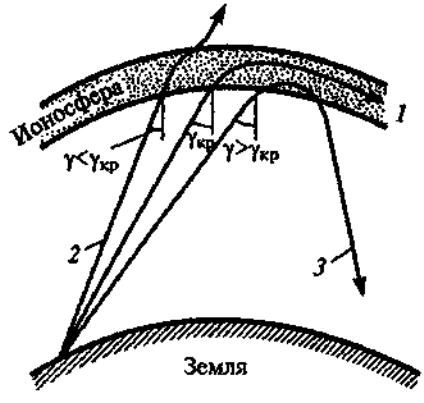



**Рис. 1.3.** Траектории распространения волн при разных углах падения

Из соотношения (1.7) следует, что чем выше частота излучаемого колебания (короче длина волны), тем меньше степень преломления радиоволн при прочих равных условиях. Очень короткие волны вообще не преломляются в ионосфере и, пройдя через нее, излучаются в космическое пространство. Такие волны применяют в системах космической и спутниковой радиосвязи. Воспользуемся формулой (1.7) и определим наивысшую частоту излучения электромагнитной волны [МГц], при которой луч, направленный с Земли вертикально ($\gamma_{кр} = 0$), еще может возвратиться на нее (отразиться от ионосферы),

$$f_{кр} = \sqrt{80{,}8N_э} \approx 9\sqrt{N_э}\,. \qquad (1.8)$$

Частота $f_{кр}$ называется *критической частотой излучения*, соответствует максимальной концентрации электронов в слоях ионосферы и достигает 2 ...10 МГц. Если радиоволна послана с поверхности Земли под наклоном, то отражение от ионосферы существеннее. Оказывается, что частота колебаний электромагнитной волны, посланной с земной поверхности по касательной к горизонту, в 3...5 раз выше критической частоты. Такую частоту называют *максимально применимой частотой* (МПЧ).

Радиоволна, излучаемая с Земли с частотой $f >$ МПЧ, теоретически ни при каких условиях не отразится от слоев ионосферы и проникнет в космическое пространство. В радиотехнике радиоволны, распространяющиеся путем отражения (а также преломления) от ионосферы, называют *пространственными волнами* (*пространственными лучами*). Траектории распространения радиоволн с различными частотами излучения показаны на рис. 1.4.

Совместное направляющее действие земной поверхности и ионосферы наиболее заметно для длинноволновой части радиодиапазона. Эти радиоволны могут распространяться как за счет дифракции (от лат. *diffractus* — разломанный), т. е. огибания, так и путем отражения от ионосферы. Причем некоторые типы радиоволн могут распространяться в результате многократных попеременных отражений от ионосферы и Земли, вследствие чего в точку приема могут прийти несколько лучей.

Интересно явление перекрестной модуляции радиоволн в ионосфере. Перекрестная модуляция (кроссмодуляция — *cross-modulation* — взаимная модуляция двух или большего числа колебаний) — явление, наблюдавшееся *при распространении радиоволн в ионосфере*, состоящее в том, что сильное электрическое поле мощной радиоволны, изменяя («возмущая») скорость движения электронов ионосферы с частотой своей амплитудной

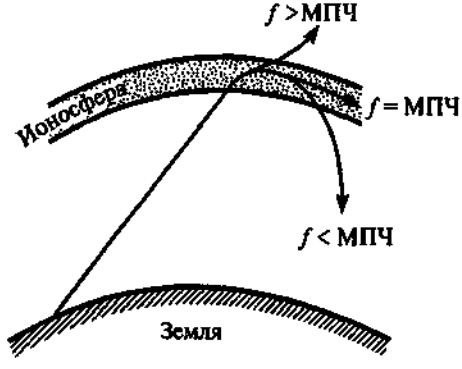



**Рис. 1.4.** Траектории распространения волн при разных частотах излучения

модуляции, вызывает *амплитудную модуляцию* других радиоволн, проходящих через эту возмущенную область ионосферы. Это явление принято называть *Люксембург-Горьковским эффектом*. Впервые данный эффект наблюдали в 1933 г. в г. Эйндховене (Нидерланды), где при обычном приеме швейцарской радиостанции прослушивали работу лежащей на пути мощной станции «Люксембург». Аналогичное явление наблюдали в г. Горьком (теперь Нижний Новгород), где при приеме радиостанций, расположенных на западе от Москвы, прослушивали мощные московские станции.

Степень практического освоения различных диапазонов волн в разных областях радиотехники и системах связи далеко не одинакова и зависит от множества факторов. Рассмотрим распространение наиболее используемых в связи и вещании диапазонов волн.

**Декамегаметровые, мегаметровые, гектокилометровые и мириаметровые (их длинноволновая часть) волны** обладают важнейшим свойством — они в меньшей степени поглощаются морской (и речной) водой и поэтому используются в системах подводной связи (в частности, для связи с подлодками).

Ослабление волн [дБ/м] этих диапазонов в морской воде определяется как

$$\alpha(f) = 0{,}0345\sqrt{f}, \qquad (1.9)$$

где частота $f$ выражена в герцах.

Для связи с глубоко погруженными лодками необходимо использовать частоты, не превышающие нескольких десятков герц. Например, для частоты связи $f$ = 100 Гц затухание на глубине 300 м, согласно (1.9), составляет более 100 дБ. Данные системы применяют лишь для предупреждения подводных лодок о необходимости всплытия и принятия более подробной информации через систему космической связи. Малая информативность связи на этих волнах, сложность и дороговизна приемопередающей аппаратуры компенсируется глубиной их проникновения в морскую воду. Следует отметить, что естественные электрические помехи в этих диапазонах достаточно стабильны и предсказуемы.

На мириаметровых волнах многие годы ведут интенсивные исследования в области подземной радиосвязи, поскольку уровень шумов на подземной трассе значительно ниже, чем на наземной (10 дБ против 130 дБ атмосферного шума). Для примера отметим, что подземная трасса длиной 1 700 км обеспечивает такое же отношение сигнал/шум, как и наземная трасса длиной 1 000 км. Волны этого диапазона (кроме радиосвязи) используют при создании аппаратуры радиоразведки полезных ископаемых, геологических исследованиях строения Земли и распространения радиоволн по естественным подводным линиям и др.

**Километровые волны** распространяются над поверхностью Земли по двум траекториям — поверхностными и пространственными лучами. Поверхностные волны в этом частотном диапазоне затухают сравнительно медленно (слабо поглощаются Землей) и распространяются вследствие дифракции на расстояния до 2 500 км. На бóльшие расстояния километровые волны распространяются за счет отражения от нижних слоев ионосферы; суша и вода достаточно хорошо их

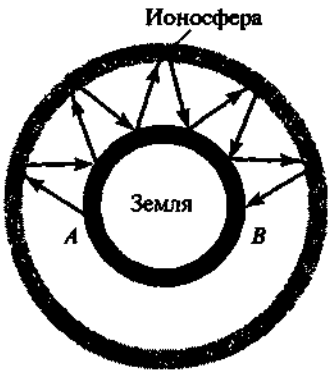



**Рис. 1.5.** Скачковое распространение волн пространственными лучами

отражают. В результате происходят многократные попеременные отражения, и километровые волны распространяются в своеобразном сферическом волноводе (двойной трубе), образованном нижними слоями ионосферы и полупроводящей поверхностью Земли (рис. 1.5). Этот способ распространения радиоволн из передающего пункта *A* в приемный пункт *B* называют *скачковым*. Связь на километровых волнах весьма устойчива и слабо зависит от времени года и суток.

**Гектометровые волны** так же, как и километровые электромагнитные колебания, распространяются и поверхностными, и пространственными лучами, однако в этом случае имеют место некоторые специфические особенности. Радиоволны этого диапазона разными путями достигают точки приема в дневное и ночное время суток. Днем в основном распространяются поверхностные лучи. Однако существенное поглощение энергии на повышенных частотах этого диапазона ограничивает дальность распространения пределами сотен километров.

Пространственные лучи в дневное время суток практически полностью поглощаются слоями ионосферы вследствие сильной ионизации газовых образований солнечным излучением. Ночью, когда ионизация газа в ионосфере существенно уменьшается, интенсивность пространственного луча возрастает, что способствует эффективному распространению гектометровых волн на расстояния до нескольких тысяч и даже десятков тысяч километров. На очень больших расстояниях от передающей антенны системы связи в ночное время созданное электромагнитное поле в точке приема сообщений может образоваться как результат взаимной интерференции поверхностной и пространственной радиоволн.

На гектометровых волнах в ночное время в некоторых местах приема наблюдается *эффект замирания* волн или *фединг* (от англ. *fading* — постепенно ослабевать, исчезать) — ослабление или полное прекращение приема в результате изменения условий их распространения. Сущность этого явления заключается в следующем. Слои ионосферы неоднородны по структуре, т. е. представляют хаотические чередования пространственных областей с различной концентрацией электронов. Эти области перемещаются под действием сильных ветров, имеющих место на больших высотах. Вследствие этого в точку приема возможен одновременный приход несколько лучей, отраженных от разных неоднородностей. Это свойство распространения волн называют *многолучевостью*.

Многолучевое распространение может вызывать ряд нежелательных эффектов. Радиосигналы, доходя до получателя разными путями, будут испытывать соответственно разные временны́е задержки. Длины их путей различны, поэтому различны и фазы пришедших сигналов, что приводит к интерференции волн. Если волны в точке приема синфазны, то происходит их сложение, если же противофазны — они взаимно компенсируются.

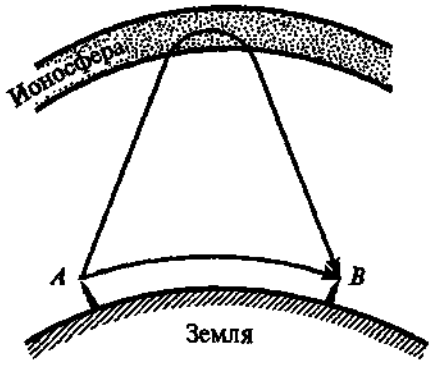



**Рис. 1.6.** Возникновение эффекта замирания

При этом амплитуда сигнала в точке приема может меняться в десятки и даже сотни раз — от очень мощного сигнала до его практически полного отсутствия. В простейшем случае интерферировать могут пространственная и поверхностная волны, пришедшие в точку приема *B* из точки излучения *A* по разным траекториям (рис. 1.6).

Особенно заметна многолучевость распространения радиоволн в системах подвижной связи. Это связано с двумя важными эффектами.

Эффект первый. При перемещении получателя сообщений условия приема периодически изменяются, так как сложение волн, приходящих с разных направлений, создает пространственную интерференционную картину. Этот эффект особенно сильно сказывается на мобильных пользователях и типичен для узкополосных систем связи. Для широкополосных систем связи действие этого эффекта сильно ослабляется из-за того, что на разных частотах в пределах широкого спектра создаются разные интерференционные картины, что и вызывает выравнивание результирующего сигнала.

Эффект второй. Время задержки радиосигналов при прохождении больших расстояний может меняться из-за изменений характеристик среды распространения, причем это по-разному сказывается на сигналах, приходящих разными путями, что при сложении радиоволн вызывает временные флуктуации уровня (громкости) принятого сигнала, приводящие к замираниям сигнала. Несинфазное изменение уровня сигнала на разных частотах спектра широкополосного сигнала приводит к сильному ослаблению влияния этого эффекта на подобные системы.

**Декаметровые волны** позволяют при относительно малых мощностях передатчика получить трассы распространения больших протяженностей. При этом возможно создать антенные устройства, обладающие заметными направленными свойствами. Декаметровые волны распространяются как поверхностными, так и пространственными лучами. Однако распространение поверхностных волн вдоль Земли ограничивается пределами прямой видимости, что обусловлено слабой дифракцией и резким возрастанием поглощения их энергии земной поверхностью.

Декаметровые волны распространяются на очень большие расстояния вдоль земной поверхности путем однократного или многократного отражения от ионосферы и Земли. Декаметровые волны достаточно хорошо отражаются от Земли и глубоко проникают в ионосферу, где за счет явления рефракции (от лат. *reffractus* — преломление) волна постепенно изменяет направление распространения и возвращается к Земле. Далее процесс повторяется, и после многократных отражений от Земли и ионосферы волна попадает на радиоприемное уст-

ройство. При этом время обхода волной земного шара составляет примерно 0,13 с. На распространение декаметровых волн большое влияние оказывают процессы в ионосфере, которые зависят от состояния Солнца.

Связь между двумя приемопередающими пунктами на декаметровых волнах имеет ряд недостатков. В частности, характер распространения электромагнитных колебаний этого диапазона является многолучевым: в точку приема приходят волны, прошедшие разный путь. Складываясь, они интерферируют, что приводит к частым замираниям сигнала. Кроме того, при радиосвязи в этом диапазоне образуются так называемые *зоны молчания* (отсутствие сигнала), где поверхностных волн уже нет, а пространственные волны, отражаясь от ионосферы, еще не действуют на этих дальностях. Для исключения этого явления вместо одной приемной антенны в пункте приема используются как минимум две антенны, размещенные на некотором расстоянии друг от друга. К настоящему времени накоплен большой материал по распространению волн этого диапазона, что обеспечивает чрезвычайно широкое применение коротковолнового диапазона в радиосвязи, радиовещании, радионавигации. Типичная картина хода лучей в пространстве Земля — слои ионосферы на декаметровых радиоволнах показана на рис. 1.7.

Одной из интересных возможностей применения многократного (скачкового) отражения волн декаметрового диапазона от ионосферы и Земли является так называемая *загоризонтная радиолокация* (радиосвязь), когда цель обнаруживают за пределами прямой видимости. Это стало возможным благодаря открытому в 1946 г. отечественным ученым Н.И. Кабановым явления сверхдальнего коротковолнового рассеивания радиоволн земной поверхностью. Это явление названо *эффектом Кабанова*, использование которого для исследования слоев ионосферы позволяет определять условия распространения волн в радиусе 9 000...12 000 км, что составляет около четверти длины экватора.

Идея загоризонтного обнаружения низколетящих самолетов и ракет, кораблей, а также баллистических и космических ракет, стартующих на удалении в несколько тысяч километров с помощью пространственных радиоволн декаметрового диапазона, была реализована в ряде стран мира. Работы по созданию и совершенствованию загоризонтных радиолокационных станций (РЛС) различного назначения продолжаются по сей день. Следует отметить, что эффект позволяет не только выявить те или иные изменения на обследуемых участках земной поверхности, но и быстро определить оптимальную длину волны для установления связи с любым пунктом Земли. Направляя в заранее рассчитанное место радиосигнал на той

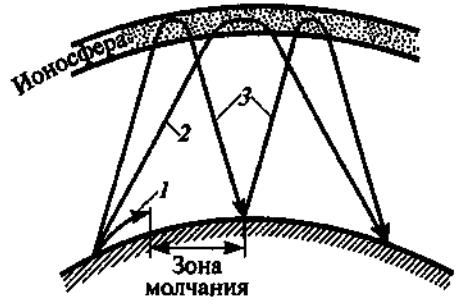



**Рис. 1.7.** Распространение декаметровой волны:
*1* — поверхностный луч; *2* — однократно отраженный луч; *3* — двукратно отраженный луч

или иной радиоволне и регистрируя интенсивность радиоотражения, несложно установить, какая из посланных волн дает лучший результат.

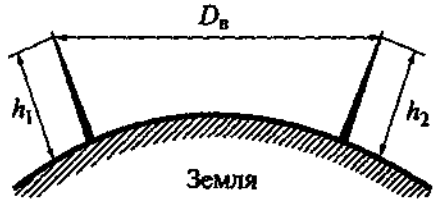



**Рис. 1.8.** Распространение метровых волн

**Метровые, дециметровые и сантиметровые волны** при распространении над земной поверхностью практически не отражаются от ионосферы, поэтому связь на них за счет пространственных волн физически почти невозможна. Исключение составляют трассы так называемого *дальнего тропосферного распространения* (за счет рассеяния волн на турбулентностях тропосферы, отражения и рассеяния на достаточно протяженных и стабильных слоях тропосферы). Однако радиоволны этого диапазона обладают очень малой дифракцией и не могут огибать препятствия на пути распространения. Поэтому волны длиной до 10 м и менее распространяются только в пределах прямой видимости между передающей и приемной антеннами систем связи (рис. 1.8). Известна приближенная формула для определения расстояния радиосвязи $D_в$ между передающей и приемной антеннами с высотами $h_1$ и $h_2$, м:

$$D_в = \sqrt{2R_З}\left(\sqrt{h_1} + \sqrt{h_2}\right) \approx 3{,}57\left(\sqrt{h_1} + \sqrt{h_2}\right) \qquad (1.10)$$

Здесь параметр $R_З$ — радиус Земли (около 6 370 км). В общем случае зона охвата для данного типа волн ограничена кривизной поверхности Земли.

В принципе дальность связи на таких волнах не так уж и мала. Теоретически два человека выше среднего роста (около 1,9 м) на равнине могут видеть друг друга на расстоянии порядка 10 км — это есть дальность связи в этих диапазонах с высотами антенн около 1,7...2 м. Для увеличения дальности связи между двумя пунктами на метровых и более коротких волнах в системах передачи информации передающие и приемные антенны поднимают высоко над поверхностью Земли. Наглядное подтверждение — Останкинская телебашня (например, телевизионная антенна, установленная на высоте 300 м, обеспечивает покрытие территории радиусом до 70 км).

В случае распространения радиоволн в пределах прямой видимости для увеличения дальности радиосвязи можно также создавать цепочки из приемопередающих устройств (ретрансляторов, см. далее), расположенных на некотором (порядка 50 км) расстоянии друг от друга. Это так называемые *многоканальные радиорелейные линии связи*.

**Пример 1.3.** Определить, чему равна дальность радиосвязи на метровых волнах, если высота передающей антенны 50 м, а приемной — 16 м?

Р е ш е н и е. По формуле (1.10) находим искомую дальность радиосвязи:

$$D_в = 3{,}57\left(\sqrt{50} + \sqrt{16}\right) = 39{,}3 \text{ км}.$$

Электромагнитные колебания *метровых, дециметровых и сантиметровых диапазонов волн* представляют большие возможности для осуществления основной задачи радиотехники — передачи информации. Радиоволны данных диапазонов обладают большей частотной емкостью (частотным наполнением), что дает возможность создавать радиотехнические системы с частотной, импульсной и импульсно-кодовой (цифровой) и другими видами широкополосной модуляции. Это обеспечивает высокую помехозащищенность линий радиосвязи. По ним можно передавать широкополосные радиосигналы, например сигналы телевизионного вещания, служебной и подвижной связи.

Последние десятилетия интенсивно изучаются и одновременно осваиваются в радиотехнике *диапазоны миллиметровых, децимиллиметровых и оптических волн.* Причинами повышенного внимания к этим диапазонам волн является их по существу неисчерпаемая частотная емкость и относительная легкость получения остронаправленного излучения (приема) вследствие малых размеров антенн. Особенно велики возможности волн оптического диапазона.

**Электромагнитная совместимость радиотехнических систем**

Одной из важных проблем современной радиотехники и систем связи является проблема электромагнитной совместимости (ЭМС) систем передачи информации. Лавинообразный рост радиотехнических систем всех диапазонов радиоволн и различного назначения делает эту проблему чрезвычайно важной. Прежде всего имеются в виду взаимные радиопомехи между системами радиосвязи, размещенными как на одном узле, объекте или территории, так и разнесенными на большие расстояния. Тенденция к укорочению длины рабочей волны приводит к возможности резкого увеличения числа радиосистем, работающих на одинаковых или близких частотах. При этом причинами, обусловливающими взаимное влияние каналов систем передачи информации в диапазоне частот 1...300 ГГц, являются дифракция на неоднородностях трассы связи, рассеяние радиоволн на гидрометеорах и турбулентностях, а также образование паразитных волноводов. Эти причины проявляются по-разному. Видимо, наиболее серьезной из них является образование паразитного волновода (например, между Землей и атмосферой в скоплениях гидрометеоров и т. д.), способствующего связи между разными радиосистемами. Другие причины и их воздействие на электромагнитную совместимость существенно зависят от рельефа местности, района земного шара, времени года, суток и др.

Существенное влияние на распространение радиоволн и оптических волн в диапазоне 1...300 ГГц оказывает атмосфера Земли. Так, область частот в диапазоне 50 ГГц соответствует максимуму поглощения радиоволн атмосферой. В этих условиях работа радиотехнических систем передачи информации возможна только в так называемых *окнах прозрачности*, где поглощение радиоволн атмосферой минимально. Поэтому, например, наиболее сильное влияние на космические трассы протяженностью во многие миллионы километров оказывает атмосферный участок линии радиосвязи. На непрозрачном участке атмосферы волны полностью поглощаются и связь практически невозможна.

## 1.3. Антенны

Направленность излучения электромагнитных волн имеет большое значение в радиотехнике, и в частности в радиолокации, радионавигации, связи. В общем случае практически любой отрезок металлического проводника, по которому протекает переменный ток, создает в окружающем пространстве электромагнитное поле. Точно так же в любом отрезке проводника, находящемся в электромагнитном поле, индуцируется переменная электродвижущая сила (ЭДС). Уровень энергии электромагнитного поля, возникающий в проводнике, существенным образом зависит от его конфигурации, соотношения размеров проводника и длины волны электромагнитных колебаний. Поэтому для излучения и приема электромагнитных колебаний, переносящих информацию, используются специфические радиотехнические устройства, называемые *антеннами*. Направленность излучения можно обеспечить, если антенное устройство по размерам существенно превышает длину излучаемой или принимаемой волны.

Теория и экспериментальные исследования показывают, что простейшую излучающую (равно и приемную) антенну (рис. 1.9, *а*) можно сделать из отрезка электрического проводника размером в половину длины излучаемой волны (аналог колебательного контура), в середину которого включен генератор колебаний высокой частоты. Антенну, полученную из отрезка проводника размером в половину длины волны излучения, называют *полуволновым вибратором*.

Интенсивность излучения электромагнитных колебаний полуволнового вибратора в различных направлениях оказывается неодинаковой, а это имеет существенное значение для антенн. Наглядное представление о характере излучения и эффективности работы любой антенны дает *диаграмма направленности* (ДН), отражающая зависимость плотности потока мощности от направления в пространстве. Простейший вид диаграммы направленности формируется элементарным диполем, у которого наибольшая интенсивность излучения имеет место в направлении, перпендикулярном проводнику; вдоль проводника излучение практически отсутствует. Увеличение длины излучающего проводника приводит к существенному изменению электромагнитного поля, поскольку ка-

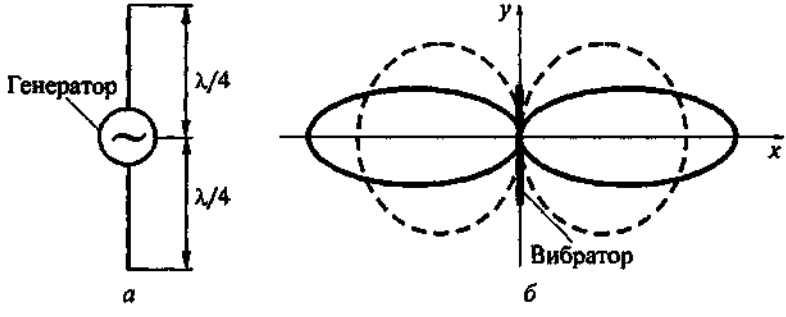



**Рис. 1.9.** Полуволновый вибратор:
*а* — устройство; *б* — диаграмма направленности

ждый элементарный излучатель (диполь) создает свое поле излучения. Полное электромагнитное поле излучения в любой точке пространства определяется суперпозицией элементарных полей. На рис. 1.9, *б* сплошной линией показана диаграмма направленности вертикально расположенного полуволнового вибратора, построенная в меридиональной (азимутальной) плоскости. Здесь же штриховой линией показана диаграмма направленности элементарного диполя. Отметим очевидное — для вертикально расположенного полуволнового вибратора диаграмма направленности в горизонтальной плоскости имеет вид окружности (круговая ДН), а в вертикальной — вытянутых восьмерок. Подобная форма диаграммы направленности в горизонтальной плоскости свидетельствует о всенаправленности излучения полуволнового вибратора вдоль поверхности Земли.

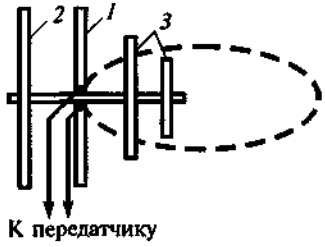



**Рис. 1.10.** Антенна волновой канал

Концентрация излучаемой энергии в узком секторе позволяет резко увеличить дальность распространения волн. Это можно осуществить с помощью многовибраторных антенн (рис. 1.10). Если параллельно излучающему вибратору *1* на расстоянии чуть менее $\lambda/4$ расположить другой полуволновый вибратор *2* (*рефлектор*), то в нем под воздействием поля излучения будет индуцироваться переменный ток, который возбудит свое электромагнитное поле (рефлектор действует как отражатель). Данное электромагнитное поле, суммируясь с электромагнитным полем излучающего вибратора, сузит диаграмму направленности со стороны последнего. Если перед излучающим вибратором антенны дополнительно расположить несколько вспомогательных вибраторов *3* (*директоров*), то, подбирая их длину и расстояние между ними, можно еще более сузить диаграмму направленности в направлении излучения и ослабить ее с противоположной стороны. Такую конструкцию называют *антенной волновой канал, директорной антенной, антенной Уда–Яги* (по имени впервые описавших ее в 1921 г. японских изобретателей С. Уда, Х. Яги) и широко используют в качестве приемной телевизионной антенны. Отметим такой важный параметр антенны, как *коэффициент направленного действия* (КНД), показывающий во сколько раз средняя мощность, излучаемая (принимаемая) в заданном направлении, больше средней мощности, излучаемой (принимаемой) по всем направлениям. Параметр, равный произведению КНД и КПД, характеризует *коэффициент усиления антенны*.

В диапазонах километровых волн габаритные размеры даже одновибраторных приемо-передающих антенн оказываются столь большими, что возникают существенные трудности в их технической реализации. В частности, поднимать излучатель на высоту, сравнимую, например, с длиной волны в 3...5 км, практически невозможно. Оказалось, что имеется очень простое решение этой проблемы. Поскольку Земля для этих диапазонов радиоволн является достаточно хорошим проводником, то в качестве антенн можно использовать не полуволновый, а четвертьволновый заземленный вибратор.

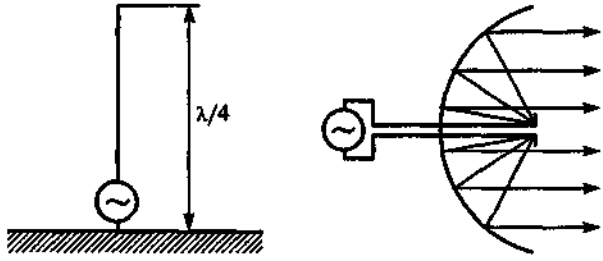



**Рис. 1.11.** Простейшие антенны:
*а* — четвертьволновая вибраторная; *б* — параболическая

Такую антенну называют *четвертьволновой вибраторной*: она представляет собой вертикальную излучающую конструкцию, между основанием которой и Землей включен источник высокочастотных колебаний (рис. 1.11, *а*). В этом случае отраженные от Земли волны будут создавать электромагнитное поле, аналогичное тому, которое возбуждается незаземленным четвертьволновым элементом полуволнового вибратора. Результирующее же электромагнитное поле над поверхностью Земли будет практически таким же, как и излученное одним полуволновым вибратором. Однако мощность такого излучения оказывается почти в два раза меньше.

Для создания остронаправленного излучения электромагнитных колебаний в диапазоне длин волн менее метра используют зеркальные антенны с параболическим отражателем (изобретена в 1888 г. Г. Герцем). Излучатель зеркальной антенны должен быть расположен точно в фокусе параболического зеркала. В этом случае подобная антенна будет концентрировать (фокусировать) энергию электромагнитного излучения в достаточно узком секторе пространства, подобно отражающему зеркалу оптического прожектора или простого электрического фонарика (рис. 1.11, *б*).

В последние годы, особенно в специальных радиотехнических системах передачи информации (прежде всего военного назначения) и системах подвижной и спутниковой связи, широко применяют так называемые *фазированные антенные решетки* (ФАР; первая ФАР создана в 1937 г. Г.Т. Фельдманом — *G.T.Feldman* и Г.Б. Фрисом — *G.B.Fris*, США). Подобные антенны представляют собой совокупность (решетку) определенным образом расположенных в одной плоскости (в площади квадрата или прямоугольника) отдельных элементарных излучателей электромагнитных волн, питаемых параллельно через индивидуальные устройства сдвига фаз (фазовращатели) одним источником высокочастотных колебаний (рис. 1.12) или системой *когерентных* (*сфазированных*) источников энергии. Электромагнитные поля, создаваемые каждым отдельным излучателем, суммируясь в пространстве вблизи антенны, образуют единый электромагнитный фронт волны. Это поле обычно представляет узконаправленный луч энергии — требуемую диаграмму направленности (см. рис. 1.12).

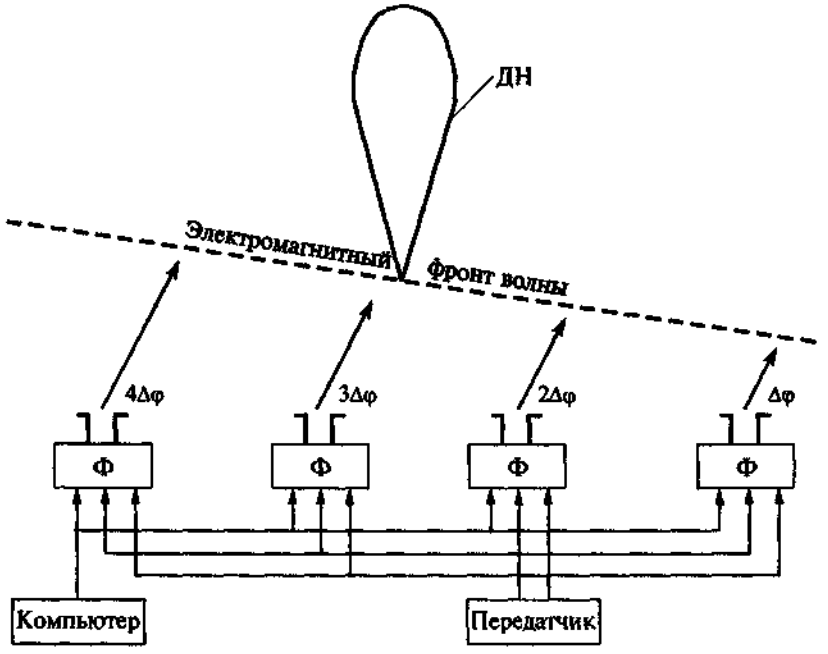



**Рис. 1.12.** Структурная схема фазированной антенной решетки

К важнейшему свойству ФАР относится возможность электронным способом, практически мгновенно, изменять положение диаграммы направленности (*сканировать* от англ. *scan* — поле зрения) в пространстве путем одновременного изменения с помощью фазовращателей Ф определенных значений фаз элементарного излучателя (сдвиг фазы составляет от 0° до $n\Delta\varphi°$; $n$ — число излучателей; $\Delta\varphi°$ — *дискретный сдвиг фазы* одного фазовращателя) высокочастотных колебаний, подводимых к каждому излучателю. Излучателей может быть много (их число может достигать 10 000) и они управляются с помощью компьютера.

Применение ФАР для создания остронаправленных диаграмм излучения позволяет реализовать высокую скорость обзора пространства и способствует увеличению количества получаемой информации о местоположении в пространстве различных объектов, таких как ракеты, самолеты, корабли и т. д. Сейчас ФАР все чаще применяют в подвижной и спутниковой радиосвязи в виде так называемых *интеллектуальных*, или *умных антенн* (*smart-antennas*), позволяющих изменять и положение луча, и мощность излучения в различных секторах. Смарт-антенна автоматически находит наиболее сильный сигнал для мобильного телефона.

В принципе практически все антенны обладают свойством *обратимости* (*взаимности*). Это свойство заключается в том, что передающие и приемные антенны могут использоваться (особенно распространено в различных системах передачи информации, и в частности, в системах подвижной связи) как на передачу, так и на прием радиоволн. При этом остаются неизменными все характеристики и параметры антенн.

## 1.4. Основные сведения о радиотехнических системах и сетях связи

Понятие «система» стало одним из наиболее распространенных в области радиотехники последних трех-четырех десятилетий. По используемым в качестве переносчиков типам колебаний физической среды и диапазону излучений волн системы подразделяются на сейсмические, акустические и электромагнитные (включая оптические). Любую техническую систему, действие которой основано на непосредственном использовании высокочастотных электромагнитных колебаний радиодиапазона для сбора, передачи, извлечения, обработки или хранения информации, называют *радиотехнической системой* (упрощенно радиосистемой). Упрощенную структуру построения радиотехнической системы можно представить в виде своеобразной пирамиды (рис. 1.13).

Фундаментом пирамиды служит *элементная база*, содержащая резисторы, катушки индуктивностей, конденсаторы, трансформаторы, диоды, биполярные и полевые транзисторы, аналоговые и цифровые микросхемы, микропроцессоры, резонансные цепи, монолитные фильтры, элементы СВЧ-техники и пр. Кстати, элементная база любой радиотехнической системы определяет технический уровень радиоаппаратуры.

Из элементов составляют второй уровень пирамиды — *цепи* (колебательные контуры, дифференцирующие и интегрирующие цепи, фильтры, ограничители уровня, формирующие цепи и пр.).

Узлы конструктивно и технологически объединяют в достаточно сложные радиотехнические цепи — *каскады*: автогенераторы, модуляторы, демодуляторы, преобразователи частоты, усилители сверхвысокой, высокой, промежуточной и низкой частоты и т. д.

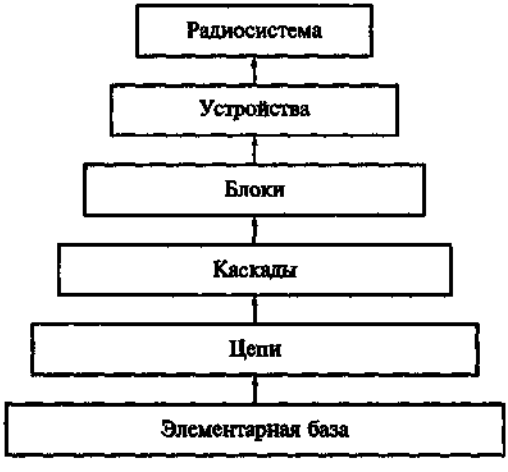



**Рис. 1.13.** Упрощенная структура построения радиотехнической системы

Следующий уровень — *блоки*, к которым относят антенно-фидерный тракт, малошумящий СВЧ-усилитель приемника, каскады усиления мощности высокочастотных и СВЧ-колебаний, кодек, модем, линейный тракт приемника, устройство цифровой обработки принимаемого сигнала, систему управления и пр.

Наиболее сложный уровень пирамиды включает функционально законченные *устройства* — приемники, передатчики и другую аналогичную радиоаппаратуру, которые работают самостоятельно в составе различных радиосистем.

Венцом пирамиды является — *радиосистема*.

В последние годы в элементной базе резко возросла доля микросхем большой и сверхбольшой интеграции. Поэтому в радиотехнических устройствах зачастую применяют только интегральные микросхемы, что позволяет три нижних уровня пирамиды (см. рис. 1.13) технологически объединить в один.

По информационному назначению радиотехнические системы делят на четыре основных класса:

- передача информации (радиосвязь, радиовещание, телевидение);
- извлечение информации (обнаружение и измерение — радиолокация, радионавигация, радиоастрономия, радиоизмерения и т.д.);
- радиотелеуправление (беспилотные летательные аппараты и др.);
- разрушение информации (радиопротиводействие).

В настоящее время широкое применение находят радиотехнические комплексы, состоящие из нескольких радиосистем, в которых для обработки информации и управления различными объектами используют мощные компьютеры. К ним относят системы спутниковой и космической связи, глобальные системы связи, системы контроля и управления воздушным движением, ракетные и космические комплексы и т. д. Такие системы можно назвать *комплексными*.

Большинство радиотехнических систем предназначено для доставки получателю определенной информации. Поэтому практически все радиотехнические системы можно называть *информационными*. Конкретное содержание информации, метод ее получения и использования получателем различны, методы же обработки, представления и передачи в значительной степени являются общими. Во всех случаях приходится иметь дело с сообщением о каком-либо физическом процессе и принимаемым сигналом, отображающим сообщение, а также естественной или искусственной помехой, искажающей принимаемый сигнал.

В радиотехнических системах передачи информации сообщения, подлежащие передаче, поступают извне, от каких-либо источников, и роль системы состоит в том, чтобы передать их получателю. Все радиосистемы этого класса начинаются с преобразователей сообщений, задача которых — преобразовать поступающие сообщения в электрические сигналы, наиболее эффективные для передачи. На выходе приемника радиотехнической системы обычно также имеются преобразователи, выдающие сообщения в виде, удобном для получателя.

В радиотехнических системах извлечения информации интересующая получателя информация не содержится в сигнале передатчика, а иногда передатчик в системе информации вообще отсутствует. В этих системах большей частью имеют дело с информацией, характеризующей параметры трассы, по кото-

рой распространяются электромагнитные волны. Информацию несут направление трассы, ее длина, а также скорость изменения этих параметров во времени. По значениям этих величин можно получать сведения о положении и характеристиках движения излучающих или отражающих объектов (радиолокационные системы, системы траекторных измерений). В других задачах информацией являются такие параметры трассы, как показатели поглощения и преломления (радиометеорологические системы, радиометрические системы). В некоторых случаях полезная информация заключается в самой структуре принимаемого радиосигнала (радиоастрономические и разведывательные радиосистемы).

Значительного развития достигли системы радиолокации, широко применяемые в комплексах военного и гражданского назначения. Особое место заняли сегодня также километровые (длинноволновые) и метровые (ультракоротковолновые) системы наземной и спутниковой навигации, без которых ныне немыслимо осуществлять безопасное мореплавание и полеты самолетов.

Системы радиотелеуправления предназначены для обеспечения движения летательных аппаратов по заданной траектории или в заданном районе. Такие системы широко применяют для решения задач дистанционного управления движением космических объектов.

Радиосистемы разрушения информации служат для создания помех нормальной работе конкурирующей радиосистемы путем излучения мешающего сигнала или приема, умышленного искажения и переизлучения сигнала.

Основной частью практически всех радиотехнических систем является *радиотехнический канал* (радиоканал), состоящий из радиопередающего (передающего) и радиоприемного (приемного) устройства и линии связи. Упрощенно, *передающим* называют устройство, преобразующее сообщение в передаваемый сигнал, а *приемным* — устройство, преобразующее принятый сигнал в исходное сообщение. В пассивных радиотехнических системах, например в теплолокаторах (ИК-пеленгаторах, тепловизорах), передающее устройство отсутствует.

Важной частью радиоканала системы связи является линия связи (среда распространения), которая оказывает существенное влияние на достоверность и качество принимаемого сообщения.

*Линией связи* называют физическую среду (космическое пространство, свободное пространство — воздух в нейтральном или ионизированном состояниях, земная поверхность, морская вода, волноводы, кабели, волоконно-оптические линии и пр.) и совокупность аппаратных средств, используемых для передачи сигналов от передатчика к приемнику. Простейшая линия связи может представлять собой два провода, соединяющие передатчик и приемник. В этом случае речь идет о проводной, например телефонной, связи. Сейчас все чаще осуществляют связь и без проводов с помощью электромагнитных или световых волн (через световоды). В системах радиосвязи линией связи является, как правило, область пространства, в котором распространяются электромагнитные волны от передающего устройства к приемному устройству.

Обобщенная структурная схема радиоканала системы передачи информации показана на рис. 1.14.




**Рис. 1.14**. Обобщенная структурная схема радиоканала

Отметим, что составными частями практически любой радиотехнической системы являются также устройства питания, антенно-фидерные, электронно-вычислительные и другие устройства.

Основными задачами, решаемыми радиосистемой при приеме информации, являются: обнаружение сигнала на фоне помехи; различение сигналов на фоне помехи; оценка параметров сигнала; воспроизведение сообщения.

### *Радиотехнические системы передачи информации*

Для систем передачи информации важна физическая природа ее восприятия. По этому признаку информацию можно разделить на слуховую, зрительную и «машинную». Первые два вида соответствуют наиболее емким каналам восприятия информации человеком. Пропускная способность слухового канала человека составляет тысячи десятичных единиц информации, а зрительного — миллионы. «Машинная» информация, как правило, предназначена для обработки ее компьютерами. Здесь пропускную способность каналов передачи информации надо согласовывать со скоростью обработки ее компьютером — до нескольких десятков миллионов битов в секунду. С помощью компьютеров обычно возможна прямая цифровая обработка слуховой и зрительной информации.

Радиотехнические системы передачи информации (сообщений) предназначены для передачи информации на расстояние с помощью электромагнитных колебаний. К ним прежде всего относятся различные телекоммуникационные системы, или системы электросвязи. Это системы связи, радиовещания, телевидения, видеоконференцсвязи и т. д.

Характерным для радиосистем передачи информации любого назначения (связных, телеметрических, вещательных, телевизионных, командных и т. д.) является пространственное разнесение передающего и приемного устройств.

Функционирование подавляющего большинства радиотехнических систем передачи информации основано на свободном распространении электромагнитных колебаний, которые передающими антеннами излучаются в окружающее пространство. К передающим антеннам от радиопередатчика подводятся колебания высокой (несущей) частоты, один из параметров которых изменяется (промодулирован) по закону передаваемого сообщения. Распространяясь в определенных направлениях, радиоволны достигают приемной антенны, в которой под их воздействием наводится ЭДС (токи высокой частоты), переносящая передаваемую информацию.

### Общие понятия о телекоммуникационных сетях и системах

Для обмена сообщениями между многими территориально разнесенными пользователями (абонентами) создают *сети связи*, обеспечивающие с установленным качеством и в заданное время передачу и распределение сообщений по заданным адресам.

Основой электросвязи любого государства (в том числе Российской Федерации) является *взаимоувязанная сеть связи* (ВСС) — совокупность технологически сопряженных сетей связи общего пользования (ОП), ведомственных, корпоративных (для определенного круга пользователей) и других сетей электросвязи на территории страны независимо от ведомственной принадлежности и форм собственности, обеспеченная общим централизованным управлением, которое строится по иерархически-территориальному принципу.

ВСС Российской Федерации имеет разнообразную *топологию*. Конфигурация соединения элементов сети представляет ее топологию. *Магистральные линии* действующей сети образуют сетевидную структуру: между любыми сетевыми узлами имеются по крайней мере два-три независимых пути соединения. Это обеспечивает экономичность и высокую надежность сети. *Зоновые сети* имеют радиально-узловую топологию, к ней добавляется соединение по принципу «каждый с каждым» для групп АТС (автоматическая телефонная станция), звездообразное подключение абонентов к станции, а также *рокадное* соединение — *круговой принцип* прохождения связей поперек радиусов в обход узлов. Развитие ВСС идет в направлении широкого внедрения новых кольцевых структур, особенно на строящихся телекоммуникационных сетях — локальных сетях.

Связь Российской Федерации представляет собой совокупность сетей и служб связи и функционирует на территории Российской Федерации как взаимоувязанный производственно-хозяйственный комплекс. Структура связи Российской Федерации (рис. 1.15) включает в себя электросвязь и почтовую связь.

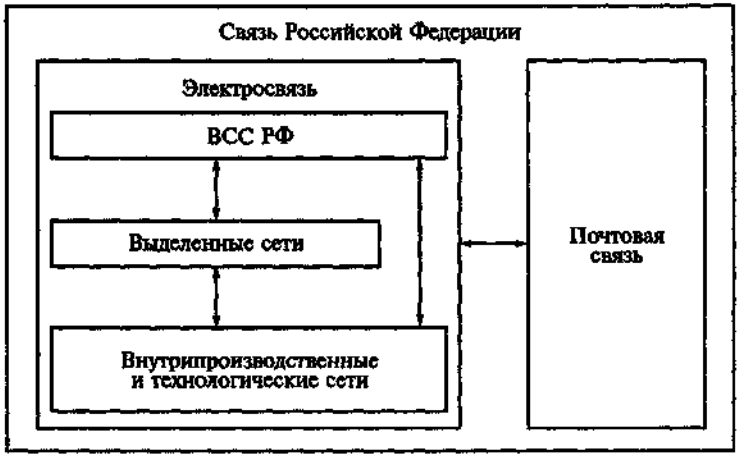



**Рис. 1.15.** Структура связи Российской Федерации

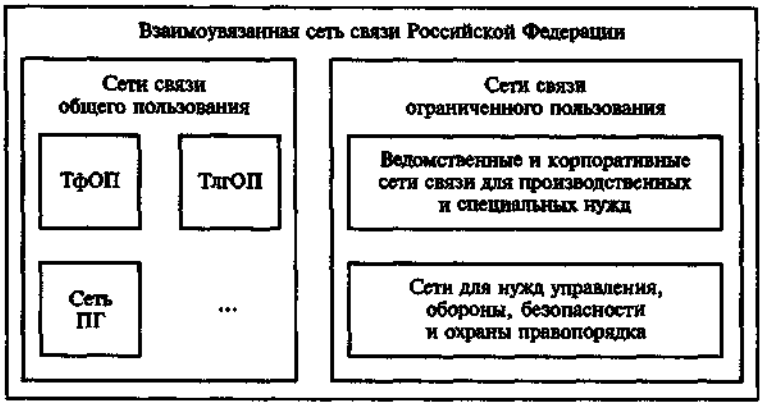



**Рис. 1.16.** Состав взаимоувязанной сети связи Российской Федерации

Эти виды связи в своей основе действуют независимо друг от друга, а имеющееся взаимодействие (служебная электросвязь в почтовом ведомстве, электронная почта, почтовые уведомления в службах электросвязи) показаны стрелками на рис. 1.15.

Управление связью Российской Федерации регулируется Министерством связи и массовых коммуникаций. Состав взаимоувязанной сети Российской Федерации показан на рис. 1.16. В состав ВСС входят сети связи ОП, в том числе телефонные сети общего пользования (ТфОП; *Public Switched Telephone Network — PSTN*), телеграфная сеть общего пользования (ТлгОП; *Public Switched Telegraphe Network*), сеть передачи газет (ПГ) и другие вторичные сети по видам услуг связи, а также *ведомственные* и *корпоративные сети* ограниченного пользования для определенных контингентов абонентов. Это *наложенные сети*, и они используют каналы и тракты ВСС. К ним относятся ведомственные и корпоративные сети для производственных и специальных нужд (например, рыбного хозяйства, службы почтовой связи, нефте-, газо- и угледобыч и пр.) и сети связи для нужд управления, обороны, безопасности и охраны правопорядка (например, МЧС, пожарной службы, милиции и т. д.).

В ВСС Российской Федерации могут не входить, т. е. не иметь централизованного управления, *выделенные сети* (сети ведомств, организаций и фирм, имеющих собственные оборудованные линии, например, сеть связи на железнодорожном транспорте, на газопроводах, компьютерные сети банков и др.), *внутрипроизводственные* и *технологические сети* (сети учреждений и организаций, например локальные вычислительные сети по управлению производственными процессами, телефонные сети абонентов учрежденческих АТС и т. д.).

Основные органы управления ВСС Российской Федерации — Главный (национальный) центр управления (ГЦУ), региональные центры управления (РЦУ) в выделенных зонах на территории одного или нескольких субъектов Российской Федерации, а также узловые территориальные пункты управления (ТПУ) и информационные исполнительные пункты (ИИП).

Распределение потоков сообщений по заданным адресам осуществляется на *узлах связи* с помощью коммутационных устройств. *Коммутация* — это процесс создания последовательного соединения функциональных единиц для транспортировки информации. В целом же задачу распределения информационных потоков выполняет *система коммутации*, состоящая из собственно сети, коммутационных станций и узлов коммутации (УК), системы подключения пользователей и оконечных пунктов — терминальных устройств. Наиболее важную роль в ней играют УК, обеспечивающие установление, поддержание и разъединение соединений между терминалами (телефонными аппаратами, компьютерами, различной аппаратурой), каждому из которых присвоен адрес (номер).

По способу распределения сообщений сети связи делятся на *некоммутируемые* и *коммутируемые*. В первом случае связь между абонентами осуществляют по постоянно закрепленным каналам по принципу «каждый с каждым». Во втором случае абоненты связываются между собой не непосредственно, а через узлы коммутации. Структурно сеть связи представляет собой совокупность оконечных (абонентских) устройств, каналов связи (линий связи) и узлов коммутации. Задача оптимального построения телекоммуникационных сетей является одной из важнейших задач теории и техники связи. Задачу решают с помощью теории массового обслуживания — теории *трафика* (от англ. *traffic* — информационная нагрузка сети связи, максимальные передаваемые потоки информации, количество информации, поступающей через сеть связи) и теории графов.

В зависимости от скорости передачи информации отечественные каналы связи подразделяются на три вида:

- цифровая интегральная сеть ЦИС-32;
- узкополосная цифровая сеть интегрального обслуживания — ЦСИО-У (ISDN-N; англ. ISDN — *Integrated Services Digital Network*; иногда проще — IDN — *Integrated Digital Networks*);
- широкополосная цифровая сеть интегрального обслуживания ЦСИО-Ш (ISDN-B).

В цифровые сети связи типа ISDN можно включать такие виды связи и сетей: сеть передачи данных (*Packet Data Network* — PDN); сотовая связь; служба обработки сообщений — электронная почта (*E-mail*); всемирная компьютерная сеть Internet. Ряд сетей связи могут функционировать как выделенные сети со своими оконечными терминалами, цифровыми каналами. Они могут быть включены в ЦСИО-У, если оконечные терминалы будут работать со скоростью передачи не выше 64 кбит/с. Напомним, что, как и телеграфные сети, сети передачи данных по скорости передачи разделяют на низкоскоростные — до 200 бит/с; среднескоростные — 600...1200 бит/с и высокоскоростные 2,4...96,0 Кбит/с и более. В цифровой интегральной сети ЦИС-32 скорость передачи информации составляет 32 Кбит/с, а в сети ISDN-B — 8 ... 565 Мбит/с и более.

Сетевые цифровые технологии развивались до последнего времени параллельно для глобальных и локальных сетей. Технологии глобальных сетей были направлены в основном на развитие цифровых телефонных сетей, используемых для передачи голоса. Технологии локальных сетей — напротив, использовались, в основном, для передачи данных.

Прогресс в развитии средств связи и компьютерной техники и технологий неизбежно привел к переходу в развитых странах от общества индустриального к обществу индустриально-информационному. *Международный Союз Электросвязи* (МСЭ), занимающийся стандартизацией телекоммуникационных систем и сетей, ввел новое для связи понятие — *интеллектуальная сеть* (ИС; *Intelligent Network* — IN), отличительным признаком которой является быстрое, эффективное и экономное предоставление информационных услуг массовому пользователю в любой момент времени.

**Эталонная модель взаимодействия открытых систем**

Для того чтобы взаимодействовать, люди используют общение на каком-либо языке. Если невозможно разговаривать друг с другом непосредственно, применяются вспомогательные средства для передачи сообщений. Одним из таких средств является система почтовой связи. В ее составе можно выделить определенные функциональные уровни, например, уровень сбора и доставки писем из почтовых ящиков на ближайшие почтовые узлы связи и в обратном направлении, уровень сортировки писем в транзитных узлах и т. д. Принятые в почтовой связи всевозможные стандарты на размеры конвертов, порядок оформления адресов и т. д. позволяют отправлять и получать корреспонденцию практически из любой точки земного шара. Аналогичная картина наблюдается и в области электронных коммуникаций, где рынок компьютеров, коммуникационного оборудования систем и сетей связи необычайно широк и пока не стандартизован. Поэтому создание современных информационных систем стало невозможным без использования общих подходов при разработке, без унификации характеристик и параметров их составных компонентов.

Для эффективной реализации пропускной способности (емкости) каналов передачи информации и коммутационных узлов сетей связи необходимо соблюдение определенного набора стандартных правил, которые должны быть построены по принципу некоторой иерархии, т. е. разделены на определенные функциональные уровни. Набор стандартных процедур взаимодействия и программных средств, обеспечивающих установление связи, переключение, прерывание связи при необходимости и т. п., в настоящее время в основном реализуется программно. Эти правила (стандарты) разрабатываются рядом международных организаций электросвязи.

В основу многоуровневой архитектуры связи положена концепция эталонной модели *взаимодействия открытых систем* (ВОС; OSI — *Open System Interconnection*), обеспечивающая введение единых стандартов на международном уровне для вновь создаваемых информационных сетей. Она описана стандартом ISO 7498. Эталонная модель взаимодействия открытых систем ВОС/OSI разработана в начале 80-х гг. XX в. Международной организацией по стандартизации (МОС; *International Organization of Standardization* — ISO — ИСО) совместно с Международным Союзом Электросвязи X.200. Модель ВОС/OSI как единый комплекс стандартов является основой для взаимной совместимости оборудования и программ различных поставщиков.

Эталонная модель BOC/OSI ориентирована на выполнение следующих функций: представление данных в стандартной форме; связь между процессами информационного обмена и синхронизация их работы; управления информационно-вычислительными ресурсами; контроль ошибок и сохранности данных; управление базами данных и запоминающими устройствами; поддержка программ, обеспечивающих технологию передачи и обработки данных; тестирование и др.

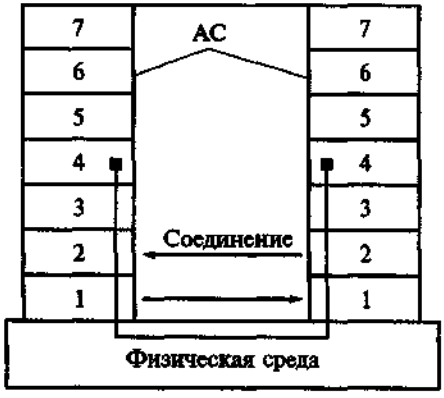
Под термином «открытая система» подразумевается система, которая может взаимодействовать с любой другой, удовлетворяющей требованиям открытой системы. В более широком смысле открытая система — это система, реализующая открытые спецификации на интерфейсы (*Interface* — программно-аппаратное средство, обеспечивающее взаимодействие двух систем или процессов в точке их сопряжения; устройство обмена информацией, или, согласно русской жаргонной терминологии, *стык*; в простейшем случае это разъем между аналоговым и цифровым узлом, или *аналого-цифровым преобразователем* — АЦП, от англ. — *Analog-to-Digital Converter*, *ADC*), службы и форматы данных, достаточные для того, чтобы обеспечить:

- возможность переноса (от англ. *portability* — мобильность) прикладных систем, разработанных должным образом, с минимальными изменениями на широкий диапазон систем;
- совместимую работу (от англ. *interoperability* — переносимость) с другими прикладными системами на локальных и удаленных платформах;
- взаимодействия с пользователями в стиле, облегчающим последним переход от системы к системе.

Основные элементы таких открытых систем, как сети связи — *абонентские станции* (АС) и физическая среда для передачи информации. АС, представляющие собой мобильные станции пользователей и устройства коллективного пользования, обеспечивают некий прикладной процесс — тот или иной вид доставки, обработки и отображения информации для нужд пользователей, например запрос и заказ билетов на концерт с домашнего персонального компьютера. Под физической средой здесь понимается совокупность цифровых каналов, позволяющих передавать биты информации.

Взаимодействие АС может иметь различный уровень иерархии преобразования информации при обмене. Сеть связи, содержащая АС, которые удовлетворяют этой модели и таким образом относятся к открытым системам, в свою очередь является открытой системой. Все современные сети связи — *открытые*. В отличие от первоначальных сетей связи, предусматривающих подключение рабочих станций только к своему центральному компьютеру, станции открытых систем могут взаимодействовать с АС любых других систем связи, выполненных по стандарту BOC/OSI.

Для упрощения разработки и реализации сетевой архитектуры каждая система разбивается на ряд функциональных уровней. При этом взаимодействие систем в сети представляется в виде взаимодействия между элементами (логическими объектами) систем одного функционального уровня (рис. 1.17).




*а*

| Уровень | Номер |
|---|---|
| Прикладной | 7 |
| Представительный | 6 |
| Сеансовый | 5 |
| Транспортный | 4 |
| Сетевой | 3 |
| Канальный | 2 |
| Физический | 1 |

*б*

**Рис. 1.17.** Эталонная модель взаимодействия BOC/OSI:
*а* — основные элементы сети; *б* — семиуровневая архитектура взаимодействия

Эталонная модель международного стандарта BOC/OSI использует семь функциональных уровней, показанных на рис. 1.17, *а*.

Семнуровневая архитектура взаимодействия АС имеет следующие функции уровней (рис. 1.17, *б*), т. е. следующие протоколы (от англ. *Protocol* — набор правил, управляющих информационным обменом взаимодействующих устройств) взаимодействия АС при соединениях систем сети между собой.

*Первый уровень — физический* (*Physical layer*), он обеспечивает интерфейс с физической средой (физическую среду часто называют *нулевым уровнем*). Реализует управление каналом связи, что сводится к его подключению и отключению и формированию сигналов, представляющих передаваемые сообщения и/или данные. На физическом уровне биты информации или данных передают по физическим каналам связи, таким, например, как коаксиальный кабель, витая пара, оптический кабель, радиоканал или цифровой канал связи. Этот уровень определяет уровень физических средств передачи, такие как полоса пропускания (*Bandwidth*), помехозащищенность, волновое сопротивление, потери и т. п. На физическом уровне задают характеристики сигналов электросвязи — электрических и оптических, такие как уровни сигналов, тип кодирования, скорость передачи и др. На этом же уровне стандартизуются собственно физические интерфейсы — в частности, типы разъемов и назначение каждого контакта.

*Второй* (*канальный*) *уровень* — формирование пакетов данных (кадров). Обеспечивает надежную передачу сообщений через физический канал, организуемый на первом уровне. Для обеспечения надежности используется средство контроля принимаемых сообщений, позволяющие выявлять ошибки в поступающих сообщениях. Уровень управления каналом через недостаточно надежный физический канал обеспечивает передачу сообщений с необходимой достоверностью. Канальный уровень обеспечивает передачу сообщений в виде временных блоков данных — кадров (*frame*) — фреймов через физический канал. Основное назначение этого уровня — прием и передача кадров в сеть. Он физи-

чески адресует передаваемые сообщения, обеспечивая правильность использования физического канала, выявляет неисправности и ошибки передачи, синхронизации кадров и управления потоками сообщений.

*Третий (сетевой) уровень* — сегментирование и объединение блоков. Сетевой уровень обеспечивает организацию диалога между абонентами сети, т. е. управление очередностью передачи данных, их приоритетом, процедурой восстановления и т. д.

*Четвертый (транспортный) уровень* создает связь между нижней и верхней группами уровней системы. Транспортный уровень обеспечивает сквозную передачу данных между абонентами используемой сети с заданным качеством обслуживания, которое является составным параметром, определяющим характеристики взаимодействия абонентов: максимальное время установления соединения, пропускная способность, время задержки сигнала, вероятность ошибки при передаче сообщений и т. п. Реализует процедуры и протоколы соединения пользователей сети через базовую (магистральную) сеть. На этом уровне возможны стандартное сопряжение различных систем с сетью и организация обмена сообщениями между сетью и узлами или системами сети. Создает связь между нижней и верхней группами уровней: обеспечивает сквозной обмен информацией между системами.

*Пятый (сеансовый) уровень* обеспечивает организацию сеансов связи на период взаимодействия сетевых узлов. На этом уровне по запросам в сети создаются порты (*Port* — интерфейс для подключения линий приема-передачи данных к сетевому устройству) для приема и передачи сообщений и организуются соединения — логические каналы.

*Шестой (представительный)* уровень согласует форму представления информации (изображение, распечатка, строка символов и т. д.). Представительный уровень управляет и преобразует синтаксис блоков данных, которыми обмениваются оконечные пользователи (коды, форматы данных, сжатие данных, машинные языки и т. п.).

*Седьмой уровень — прикладной*. Его главная задача — предоставить уже переработанную (принятую) информацию. С этим обычно справляется системное и пользовательское прикладное программное обеспечение. Прикладной уровень обеспечивает интерфейс с прикладным процессом и служит для выполнения всех информационно-вычислительных процессов, предоставляемых пользователю через транспортную сеть: электронная почта, телетекст, факс, пакетная передача речевых сообщений и др.

Четыре нижних уровня (физический, канальный, сетевой и транспортный) предоставляют сетевые услуги, три верхних (сеансовый, представительный и прикладной) — услуги самим оконечным пользователям. Как правило, соединение между АТС разных типов сетей осуществляется на нижних уровнях.

Понятие о протоколах. Правила взаимодействия объектов одного уровня, называемые протоколами, определяют логическое взаимодействие объектов. В эталонной модели ВОС принята концепция, в соответствии с которой взаимодействие объектов одного уровня обеспечивается предоставлением ему

услуг смежным нижним уровнем. Правила взаимодействия объектов смежных уровней в одной системе сети, а также межсетевого обмена информацией называют *интерфейсами* ВОС. В настоящее время наиболее проработаны основные стандарты МСЭ и МОС/ISO и на них ориентированы проекты крупных сетей.

### Системы связи

Напомним, что системы связи (в том числе и радиосвязи) предназначены для передачи *сообщений* из одной точки пространства (местоположения) в другую через *канал связи*. При этом канал связи должен обладать определенными свойствами — в частности, пропускать лишь заданную полосу частот, иметь одну или несколько несущих и т. д. Для решения задачи связи приходится осуществлять целый ряд преобразований сообщений (информации). Системы связи включают в себя все основные устройства, применяемые в большинстве радиотехнических систем передачи информации. В теории связи принято следующее определение: *системой связи называется совокупность технических средств для передачи сообщений от источника к потребителю*. Иногда в понятие система связи включают такие элементы, как источник сообщений и потребитель.

По виду передаваемых сообщений различают системы: передачи речи (телефония); передачи данных, радиовещание, передачи подвижных изображений (телевидение); передачи текста (телеграфия); передачи неподвижных изображений, рисунков и текста (фототелеграфия); телеизмерения; телеуправления.

По назначению телефонные, радиовещательные и телевизионные системы делят на *вещательные*, отличающиеся высокой степенью качества воспроизведения сообщений, и *профессиональные*, имеющие специальное применение (служебная связь, промышленное телевидение и т. д.).

В системе телеизмерения физическая величина, подлежащая измерению (напряжение, температура, скорость и т. п.), с помощью датчиков преобразуется в электрический сигнал (этот сигнал называют *первичным*), поступающий на передающее устройство. На приемном конце переданную физическую величину или ее изменения выделяют из сигнала и наблюдают или регистрируют с помощью записывающих приборов. В системе телеуправления осуществляется передача команд для автоматического выполнения определенных действий. Часто эти команды формируют по результатам измерений, переданных телеметрической системой.

Повсеместное внедрение высокоэффективных и скоростных компьютеров привело к необходимости быстрого развития систем передачи данных, обеспечивающих обмен информацией между вычислительными средствами и объектами автоматизированных систем управления и измерения. Этот вид связи по сравнению с другими отличается более высокими требованиями к скорости и верности передачи информации.

**Назначение систем связи**. Рассмотрим общие принципы построения систем радиосвязи (радиоканала). Достаточно условно все существующие системы радиосвязи можно разделить на два больших класса: симплексные и дуплексные системы связи.

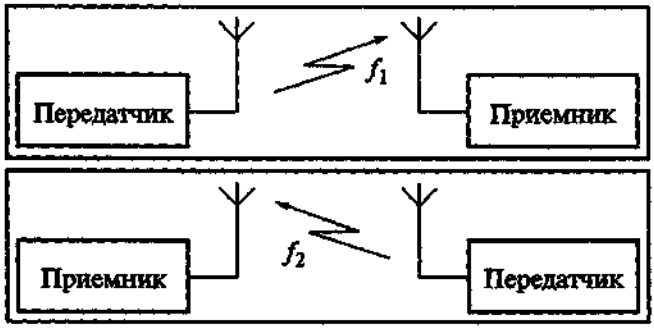



**Рис. 1.18.** Структурная схема организации дуплексной связи

Под *симплексной связью* (*simplex* — односторонний; связь «один-ко-всем») понимают связь между двумя пунктами, при которой в каждом из них передача и прием сообщений ведутся поочередно на одной несущей частоте. Часто симплексную связь используют для передачи информации только в одном направлении, например радиовещание, телевидение, оповещение и т. д.

*Дуплексная связь* (*duplex* — двусторонний; связь «один-на-один») — двусторонняя связь между двумя пунктами, при которой передача и прием сообщений осуществляют одновременно на разных несущих частотах (рис. 1.18). Сейчас применяют такую разновидность симплексной радиосвязи, как *полудуплексная* (*half-duplex*) *связь* или *двухчастотный симплекс*, когда система связи обеспечивает поочередно передачу и прием информации на двух разных несущих частотах с использованием ретрансляторов.

Отметим, что ретранслятор (от лат. *translator* — переносчик) — радиотехническое устройство, используемое как промежуточный приемопередающий пункт линии радиосвязи.

По числу используемых каналов различают *одноканальные* и *многоканальные* системы связи (системы передачи информации). Об одноканальных системах связи уже в принципе и говорилось.

Система связи называется *многоканальной*, если она способна обеспечить передачу нескольких сообщений по одной общей линии связи (каналу). Основная задача многоканальных систем связи — одновременная передача сообщений от многих источников, т. е. увеличение пропускной способности (часто используется термин «емкость»). Повышение эффективности использования канала связи достигается путем применения разных методов уплотнения каналов связи, за счет сокращения избыточности сообщений и организации так называемого *многоканального* и *многостанционного доступа* абонентов.

Для увеличения пропускной способности большинства систем связи применяют временно́е и частотное уплотнение (*multiplexing*; от лат. *multiplex* — сложный, многократный) сигналов (рис. 1.19).

Амплитудная, частотная и фазовая модуляция несущих колебаний позволяет строить многоканальные радиоэлектронные системы с частотным уплотнени-

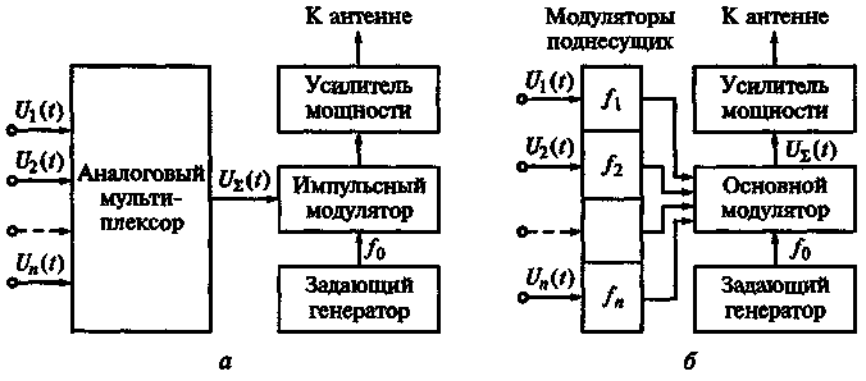



**Рис. 1.19.** Структурные схемы модуляторов систем связи с уплотнением: *а* — временны́м; *б* — частотным

ем (разделением) каналов (ЧРК), обусловленным использованием несущих колебаний с различными частотами. Достоинством системы с ЧРК является сравнительная простота и возможность передачи весьма широкополосных сообщений, например телевизионных.

Импульсная модуляция несущего колебания дает возможность разрабатывать многоканальные радиотехнические системы связи с временны́м уплотнением (разделением) каналов (ВРК), обладающие заметными преимуществами перед системами связи с ЧРК. К этим достоинствам относится высокая точность передачи сигналов (лучшая помехозащищенность) и возможность передавать совместно сообщения нескольких каналов в одном частотном диапазоне, поскольку сообщению каждого канала будет соответствовать своя последовательность импульсов, не перекрывающаяся с последовательностью импульсов сообщения другого канала.

При временно́м уплотнении, благодаря тому, что сигналы передают не непрерывно, а только их отсчетами (выборками) в очень короткие временны́е интервалы, на одной несущей частоте можно передавать ряд различных сигналов. Для этого разные сигналы $U_1(t)$, $U_2(t)$, ..., $U_n(t)$, отражающие группу из $n$ передаваемых сообщений, подают на аналоговый мультиплексор (селектор или аналоговый коммутатор) (рис. 1.19, *а*). Суммарные сигналы аналогового мультиплексора $U_\Sigma(t)$ с помощью импульсного модулятора и задающего генератора переносят на частоту $f_0$ и через усилитель мощности подводят к передающей антенне.

Традиционно во многих радиотехнических системах передачи информации широкое применение находит частотное уплотнение сигналов, осуществляемое предварительно (перед основной модуляцией) дополнительной модуляцией на так называемых *поднесущих частотах* (предварительных; от англ. — *subcarrier frequency*) — $f_1, f_2, ..., f_n$ (рис. 1.19, *б*). Поднесущие частоты значительно превышают частоту передаваемого сигнала, но во много раз меньше несущей частоты. При частотном уплотнении передаваемые сигналы предварительно поступают на модуляторы поднесущих частот, где осуществляется амплитудная, частотная, фазовая или другие виды модуляции.

Необходимые элементы модуляторов поднесущих частот — полосовые фильтры (на рис. 1.19, *б* не показаны), настроенные на поднесущие частоты и подавляющие спектральные составляющие соседних каналов. Затем промодулированные сигналы с поднесущими частотами подают на основной модулятор, работающий на основной несущей частоте $f_0$, и в виде суммарного сигнала $U_\Sigma(t)$ через антенну излучают в пространство.

Системы радиосвязи принято делить на наземные и спутнико-космические.

В наземных системах радиосвязи радиоволны распространяются в пределах земной атмосферы. Такие системы служат для обеспечения связи с самолетами, кораблями, наземным транспортом и другими объектами. Они осуществляют персональную радиосвязь в рамках сотовой, транкинговой и иных видов связи.

Особенностью спутнико-космических систем радиосвязи является наличие в их составе *искусственных спутников Земли* (ИСЗ), на которых располагают ретрансляторы радиосигналов. В целом система состоит из двух основных частей, или сегментов: наземного и космического. С помощью систем космической радиосвязи, работающих, как правило, в диапазоне сверхвысоких частот, передают огромные объемы сообщений: трансляция множества телевизионных каналов, компьютерных данных, телефонных, телефаксных и иных сообщений.

Аналоговые системы радиосвязи. Упрощенная структурная схема канала аналоговой (с непрерывными сигналами) системы радиосвязи (радиоканала) с так называемой *амплитудной модуляцией* (АМ; от англ. — *amplitude modulation, AM*) несущего колебания представлена на рис. 1.20.

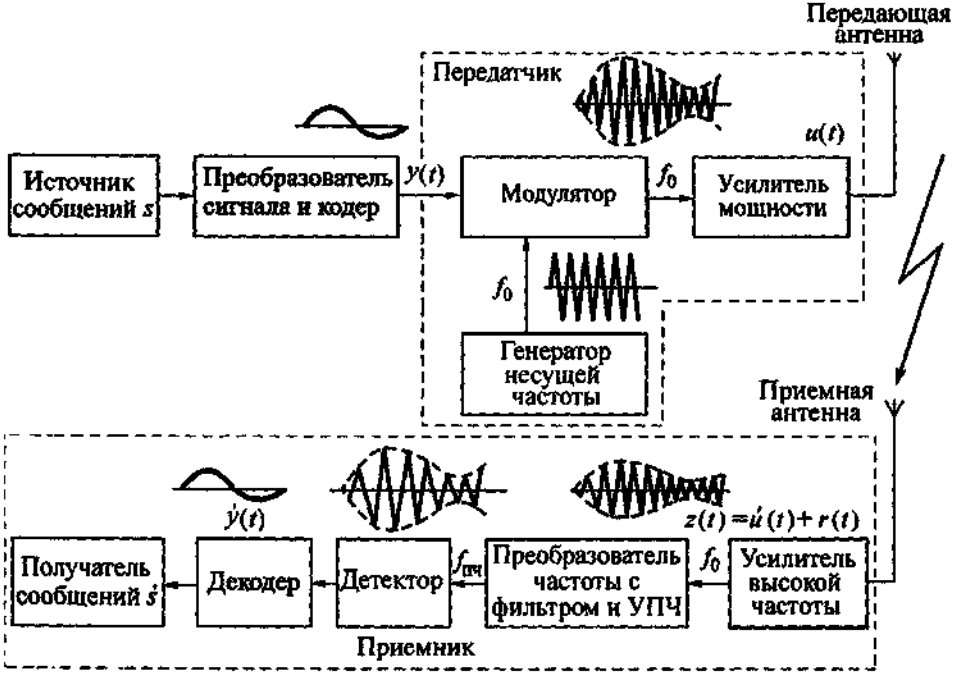



**Рис. 1.20.** Упрощенная структурная схема канала аналоговой системы радиосвязи

На структурной схеме для наглядности упрощенно показаны эпюры сигналов в некоторых ее характерных точках.

Одним из важных звеньев любой системы связи является *источник сообщений*, подлежащих передаче. Источник сообщений (источник информации; *information source*) может быть *аналоговым* или *дискретным*. Выход аналогового источника может иметь любое значение из непрерывного диапазона амплитуд, тогда как выход источника дискретной информации — значения из конечного множества амплитуд. При использовании цифровой (дискретной) связи источники аналоговой информации преобразуются в источники цифровой информации посредством *дискретизации*, *квантования* и *оцифровки*. Многие устройства для ввода данных в компьютер осуществляют подобные операции. Например, звуковая карта компьютера оцифровывает сигнал с микрофона, сканер дискретизирует сигнал, поступающий с фотоэлемента, и т. д.

В общем случае исходное сообщение $s = s(t)$ не является электрическим, может иметь любую физическую природу (подвижное изображение, звуковое колебание и т. п.), и поэтому его необходимо преобразовать в электрический (первичный) сигнал $y(t)$ с помощью *электрофизического преобразователя сигнала* (ЭФПС), проще *преобразователя сигнала*, который часто совмещают с *кодирующим устройством — кодером*. Источником сообщения при телефонной передаче является говорящий; при телевизионной — передаваемое изображение и т. д. При передаче речи и музыки преобразователем сигнала и кодером служит микрофон; при передаче изображения — передающие телевизионные трубки, или специальные матрицы. В телеграфии при преобразовании сигнала последовательность элементов письменного сообщения (букв) с помощью телеграфного аппарата заменяется последовательностью кодовых символов (0, 1 или точка, тире), которая одновременно преобразуется в последовательность электрических импульсов постоянного тока разной длительности, полярности и т. д.

Если рассматривать простейший, телефонный проводной канал связи, то в нем линия связи, по которой передается сообщение, — телефонный провод. Часть трубки, которую мы подносим к уху, выполняет роль *декодера* и преобразователя сигнала (электрические сигналы снова преобразуются в звуки). Сообщение поступает в «принимающее устройство» — ухо человека на другом конце провода. Канал включает в себя телефонные аппараты (устройства), провода (предметы) и аппаратуру АТС (устройства). Особенностью этого информационного канала является то обстоятельство, что при поступлении в него сообщение, представленное в виде звуковых волн, преобразуется в электрические колебания и затем передается.

В последнее время в структурных (и принципиальных) схемах радиоканала источник сообщения и преобразователь сигнала объединяют в одно звено, называемое *источником первичных сообщений*, а кодер включают отдельно.

Как уже отмечалось, передаваемый (первичный) сигнал является низкочастотным (*baseband signal*). Однако термин «низкочастотный» здесь достаточно условен, в частности, телевизионный сигнал имеет спектр с полосой порядка 0...6 МГц. Поэтому в ряде случаев первичный сигнал непосредственно передают

по линии связи. Так поступают, например, в обычной городской телефонной сети. Для передачи на большие расстояния (с помощью кабеля, оптического волокна или радиоканала) первичный сигнал преобразуют в высокочастотный.

Необходимым условием преобразования сообщения в электрический сигнал является то, что оно должно быть обратимым. В этом случае по выходному сигналу можно восстановить входной первичный сигнал, т. е. получить всю информацию, содержащуюся в переданном сообщении. В противном случае часть информации будет потеряна при передаче сигнала.

Передающее устройство включает в себя, кроме преобразователя сигнала, *передатчик* (содержащий *модулятор, генератор несущей частоты* и *усилитель мощности*) и *передающую антенну*. Для передачи сообщения отражающий его сигнал необходимо предварительно ввести в несущее высокочастотное электромагнитное колебание. Это осуществляется в модуляторе передатчика. Несущее колебание вырабатывается *генератором несущей частоты*. В последние годы в качестве генераторов несущих частот используют так называемые *синтезаторы частот*. Кстати, отметим, что генераторы формируют когерентное несущее, имеющее стабильные параметры. В частности, положение начальной фазы их колебаний в течение длительного времени настолько постоянно, что приемные устройства используют это свойство и извлекают из него пользу. Без свойства когерентности несущего колебания не могли бы успешно функционировать мощные радиотехнические системы связи.

Напомним, что процесс, в результате которого один или несколько параметров несущего колебания изменяется по закону передаваемого сообщения, называется *модуляцией*. Модулированное высокочастотное колебание относят к *вторичным сигналам* и называют *радиосигналом*. Один из известных отечественных ученых в области радиотехники, академик А.А. Харкевич (1904–1965) так образно оценил значение модуляции при передаче информации: «Излучение радиоволн без модуляции подобно чистой странице; модулированное колебание подобно странице, на которой нанесены знаки и буквы».

Образно говоря, модулятор (и кодирующее устройство — кодер) «записывает» передаваемое сообщение в изменениях параметров несущего колебания (радиоволны). Следовательно, совместное действие модулятора и генератора несущей частоты позволяет превратить передаваемые низкочастотные процессы — сообщения (первичные сигналы) в высокочастотные колебания, которые излучаются антенной и распространяются в свободном пространстве. Если перестраивать генератор несущей частоты или использовать синтезатор частот, то можно найти диапазон частот, свободный от посторонних радиосигналов. Без процесса модуляции низкочастотные сигналы, отражающие сообщения различных отправителей, мешали бы друг другу из-за перекрытия их спектров; модуляция же позволяет отвести каждому отправителю сообщений индивидуальную полосу частот в соответствующем диапазоне волн.



А.А. Харкевич

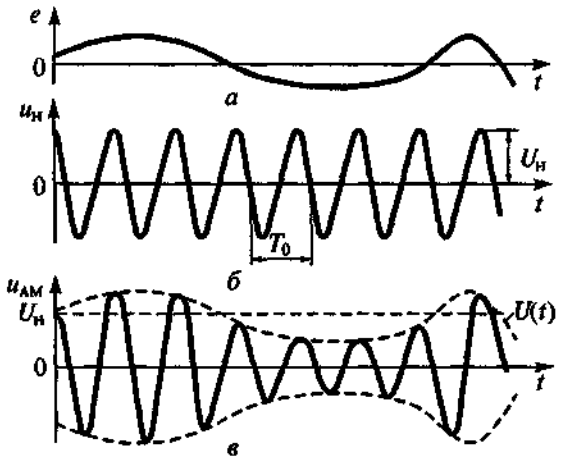



**Рис. 1.21.** Временны́е диаграммы к процессу амплитудной модуляции:
*а* — модулирующий сигнал; *б* — несущее колебание; *в* — АМ-сигнал

Наиболее простым и распространенным способом введения передаваемого сообщения (модулирующего сигнала) в высокочастотное несущее колебание служит амплитудная модуляция (рис. 1.21). При амплитудной модуляции по закону передаваемого сообщения изменяется только амплитуда несущего колебания при неизменных остальных его параметрах.

Заметим, что и природа использует принцип модуляции (в данном случае амплитудной) при издании звука человеком. Голосовыми связками возбуждают звуки относительно высоких частот (несущее колебание), которые модулируют мускулами губ и полости рта (модулирующий сигнал), делающими не более 10 движений в 1 с. Строение рта человека таково, что он не может мускулами губ непосредственно издавать звуковые колебания с частотами выше 10 Гц.

Положим, что передаваемое сообщение отражено непрерывным модулирующим сигналом (рис. 1.21, *а*), представляющим собой некую непрерывную функцию времени $e(t)$. В качестве несущего колебания в радиотехнике практически всегда используют высокостабильный высокочастотный гармонический сигнал (рис. 1.21, *б*)

$$u_н(t) = U_н\cos(\omega_0 t + \varphi_0) = U_н\cos\psi(t), \qquad (1.11)$$

где $U_н$ — амплитуда (максимальная высота синусоиды; заметим, что *амплитудой* сигнала называют модуль наибольшего его отклонения от нуля, следовательно, амплитуда всегда положительна) в отсутствие модуляции (амплитуда несущего колебания); $\omega_0$ — угловая (круговая) частота; $\varphi_0$ — начальная фаза; $\psi(t) = \omega_0 t + \varphi_0$ — полная (текущая или мгновенная) фаза.

Круговая частота $\omega_0$, период колебаний $T_0$ и циклическая частота $f_0 = 1/T_0$ связаны между собой соотношением

$$\omega_0 = 2\pi/T_0 = 2\pi f_0. \qquad (1.12)$$

При амплитудной модуляции огибающая амплитудно-модулированного сигнала (АМ-сигнала) $U_н(t)$ совпадает по форме с модулирующим сигналом, (рис. 1.21, *в*). Поэтому выражение (1.11) примет вид:

$$u_{АМ}(t) = U_н(t)\cos(\omega_0 t + \varphi_0) = [U_н + k_A e(t)]\cos(\omega_0 t + \varphi_0). \qquad (1.13)$$

Здесь $k_A$ — безразмерный коэффициент пропорциональности, такой, что всегда $U_н(t) \geq 0$. Как уже отмечалось, передача и прием модулированных электромагнитных колебаний (радиосигналов) осуществляется с помощью антенн. Полученный на выходе модулятора передающего устройства радиосигнал усиливается в усилителе мощности до необходимого уровня и поступает в передающую антенну и излучается в окружающую среду (поступает в линию связи).

Высокочастотные сигналы, улавливаемые приемной антенной, поступают в *приемник*, структура которого является зеркальным отражением структуры передатчика — сигнал проходит через блоки, в обратном порядке осуществляющие преобразования, т. е. «обратные» по отношению к тем, что проводились в передатчике. Приемная антенна улавливает очень малую долю энергии, излученную передающей антенной. Поэтому принятые антенной модулированные колебания, после отфильтровывания их с помощью селективных (избирательных) цепей от помех и сигналов других радиостанций (цепи фильтрации на рисунке для упрощения не показаны), подают на *усилитель высокой частоты* (УВЧ), обычно являющимся малошумящим усилителем (*low-noise amplifier — LNA*). УВЧ, помимо усиления, также селектирует (выделяет) полезный сигнал от совокупности многих радиосигналов и помех, одновременно поступающих на приемную антенну.

Здесь важно отметить, что на всех этапах прохождения сигнала, несущего информацию от источника к приемнику, на сигнал воздействуют помехи (шумы). Пусть требуется передать некоторое сообщение $s$, которое преобразуется преобразователем сигнала в электрическое колебание $y(t)$, а на выходе передающего устройства принимает вид сигнала $u(t)$. В процессе передачи сигнал $u(t)$ практически всегда искажается, поскольку на него накладываются помехи $r(t)$. В результате воздействия помех количество информации, поступающей к приемнику, уменьшается. В частности, при передаче речи в результате воздействия помех, кроме переданного слова, сигнал $u(t)$ содержит шум, который может сделать слово неразборчивым. Приемное устройство обрабатывает принятое колебание $z(t) = \acute{u}(t) + r(t)$, представляющее собой сумму пришедшего искаженного сигнала $\acute{u}(t)$ и помехи $r(t)$, и восстанавливает по нему, после преобразования и детектирования из сигнала $\acute{y}(t)$, сообщение $\acute{s}$, которое с некоторой погрешностью отображает переданное сообщение $s$. Другими словами, приемник должен на основе анализа колебания $z(t)$ определить, какое из возможных сообщений передавалось. Поэтому приемное устройство является одним из наиболее ответственных и сложных элементов системы связи.

Следует помнить, что в системах радиосвязи с помощью передающих антенн из выходных колебаний передатчика формируется электромагнитная волна (электромагнитное поле) — пространственно-временно́й сигнал $u(t, \mathbf{r})$, который

зависит не только от времени $t$, но и от многих пространственных координат точки приема $\mathbf{r} = x, y, z$. Полезный пространственно-временно́й сигнал, но уже совместно с помехами $z(t, \mathbf{r}) = u(t, \mathbf{r}) + r(t, \mathbf{r})$, возникает в месте приема (на входе приемного устройства). Обычно это поле сначала посредством приемной антенны превращается в чисто временно́й сигнал $z(t)$, который в дальнейшем подвергается уже временно́й обработке. Вопросы формирования и обработки пространственно-временны́х сигналов в настоящей книге не рассматриваются, т. е. будем считать, что устройства преобразования *временно́й сигнал – электромагнитное поле* на передаче и *электромагнитное поле – временно́й сигнал* на приеме включены внутри заданной «линии связи». Эти вопросы рассматриваются в специальных дисциплинах.

Усиление радиосигналов осуществляется и в последующих каскадах приемника. При этом непосредственное усиление сигнала используется крайне редко. Дело в том, что при переходе на прием другой станции требуется перестраивать избирательный усилитель, сохраняя высокую частотную селекцию (избирательность), т. е. выделять полезный сигнал из других сигналов и помех. Эта сама по себе нелегкая проблема становится чрезвычайно сложной, когда требуется большое усиление и, следовательно, применение нескольких усилительных каскадов. Задача существенно упрощается, если в приемнике используется *преобразователь* (здесь *преобразователь частоты*), в котором разные несущие частоты сигналов, поступающих на вход его *смесителя*, преобразуются (точнее, переносятся) с помощью вспомогательного многочастотного генератора (гетеродина) в сигналы с одинаковой, более низкой несущей частотой $f_{\text{пч}}$, называемой *промежуточной*. На выходе преобразователя частоты включают фильтр, который выделяет полезный сигнал. При этом дальнейшее усиление информационных сигналов будет происходить на одной частоте без перестройки схем в *усилителе промежуточной частоты* (УПЧ), который производит основное усиление в приемнике и улучшает селекцию по частоте полезного сигнала. Такой приемник называется *супергетеродинным* (см. гл. 9).

*Детектор* (от лат. *detectio* — обнаруживать, выделять) или *демодулятор* осуществляет процесс, обратный модуляции, — выделяет из принятого, усиленного и преобразованного высокочастотного модулированного колебания передаваемый сигнал. Основная задача детектирования — по возможности полное восстановление информации, содержащейся в модулирующем сигнале (искаженном при передаче помехой) $\dot{y}(t)$, поступившем с преобразователя частоты. Поэтому главное требование к детектору — точное воспроизведение формы передаваемого сигнала, чтобы он поступал к получателю неискаженным.

*Получатель сообщений* приемника преобразует низкочастотный электрический сигнал детектора в форму информации, удобную для получателя. Как правило, источник первичных сообщений и оконечное устройство в структурную схему системы радиосвязи не включают.

Степень закрытости сообщения. В последние годы даже в аналоговых радиоканалах систем передачи информации применяют кодирование сообщений. Для этого в структуры схем передатчика и приемника сообщений

включают кодер (простейшим является АЦП), а в оконечное устройство приемника — декодер, устанавливая их соответственно перед модулятором и после детектора. С помощью такого вида кодирования и модуляции источник сообщений *согласуют* с каналом связи.

Идея кодирования возникла давно и преследовала в основном цели быстроты и секретности передачи информации. В настоящее время кодирование используют и для создания условий, обеспечивающих надежную и экономичную передачу сообщений по каналам связи. В общем случае считают, что кодирование определяет закон построения дискретного управляющего сигнала, а модуляция — вид формируемого радиосигнала, который передается по каналу связи.

Часто кодирование применяют и для обеспечения секретности (закрытости) передаваемого сообщения, т. е. его правильного понимания только тем получателем, которому оно адресовано. Такое требование обязательно, например, в военном деле: противник не должен перехватить и рассекретить передаваемую информацию. Обычно требуется закрывать информацию экономического характера, чтобы обезопасить себя от конкурентов. При передаче банковской информации также надо обеспечить ее конфиденциальность. Секретность передаваемого сообщения обеспечивают не только кодированием, но и шифрованием, которое используют для обеспечения секретности связи, предотвращает понимание сообщения несанкционированным пользователем и введение в систему ложных сообщений. Шифрование обеспечивает более сложный способ засекречивания сообщения согласно определенному алгоритму и ключу.

К общим признакам аналоговых каналов связи относятся следующие. Во-первых, большинство таких каналов можно считать *линейными*: в них выходной сигнал является суммой откликов отдельных входных сигналов и помех (применим принцип суперпозиции), а продукты нелинейных преобразований в канале малы по сравнению с выходными сигналами. Во-вторых, на выходе канала даже в отсутствие полезного сигнала всегда имеются помехи. В-третьих, сигнал при передаче по каналу претерпевает задержку по времени и затухание по уровню. И наконец, в реальном аналоговом канале всегда имеют место искажения сигнала, обусловленные несовершенством характеристик канала и нередко изменениями его параметров во времени.

Отметим, что структурная схема радиотехнической системы для локации объектов отличается от приведенной на рис. 1.20 размещением источника первичных сообщений и детектора с оконечным устройством приемника в одном пункте с передающим устройством. Кроме того, сигнал в этом случае будет определяться эффектом рассеяния радиоволн от облучаемого ими объекта. Что же касается других видов структурных схем радиосистем, то они могут быть получены на основе рис. 1.20 путем комбинации источника первичных сообщений и детектора с оконечным устройством приемника в разных местах базирования и в том числе при работе на нескольких длинах волн.

Цифровые (дискретные) системы радиосвязи (*digital communication system* — DCS). Вплоть до середины 60-х гг. XX в. (в Российской Федерации и до середины 70-х гг. XX в.) все существовавшие системы связи практи-

чески были аналоговыми. До этого цифровых систем связи практически не существовало, несмотря на то, что *импульсно-кодовая модуляция* (ИКМ; от англ. — *Pulse code modulation* — PCM) была известна с 1938 г. (изобретателем ИКМ является французский инженер А. Ривс). Первое упоминание о разработанной фирмой «Белл» реальной дискретной системе связи с ИКМ появилось в 1947 г.

Под *дискретной системой связи* понимают систему, в которой и передаваемый и принимаемый сигналы являются последовательностями дискретных символов. Типичным примером такой системы является телеграфия, в которой и сообщение, и сигнал являются последовательностями точек, тире и промежутков между ними. Импульсные методы модуляции интенсивно развивались с начала 40-х гг. XX в. в связи с развитием радиолокации, однако ИКМ не находила широкого практического применения ввиду громоздкости цифрового оборудования, вплоть до появления в 1959 г. компьютеров второго поколения, использующих в качестве активной элементной базы транзисторы.

Чаще всего кодирование здесь сводится к записи номера уровня в двоичной системе счисления, т. е. в случае применения ИКМ дискретные значения непрерывного сигнала передаются в виде кодовых комбинаций «1» и «0» или, иначе говоря, двоичным кодом. В дальнейшем будем рассматривать в основном цифровые системы, в которых непрерывное сообщение преобразовано в последовательность кодовых комбинаций, составленных из двоичных символов. Начало использования цифровых технологий в сетях передачи данных было неразрывно связано с ИКМ, а именно с системами цифровой телефонии на основе кабельных (медных) сетей связи, применяемых для передачи голоса.

Стремительное развитие микропроцессорной техники и технологии, зародившейся в 1969 г. с появлением первого микропроцессора компании Intel (микропроцессор создал американский инженер М. Хофф), сделало возможным реальное внедрение цифровой техники в системы связи и привело к широкому распространению и развитию компьютерных сетей на основе ИКМ. Итак, в цифровых (дискретных, импульсных) системах передачи информации энергия полезного сигнала излучается не непрерывно (как при синусоидальном переносчике — гармонической несущей), а в виде коротких импульсов. Это позволяет при той же общей энергии излучения, что и при непрерывном переносчике, увеличить пиковую (максимальную) мощность в соответствующем импульсе и тем самым повысить помехоустойчивость приема. В цифровых системах связи задачей приемника является не точное воспроизведение переданного сигнала, а определение на основе искаженного шумами сигнала, какой именно сигнал из конечного набора был послан передатчиком.

В качестве переносчика первичного сигнала $e(t)$ в цифровых системах радиосвязи используют периодическую последовательность видео- и радиоимпульсов (см. гл. 2). Упрощенная структурная схема современного радиоканала цифровой (дискретной) системы связи показана на рис. 1.22, где для наглядности и лучшего понимания изображены упрощенные эпюры сигналов в ряде ее характерных точек. Кстати, в подобных системах вместо генераторов несущих частот используют синтезаторы несущих частот, что отражено на рис. 1.22.

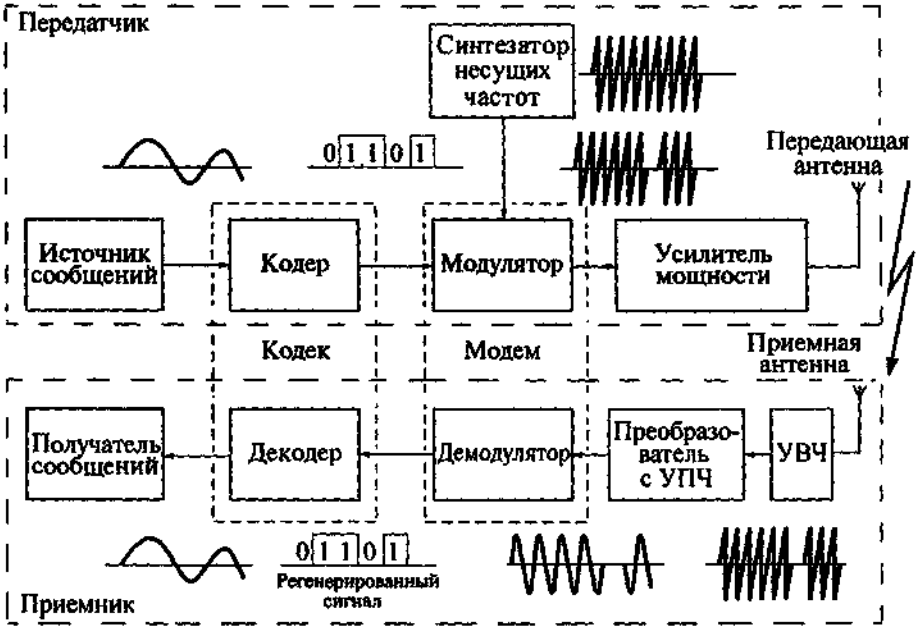



**Рис. 1.22.** Упрощенная структурная схема радиоканала цифровой системы связи

Как уже отмечалось, непрерывные сообщения можно передавать по дискретным (цифровым) радиоканалам. Для этого их преобразуют в цифровую форму с помощью операций дискретизации по времени, квантования по уровню и кодирования или оцифровки.

Методы дискретизации и квантования в последнее время стали называть *форматированием* и *кодированием источника*.

Процесс *форматирования* (*formatting*) преобразовывает исходную информацию в цифровую форму, обеспечивая таким образом совместимость информации и функций обработки сигналов с системой цифровой связи. С этого момента передаваемый первичный сигнал представляется цифровым кодом или *потоком битов* — некой последовательностью стандартных импульсов («единиц») и пауз («нулей») обычно одинаковой длительности. Помимо форматирования применяется сжатие данных, и этот процесс относят к кодированию источника.

*Кодирование источника* (*source coding* — преобразование аналогового сигнала в цифровой для аналоговых источников) и удаление избыточности в передаваемой информации. В передающем устройстве цифровой системы радиосвязи кодирование передаваемого сигнала выполняется современной цифровой логической микросхемой, называемой *кодером*.

Следующий этап передачи сообщений — *помехоустойчивое кодирование*. Идея помехоустойчивого кодирования заключается в том, что к передаваемому сообщению добавляют избыточные неинформативные символы. Эти избыточные символы связаны определенными математическими соотношениями с сим-

волами сообщения. Цель внесения избыточности — сделать возможные кодированные сообщения как можно больше отличающимися друг от друга для обеспечения исправления на приемной стороне всех или некоторых ошибок, возникших в процессе передачи. После применения помехоустойчивого кода сообщение поступает в модулятор. Цифровое сообщение в модуляторе радиопередатчика преобразуют в аналоговый модулированный сигнал, занимающий заданную полосу частот. Для этого несущее колебание модулируют полученной в кодере импульсной последовательностью. Чаще всего в цифровых системах связи используют *импульсно-кодовую модуляцию*.

Итак, в цифровой системе передачи информации превращение сообщения в радиосигнал осуществляется тремя операциями: преобразованием, кодированием и модуляцией (в аналоговой системе двумя — преобразованием и модуляцией). Отметим, что кодирование определяет математическую сторону, а модуляция — физическую сторону превращения сообщения в радиосигнал. По существу кодирование представляет собой преобразование передаваемого сообщения в последовательность кодовых символов (см., например, код 01101, на рис. 1.22), а модуляция — преобразование этих символов в сигналы, пригодные для передачи по цифровому каналу. Фактически с помощью кодирования и модуляции источник сообщений согласуется с каналом связи.

В приемнике после усиления на радиочастоте в УВЧ и преобразования частоты и усиления в блоке «Преобразователь с УПЧ», из сигнала промежуточной частоты (принятого вторичного сигнала) с помощью демодулятора извлекается и регенерируется (восстанавливается) последовательность кодовых символов (первичный сигнал). Затем производится декодирование этих символов в *декодере*. Процесс *декодирования* состоит в восстановлении переданного сообщения по принимаемым кодовым символам. С выхода декодера восстановленный аналоговый сигнал поступает к *получателю сообщений.*

В современных цифровых системах передачи информации используют два специфических аналого-цифровых устройства (микросхемы) — кодеки и модемы. *Кодеком* называют пару преобразователей *кодер-декодер* (КОдер и ДЕКодер; по первым буквам), а *модемом* — пару преобразователей *модулятор-демодулятор* (МОдулятор и ДЕМодулятор). Термин «модем» (*modem*) часто объединяет несколько этапов обработки сигналов, показанных на рис. 1.22, в этом случае модем можно представлять как «мозг» системы связи (тогда передатчик и приемник можно рассматривать как «мускулы» системы связи).

Модемы выполняют набор различных функций и в зависимости от принципов их реализации разделяются на проводные или телефонные модемы, сотовые модемы, пакетные радиомодемы, связные высокочастотные радиомодемы, цифровые модемы, факс-модемы и пр. Проводные модемы включаются в системах связи между телефонной сетью общего пользования и управляющим компьютером. На рис. 1.23 приведена упрощенная структурная схема одного из типов проводных модемов. Подобный модем может работать в дуплексном режиме, если применена четырехпроводная линия связи, или полудуплексном режиме, если используется двухпроводная линия обычной городской телефонной связи.

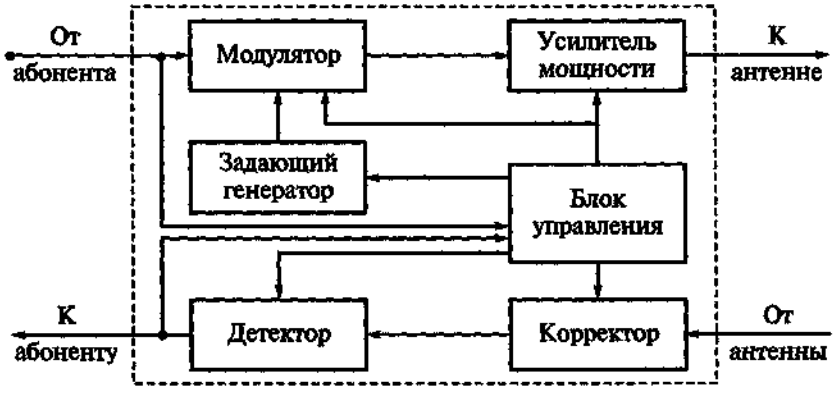



**Рис. 1.23.** Упрощенная структурная схема проводного модема

В режиме приема данных с линии связи в модеме с помощью корректора устраняются искажения передаваемых сигналов, возникающих чаще всего в результате ограниченной полосы пропускания телефонного канала и неравномерности *амплитудно-частотной характеристики* (АЧХ) и нелинейности *фазочастотной характеристики* (ФЧХ). С выхода корректора сигнал подается на детектор (демодулятор), преобразующий модулированное колебание в напряжение, форма которого воспроизводит низкочастотный передаваемый сигнал аналогового или цифрового вида. В схему модема входит также блок управления. Практически по такой же структурной схеме выполняется связной высокочастотный радиомодем, использующийся в радиоканале системы связи небольшого (менее 1 км) радиуса действия.

В радиотехнических системах передачи информации используют также дискретно-непрерывные и непрерывно-дискретные каналы, на вход которых поступают дискретные сигналы, а с выхода снимают непрерывные, или наоборот. Это значит, что канал связи может быть дискретным или непрерывным независимо от характера передаваемых сообщений. Более того, в одной и той же системе связи можно выделить как дискретный, так и непрерывный канал. Все зависит от того, каким образом выбраны вход и выход канала связи.

Следует четко представлять, что декодирование и демодуляция — это не просто операции, обратные кодированию и модуляции, выполняемые над пришедшим в приемник сигналом. В результате воздействия помех и различных искажений в линиях связи принятый сигнал может существенно отличаться от передаваемого. Поэтому всегда можно высказать ряд предположений (гипотез) о том, какое сообщение передавалось. Главной задачей приемного устройства является принятие решения о том, какое из возможных сообщений действительно передавалось источником. Для принятия такого решения необходимо проанализировать пришедший сигнал. С этой целью он подвергается различным преобразованиям, которые и называют *обработкой сигнала*. Это процесс преобразования аналоговых, дискретных и цифровых сигналов, который может осуществляться соответственно аналоговыми, дискретными и цифровыми методами.

Одной из основных проблем теории связи является отыскание правил оптимальной обработки сигнала, при которой решение о переданном сообщении оказывается наиболее достоверным. Эти правила зависят от свойств канала связи и методов передачи (кодирования и модуляции). Иногда оптимальные правила обработки сигнала оказываются сложными и для упрощения аппаратуры используют другую, не оптимальную обработку.

При обработке принятый сигнал подвергается анализу с учетом всех сведений об источнике (например, о вероятностях, с которыми источник посылает то или иное сообщение), о применяемом коде и методе модуляции, а также о свойствах радиоканала. По итогам анализа обычно можно определить условные (апостериорные) вероятности возможных гипотез и на основании этих вероятностей принять решение, результат которого и поступает к получателю. Та часть приемника, которая выполняет анализ приходящего сигнала и принимает решение о переданном сообщении, представляет собой специальную систему распознавания (систему анализа сигналов и принятия решения о виде передаваемого в данном случае сигнала, в частности, какое слово было произнесено) и называется *решающей схемой*.

В системах передачи непрерывных сообщений с аналоговой модуляцией решающая схема определяет по пришедшему искаженному вторичному сигналу наиболее вероятный переданный первичный сигнал и восстанавливает его. В таких системах решающей схемой является демодулятор. Решающая схема приемника в системах передачи дискретных сообщений состоит, как правило, из двух частей: первой решающей схемы — демодулятора и второй решающей схемы — декодера. В некоторых случаях при передаче дискретных сообщений операции демодуляции и декодирования выполняет одно устройство, которое приходящую последовательность элементов сигнала преобразует сразу в последовательность символов (букв) сообщения. Такой метод приема сообщений называют *совместной демодуляцией-декодированием* или *приемом в целом*, в отличие от поэлементного приема с двумя решающими схемами. Иногда роль решающей схемы полностью или частично выполняет человек. В частности, при приеме телеграфных сигналов на слух оператор решает, какой сигнал («точка» или «тире») был передан. Он же выполняет и операцию декодирования принятого сигнала, восстанавливая по кодовым символам передаваемое сообщение. В приемниках дискретных сообщений, предназначенных для записи полученной информации, все указанные операции выполняются в автоматическом режиме.

В некоторых случаях применяют решающие схемы с двумя порогами. В простейшем случае первая решающая схема представляет собой *пороговое устройство*, работающее по принципу «да» или «нет». Если принятый элемент сигнала выше установленного порога, выдается один символ кода (например, «1»), если ниже — другой (например, «0»). При попадании уровня сигнала между двумя порогами никакого решения не принимается — вместо сомнительного элемента сигнала выдается специальный *символ стирания*. Включение подобного стирающего символа облегчает возможность правильного декодирования принятой кодовой комбинации.

Чтобы принять решение о том, какое сообщение было передано, необходимо проанализировать принятый сигнал. Для этого его в приемнике подвергают различным преобразованиям, т. е. обрабатывают сигнал. Одной из главных проблем теории связи является отыскание *правил оптимальной обработки сигнала,* при которой решение о переданном сообщении оказывается наиболее *достоверным.* Эти правила зависят от свойств системы связи и методов передачи (кодирования и модуляции) информации. В ряде случаев оптимальные правила обработки сигнала оказываются настолько сложными, что для упрощения приемной аппаратуры используют неоптимальную обработку.

Особое влияние на качество приема сигналов оказывает точность *синхронизации* переданных и принятых сигналов. Различают *тактовую синхронизацию* (определение границ единичных элементов сигнала; *clock synchronization*), *цикловую синхронизацию* (правильное разделение кодовых комбинаций), *синхронизацию несущих частот* и др. Погрешности синхронизации передаваемых и принимаемых сигналов приводят к снижению достоверности приема сообщений, а иногда и к неверному приему всего сообщения или его части. Наиболее часто в системах передачи информации применяют *цикловую синхронизацию*. Простейшим методом, позволяющим на приеме отделить одну кодовую комбинацию от другой, является *стартстопный режим передачи,* когда в начале и конце каждой кодовой комбинации передают специальные синхросигналы («старт» и «стоп»). Такой метод передачи относят к *асинхронным,* поскольку передача очередной кодовой комбинации может начинаться в любой момент времени после окончания предыдущей комбинации. При *синхронных* методах передачи информации кодовые элементы сигнала передают непрерывно через одинаковые промежутки времени. Разделение кодовых комбинаций осуществляют в этом случае с помощью цикловой синхронизации. В общем случае вопросы синхронизации представляют достаточно сложную проблему при передаче информации (например, в системах сотовой связи с временны́м разделением сигналов) и их изучают в специальных курсах по телекоммуникациям.

Между аналоговыми и дискретными радиоканалами имеется существенное отличие. Цифровые каналы менее подвержены искажению и интерференции радиоволн, чем аналоговые. Поскольку двоичные цифровые каналы дают значимый по мощности сигнал только при работе в одном из двух состояний — включенном или выключенном — возмущение должно быть достаточно большим, чтобы перевести рабочую точку канала из одного состояния в другое. Наличие всего двух состояний передаваемого сигнала при цифровой связи (как правило, это «0» и «1») облегчает его восстановление и, следовательно, предотвращает накопление в процессе передачи шумов или других возмущений. В дискретных радиотехнических системах передачи информации и связи ошибка при передаче сообщений возникает лишь тогда, когда сигнал опознается неправильно, а это возможно только при искажениях, превышающих некоторый оптимальный порог сигнала.

Аналоговые (непрерывные) сигналы, наоборот, *не являются* сигналами с двумя состояниями; они могут принимать *бесконечное множество* форм. В ана-

логовых системах любое, даже сколь угодно малое помеховое воздействие на сигнал, вызывающее искажение модулируемого параметра, всегда влечет за собой внесение соответствующей погрешности в передаваемое сообщение. Поэтому абсолютно точное восстановление переданного сообщения в них практически невозможно. При использовании цифровых технологий очень низкая частота возникновения ошибок плюс применение процедур выявления и коррекции ошибок делают возможным высокую точность сигнала. Остается только отметить, что с аналоговыми технологиями подобные процедуры невозможны.

При цифровой обработке оказалось возможным реализовать сложные алгоритмы обработки сигналов, которые не под силу аналоговой технике. Прежде всего это касается важнейшего раздела современной радиоэлектроники — создания самонастраивающихся (адаптивных) радиосистем, изменяющих обработку сигнала при изменении его параметров.

При передаче и коммутации различные типы цифровых сигналов (данные, телеграф, сотовая связь, телефон, телевидение) можно рассматривать как идентичные: ведь бит остается битом. Кроме того, для удобства коммутации и обработки цифровые сообщения могут группироваться в автономные единицы, называемые *пакетами* (*Packet*). В цифровые технологии естественным образом внедряются функции, защищающие от интерференции и подавления сигнала либо обеспечивающие шифрование или секретность.

Какова же плата за преимущества систем цифровой связи? Цифровые системы требуют более интенсивной обработки, чем аналоговые. Кроме того, для цифровых систем необходимо выделение значительной части ресурсов для синхронизации на различных уровнях. Аналоговые системы, наоборот, легче синхронизировать. Еще одним недостатком систем цифровой связи является то, что *ухудшение качества носит пороговый характер*. Если отношение сигнал/шум падает ниже некоторого порога, качество обслуживания может скачком измениться от очень хорошего до очень плохого. В аналоговых же системах ухудшение качества происходит более плавно.

Основные характеристики и параметры систем связи. Работа любой системы связи оценивается, прежде всего, точностью передачи сообщения и скоростью передачи информации. Первое определяет качество передачи, второе — количество. В реальной системе связи качество передачи связано со степенью искажений принятого сообщения. Эти искажения зависят от свойств и технического состояния системы, а также от интенсивности и характера помех. Если система связи спроектирована правильно и технически исправна, то необратимые искажения сообщений обусловлены лишь воздействием помех. В этом случае качество передачи полностью определяется помехоустойчивостью системы.

Под *помехоустойчивостью* (*Noiseproof*) понимают способность системы связи противостоять вредному влиянию помех на передачу сообщений. Поскольку действие помех проявляется в том, что принятое сообщение отличается от переданного, то количественно помехоустойчивость при заданной помехе можно характеризовать *степенью соответствия принятого сообщения пере-*

*данному.* Эта величина характеризуется общим термином — *верность*. Количественную оценку верности выбирают по-разному, в зависимости от характера сообщения и требований получателя. Можно показать (см. гл. 9), что верность передачи зависит от отношения средних мощностей сигнала и помехи (чаще *отношение сигнал/шум*; англ. — *signal-to-noise ratio* — *SNR*; обозначают обычно это отношение как *S/N*).

В работах В.А. Котельникова и К. Шеннона показано, что при выбранном критерии и заданном множестве сигналов, принимаемых при определенной помехе (так называемом *белом шуме*; *white noise*), существует предельная (потенциальная) помехоустойчивость, которая ни при каком способе приема не может быть превзойдена. Приемное устройство, реализующее потенциальную помехоустойчивость, называют *оптимальным* (наилучшим) по данному критерию. При некой интенсивности помехи вероятность ошибки приема тем меньше, чем сильнее различаются между собой сигналы, передающие разные сообщения. Проблема состоит в том, чтобы выбрать для передачи информации сильно различающиеся сигналы. Верность передачи можно повысить за счет усложнения методов модуляции-демодуляции и введения *помехоустойчивого кодирования* сообщений. Наконец, верность передачи зависит и от способа приема сообщений. Необходимо выбрать такой способ приема, который наилучшим образом реализует различие между сигналами при данном отношении сигнал/шум.

Другим важнейшим показателем системы связи является *скорость передачи информации*. Напомним, что в системах передачи дискретных сообщений скорость передачи измеряется числом передаваемых символов в единицу времени $\nu$, Бод (или бит/с).

Как уже отмечалось, количество (объем) передаваемой информации принято измерять в битах и байтах. Широко используют и более крупные производные единицы объема информации (как впрочем и объема памяти компьютеров): килобайт, мегабайт, гигабайт, а также, в последнее время, терабайт и петабайт.

При определении количества информации исторически сложилась такая ситуация, что с наименованиями «бит» и «байт» некорректно применяли (и применяют) приставки системы СИ (в соответствии с международным стандартом МЭК 60027-2 эти единицы используют, например, так: вместо $1000 = 10^3$ принято записывают $1024 = 2^{10}$):

$$1 \text{ Кбайт} = 2^{10} \text{ байт} = 1024 \text{ байт};$$

$$1 \text{ Мбайт} = 2^{20} \text{ байт} = 1024 \text{ Кбайт};$$

$$1 \text{ Гбайт} = 2^{30} \text{ байт} = 1024 \text{ Мбайт} = 1\,048\,576 \text{ Кбайт и т. д.}$$

При этом обозначение «Кбайт» принято начинать с прописной буквы в отличие от строчной буквы «к» для обозначения множителя $10^3$.

Напомним, что количество бит или байт, передаваемых в секунду, есть скорость передачи информации, которая определяется в бит/с, Бод или байт/с. При повышенной скорости передачи она определяется в кбит/с, Мбит/с, Гбит/с, Кбайт/с, Мбайт/с, Гбайт/с, КБод, МБод, ГБод и т. д.

В последние годы для оценки скорости передачи информации используют термин «битрейт» (*bitrate*), отражающий объем передаваемой информации в единицу времени (обычно в секунду). Битрейт принято использовать при измерении эффективной скорости передачи информации по каналу, т. е. скорости передачи «полезной информации» (помимо таковой по каналу можно передавать служебную информацию, например стартовые и стоповые символы при асинхронной передаче, или контрольные символы при избыточном кодировании). Битрейт выражают битами в секунду [бит/с], а также производными величинами с приставками кило-, мега- и т. д.

При использовании не двоичных, а $m$-ичных символов максимальное количество информации, которое можно передать по каналу связи, равно $\log_2 m$ [бит]. Поэтому дискретный источник сообщений может обеспечить максимальную производительность (скорость выдачи) информации [бит/с], не превышающую:

$$R_{\text{и}} = \frac{\log_2 m}{T_{\text{и}}}, \tag{1.14}$$

где $T_{\text{и}}$ — длительность одной посылки; $m$ — основание цифрового кода.

Если $m = 2$, то $R_{\text{и}} = 1/T_{\text{и}}$ и скорость передачи информации $R_{\text{и}}$ численно равна так называемой *технической скорости* $v$. При $m > 2$ возможна скорость передачи информации $R_{\text{и}} > v$. Однако нередко в цифровых системах связи скорость передачи информации $R_{\text{и}} < v$. Такой вариант бывает, когда не все посылки используются для передачи информации, например, если часть из них служит для синхронизации или для обнаружения и исправления ошибок (при использовании *корректирующего кода*).

Как будет показано в гл. 9, максимальное количество информации, которое можно передать одним двоичным символом («1» или «0»), равно 1 биту. Теоретически каждый символ, поступивший на вход канала связи, вызывает появление одного символа на выходе, так что техническая скорость на входе и выходе канала одинакова.

Системы связи еще характеризуются средней скоростью передачи информации по каналу в секунду с заданной верностью. Существует максимально возможная (предельная) скорость передачи — *пропускная способность* канала $C$ (*channel capacity*). Это фундаментальное понятие теории информации определяет потенциальные возможности системы связи. В реальной системе связи средняя скорость передачи информации всегда меньше пропускной способности.

Сжатие передаваемой информации. При передаче информации имеют место две взаимосвязанные проблемы: устранение избыточности в передаваемой информации и сжатие последней. Избыточность (*redundancy*) сообщений — это понятие теории информации. Под избыточностью будем понимать бесполезную, лишнюю при приеме часть информации, которой все равно невозможно воспользоваться, и она фактически не нужна потребителю. Сообщения практически любого источника обладают избыточностью. Дело в том, что отдельные знаки сообщения находятся в определенной статистической связи.

Так, в словах русского языка после двух подряд стоящих согласных букв более вероятна гласная, а после трех подряд согласных наверняка будет гласная. Избыточность позволяет представлять сообщения в более экономной, сжатой форме. Мера возможного сокращения сообщения без потери информации за счет статистических взаимосвязей между его элементами определяется избыточностью. Понятие «избыточность» применимо не только к сообщениям или сигналам, но и к языку в целом, коду (алфавит любого языка и слова, составленные из его букв, можно рассматривать как код). Например, избыточность европейских языков достигает 60...80 %.

Как правило, причиной появления избыточности является невосприимчивость человеческих органов к некоторой части принятой информации. Так, например, телевизионное изображение может содержать до 16 тысяч цветовых оттенков, тогда как зрение человека, чувствительное к яркости, невосприимчиво к такой громадной гамме цветов. В лучшем случае человек может различить до нескольких сотен цветовых оттенков. Поэтому часть цветовых оттенков при передаче можно исключить без ощутимой со стороны человека потери качества цветного изображения на экране телевизора. То же самое можно сказать относительно передачи по каналу связи устной речи, верхнюю частоту спектра которой можно ограничить частотой 3 400 Гц без потери смысла принятого сообщения. Еще очень простой пример — пусть по каналу связи следует передать сведения о значениях индуктивности $L$, емкости $C$ и резонансной частоты $f$ идеализированного колебательного контура. В этом случае, очевидно, можно по каналу связи передать только значения двух величин, например, индуктивности и емкости, а резонансную частоту контура на приемном конце вычислить по известной любому студенту формуле.

Устранение избыточности в исходной информации позволяет передавать или хранить меньшее число бит. В теории информации К. Шеннон доказал теорему (см. гл. 9), согласно которой для источника без избыточности при $R_и < C$ можно найти такой способ кодирования-декодирования, при котором возможна передача сообщений по каналу связи с помехами со сколь угодно малой ошибкой. Наличие избыточности в сообщении часто оказывается полезным и даже необходимым, так как позволяет обнаруживать и исправлять ошибки, т. е. повысить достоверность воспроизведения сообщения. Если избыточность в сообщении не используется для повышения достоверности передачи, то она должна быть исключена. Это достигается использованием специального статистического кодирования. При этом избыточность сигнала уменьшается по отношению к избыточности сообщения.

Универсальным показателем системы связи является *информационная эффективность* $\eta$, характеризующая использование системой пропускной способности канала $\eta = R_и/C$.

Своевременность передачи сообщений определяется допустимой задержкой, обусловленной преобразованием сообщений и сигнала, а также конечным временем распространения сигнала по каналу связи (особенно время распространения заметно в спутниковых системах связи). Она зависит от двух показа-

телей: характера и протяженности канала и длительности обработки сигнала в передающем и приемном устройствах. Скорость передачи информации и ее задержка в линиях связи являются независимыми характеристиками, почти не связанными друг с другом.

Любой канал радиосвязи можно так же, как и передаваемый сигнал, характеризовать тремя основными параметрами: временем $T_к$, в течение которого по каналу возможна передача информации, динамическим диапазоном $D_к$ и полосой пропускания канала $F_к$.

Под *динамическим диапазоном канала* понимают отношение допустимой мощности передаваемого сигнала к мощности неизбежно присутствующей в канале помехи, выраженное в децибелах.

Обобщенной характеристикой канала связи служит его *емкость* (объем)

$$V_к = T_к F_к D_к. \tag{1.15}$$

Необходимым условием неискаженной передачи по каналу сигналов с объемом $V_с$, очевидно, должно быть

$$V_с < V_к. \tag{1.16}$$

Часто преобразование первичного сигнала в высокочастотный радиосигнал и преследует цель согласования передаваемого сигнала с каналом. В простейшем случае сигнал согласуют с каналом по всем трем параметрам, т. е. добиваются выполнения условий:

$$T_с < T_к; \qquad F_с < F_к; \qquad D_с < D_к. \tag{1.17}$$

При соблюдении этих условий объем передаваемого сигнала практически полностью «вписывается» в объем канала.

В ряде случаев неравенство (1.16) может выполняться и тогда, когда одно или два из неравенств (1.17) не выполнены. Это означает, что можно производить «обмен» длительности на ширину спектра или ширину спектра на динамический диапазон и т. д. Пусть записанный на магнитофон телефонный сигнал с шириной спектра 3,4 кГц необходимо передать через канал связи, полоса пропускания которого 340 Гц. Это можно осуществить, воспроизводя сигнал со скоростью, в 5 раз меньшей той, с которой он был записан. При этом все значения частот исходного сигнала уменьшатся в 5 раз, но и во столько же раз увеличится время передачи. Принятый сигнал также записывают на магнитофон, а затем, воспроизведя его со скоростью, в 5 раз большей, можно с высокой точностью восстановить исходный сигнал. Аналогично можно передать сигнал быстрее, если полоса пропускания канала шире спектра сигнала.

Однако наибольший интерес вызывает возможность обмена динамического диапазона канала связи на полосу пропускания. Оказывается, что при внедрении импульсно-кодовых видов модуляции (см. гл. 2), можно передать сообщение с динамическим диапазоном, например 60 дБ, по каналу, в котором сигнал превышает помеху всего лишь на 30 дБ. При этом используется полоса пропускания канала в несколько раз более широкая, чем спектр сообщения.

### Линии связи

Виды линий связи, по которым передают информацию от источника к получателю, многочисленны и разнообразны. Различают каналы проводной связи (проводные, кабельные, оптоволоконные и др.) и каналы радиосвязи.

*Кабельные линии связи* являются основой магистральных сетей дальней связи; по ним осуществляется передача сигналов в диапазоне частот от десятков килогерц до сотен мегагерц.

Одним из самых совершенных систем передачи информации являются *волоконно-оптические линии связи* (ВОЛС). Информация по таким каналам передается в виде световых импульсов, посылаемых лазерным излучателем. Они позволяют в диапазоне частот 600 ... 900 ТГц ($\lambda$ = 0,5...0,3 мкм) обеспечить чрезвычайно большую пропускную способность (примерно 120 000 каналов по паре оптических волокон) и создают надежную и скрытую связь с высоким качеством передачи информации. Основными преимуществами *оптических волокон* (ОВ), или световодов, как физической среды распространения сигналов электросвязи и конструктивной основы *оптического кабеля* (ОК) являются:

- широкая полоса пропускания, позволяющая передавать сигналы электросвязи со скоростью (битрейтом) до 2,0 ... 2,5 Тбит/с и выше; например, даже при скорости 50 Мбайт/с в течение 1 с передается объем информации, приблизительно равный содержанию 10 школьных учебников.
- низкий уровень потерь на распространение сигналов, обеспечивающих их передачу без регенерации на расстояния до 150 ... 175 км (и в перспективе до 350 км и более);
- абсолютная нечувствительность к электромагнитным помехам;
- отсутствие перекрестных помех (перекрестной модуляции) в ОК;
- малая масса и размеры ОК.

К другим достоинствам ОВ и ОК можно отнести такие, как достаточно высокая защищенность от несанкционированного перехвата передаваемой информации, пожаробезопасность, относительно невысокая стоимость ОК по сравнению с медными кабелями и практически неограниченные запасы сырья для производства ОВ. Все это делает их применение в сетях и системах связи еще более привлекательным и технически и экономически оправданным. Поэтому ОК почти полностью вытесняют в настоящее время другие виды направляющих структур в магистральных линиях цифровых первичных сетей связи.

Наряду с проводными линиями связи широко используют радиолинии различных диапазонов (от сотен килогерц до десятков гигагерц). Эти линии более экономичны и незаменимы для связи с подвижными объектами.

Для многоканальной системы радиосвязи при передаче информации на большие расстояния широко используются *радиорелейные линии* (РРЛ) связи.

*Радиорелейная связь* (радио и франц. *relais* — промежуточная станция) — радиосвязь, состоящая из группы ретрансляционных станций, расположенных на определенном расстоянии друг от друга, обеспечивающем устойчивую работу. Антенны станций линии радиорелейной связи устанавливают на мачтах

(башнях) высотой 70 ... 100 м. Протяженность линии радиорелейной связи может составлять до 10 000 км, емкость — до нескольких тысяч каналов.

История создания радиорелейных линий связи такова. В 1935 г. между городами Нью-Йорком и Филадельфией вступила в строй линия радиосвязи на ультракоротких (метровых) волнах. Она имела протяженность 150 км и две промежуточные («релейные») станции, которые принимали ослабленные радиоволны, «заменяли» их новыми, усиливали и посылали дальше. Сама радиолиния была названа «радиорелейной линией».

В зависимости от используемого метода распространения радиоволн радиорелейные линии связи можно разделить на две основные группы: *прямой видимости* и *тропосферные*.

Радиорелейные линии прямой видимости — основные наземные средства передачи сигналов телефонной связи, звукового и телевизионного вещания, цифровых данных и других сообщений на большие расстояния. Ширина полосы частот сигналов многоканальной телефонии и телевизионного вещания составляет несколько десятков мегагерц, поэтому для их передачи практически могут быть использованы диапазоны только дециметровых и сантиметровых волн, общая ширина спектра которых составляет 30 ГГц. Кроме того, в этих диапазонах почти полностью отсутствуют атмосферные и промышленные помехи.

Современные радиорелейные линии связи представляют собой цепочки достаточно мощных приемно-передающих радиостанций — ретрансляторов, последовательно принимающих, усиливающих, преобразовывающих (переносящих) сигналы на другие частоты и передающих далее сигналы от одного конца линии связи к другому (рис. 1.24). На каждой из промежуточных станций происходит восстановление и перенос сигнала на другую частоту, т. е. замена принятого слабого сигнала новым сильным, посылаемым на следующую станцию. Наиболее распространены радиорелейные линии метрового, дециметрового и сантиметрового диапазонов на частотах от 60 МГц до 15 ГГц.

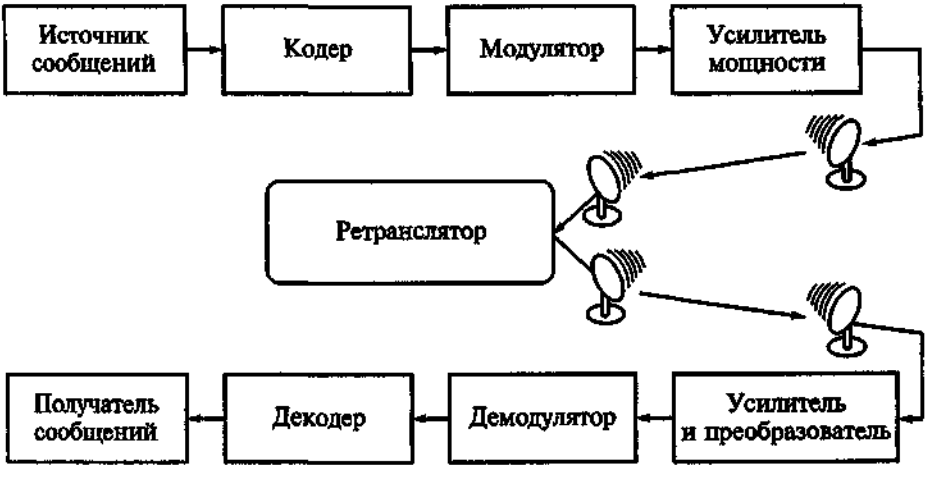



**Рис. 1.24.** Структурная схема радиорелейной линии связи

Все большее применение находят *спутниковые линии связи* — РРЛ с ретранслятором на искусственном спутнике Земли. В системах спутниковой радиосвязи используются радиоволны СВЧ-диапазона (обычно в пределах частот 1,5...14 ГГц, наиболее используемый диапазон 4...6 ГГц), пронизывающие ионосферу с минимальным затуханием. Передача информации на большое расстояние при одном ретрансляторе на ИСЗ, гибкость и возможность организации глобальной связи — важное преимущество спутниковых систем.

Основным преимуществом цифровых систем связи перед аналоговыми системами является их высокая помехоустойчивость. Это полезное качество наиболее сильно проявляется в системах передачи с *многократной ретрансляцией* (переприемом) сигналов. Типичные системы подобного типа — радиорелейные, волоконно-оптические и кабельные линии большой протяженности. В них сигналы передаются по цепи ретрансляторов, расположенных на таких расстояниях друг от друга, которые обеспечивают надежную связь. В таких системах помехи и искажения, возникающие в отдельных звеньях, как правило, накапливаются.

Для простоты положим, что радиосигнал в каждом ретрансляторе только усиливается. Тогда, если аддитивные помехи в каждом звене связи статистически независимы, их мощность на входе последнего звена равна сумме мощностей помех всех звеньев. Если система передачи информации состоит из $n$ одинаковых звеньев, для обеспечения заданной верности связи необходимо обеспечить на входе каждого ретранслятора отношение сигнал/помеха в $n$ раз больше, чем при передаче сигнала без ретрансляций. В реальных системах число ретрансляций $n$ может достигать несколько десятков, а иногда и сотен; накопление помех вдоль тракта передачи становится основным фактором, ограничивающим протяженность линии связи.

В цифровых системах передачи для ослабления эффекта накопления помех при передаче с ретрансляциями наряду с усилением применяют регенерацию импульсов, т. е. демодуляцию с восстановлением переданных кодовых символов и повторную модуляцию на переприемном пункте. При использовании регенерации аддитивная помеха с входа ретранслятора не поступает на его выход. Однако она вызывает ошибки при демодуляции. Ошибочно принятые в одном регенераторе символы в таком виде передаются и на следующие регенераторы, так что ошибки все же накапливаются. При цифровой системе передачи непрерывных сообщений можно, кроме того, повысить верность применением помехоустойчивого кодирования. Высокая помехоустойчивость цифровых систем передачи позволяет осуществить практически неограниченную по дальности связь при использовании каналов сравнительно невысокого качества.

### Системы радиовещания

Широкое распространение получили системы радиовещания — симплексные системы передачи информации, заключенной в звуковых сигналах (речь, музыка), от передатчика радиовещательной станции через линию связи — околоземное пространство — с помощью электромагнитных колебаний к приемникам слушателей (рис. 1.25).

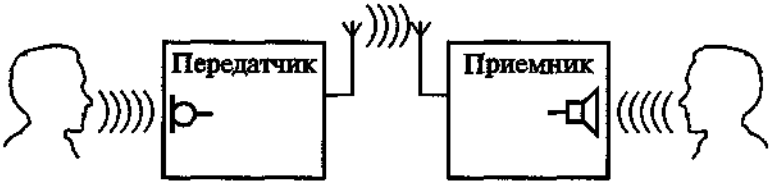



**Рис. 1.25.** Структурная схема канала радиовещания

На входе передатчика включен микрофон, который превращает звуковые волны, создаваемые диктором в студии, в электрические сигналы. На выходе приемника включен громкоговоритель (или наушники), осуществляющий обратное преобразование электрических сигналов в звуковые волны.

Необходимость перехода радиовещания к цифровым видам модуляции обусловлена, прежде всего, развитием цифровой техники, средств связи, вещания и информационных служб. Специалисты разработали различные системы цифрового радиовещания (ЦРВ), из которых широко признаны DRM (*Digital Radio Mondiale* — Всемирное цифровое радио), DAB (*Digital Audio Broadcas*ting — Радиовещание с цифровой звукозаписью), она же Eureka-147 Ю.

### *Телевизионные (телевещательные) системы*

Первый проект телевизионной системы с последовательной передачей элементов изображения предложил в 1878 г. португальский инженер А. де Пайва. Создателем первых разработок по электронной передаче изображения на расстояние был профессор Петербургского технологического института Борис Львович Розинг (1869–1938). В 1907 г. он изобрел электронную систему воспроизведения телевизионного изображения с помощью электронно-лучевой трубки, а в 1911 г. впервые продемонстрировал прием изображений простейших геометрических фигур. Его ученик, Владимир Кузьмич Зворыкин (1889–1972) в 1923 г. получил в США патент на изобретение *электронной системы телевидения* (от греч. *tele* — вдаль, далеко; в данном случае — далеко видеть).

К 1929 г. В.К. Зворыкин, существенно усовершенствовав электронно-лучевую трубку, создал *кинескоп* — приемную телевизионную трубку, а также в 1931 г. *иконоскоп* — передающую телевизионную трубку (независимо от него это сделал русский ученый Семен Исидорович Катаев; 1904–1991) и разработал



Б.Л. Розинг В.К. Зворы- С.И. Катаев

комплекс аппаратуры электронного телевидения. Большое значение имели труды В.К. Зворыкина по созданию цветного телевидения. Он признан во всем мире как «отец телевидения».

Телевизионные системы являются частным случаем радиовещательных и симплексных связных систем и между ними много общего. Однако есть ряд особенностей, выделяющих телевизионные системы из общего класса систем передачи информации. Телевизионной системой называют совокупность радиотехнических устройств, обеспечивающих передачу и прием изображений по радиоканалам, спутниковым, волоконно-оптическим или кабельным линиям связи. В настоящее время для передачи телевизионных сигналов изображения и звука по различным радиоканалам используют диапазоны метровых, дециметровых и сантиметровых (спутниковые системы) волн. Принцип действия всех телевизионных систем основан на последовательном построчном разложении изображения на мелкие элементы — *пикселы* (от англ. *pixel* — *picture element* — элемент изображения) и передачи о них информации.

Процесс построчного преобразования яркости (и цветности в цветном телевидении) элементов изображения в напряжение — *видеосигнал* (термин «видеосигнал», от англ. — *baseband signal*, определяет сигнал, спектр которого сосредоточен в диапазоне от постоянной составляющей до некоторого конечного значения, обычно не более нескольких мегагерц) — называют *растровой* (растр, от лат. *rastrum* — грабли) *разверткой*. При такой развертке площадь одного образца изображения (*кадра*) просматривают по двум взаимно перпендикулярным направлениям с достаточно высокой скоростью по горизонтали (*строчная развертка*) и более медленной по вертикали (*кадровая развертка*).

Системы черно-белого телевидения. Рассмотрим структурную схему черно-белого телевизионного передающего устройства без канала звукового сопровождения (рис. 1.26). В передающем устройстве телевизионный сигнал формируется *передающей трубкой*. Для построчной и кадровой развертки изображения на передающую трубку поступают два пилообразных напряжения от *генераторов строчной* и *кадровой разверток*. Пилообразное напряжение строчной развертки отклоняет электронный луч передающей трубки по горизонтали (по строке), а пилообразное напряжение кадровой развертки осуществляет его быстрое перемещение сверху вниз (по кадру). При совместной работе обоих генераторов луч перемещается (сканирует) по экрану слева направо, прочерчивая строки изображения, а когда прочерчена последняя строка, луч скачком возвращается к началу нового кадра. Поскольку каждая передаваемая строка должна вызывать синхронное свечение той же строки на экране телевизионного приемника, в *генераторах строчных* и *кадровых синхроимпульсов* (СИ) передатчика формируют и вместе с сигналом изображения передают прямоугольные импульсы строчной и кадровой синхронизации — *синхроимпульсы*. Они подаются на *видеоусилитель* (усилитель видеосигнала; от лат. *video* — смотрю, вижу) передающего устройства, а также синхронизируют *генераторы строчной* и *кадровой разверток* соответственно. Кроме того, для гашения луча в передающей трубке при возврате его из конца одной строки к началу другой, а

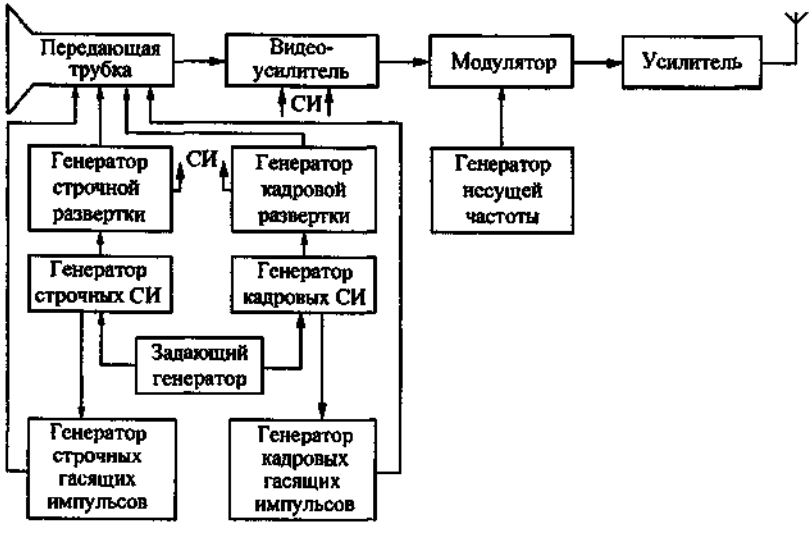



**Рис. 1.26.** Структурная схема телевизионного передающего устройства

также при смене кадров изображения генерируются *строчные* и *кадровые гасящие импульсы*. Синхронизирующие и гасящие импульсы должны быть согласованы по времени и поэтому вырабатываются из колебания одного высокостабильного по частоте *задающего генератора*. В результате в телевизионном передающем устройстве формируется *полный телевизионный сигнал*.

С помощью упрощенной временно́й диаграммы проанализируем структуру полного телевизионного сигнала (рис. 1.27).

Практически во всех телевизионных системах видеосигнал подают на катод приемной трубки (кинескопа). Вследствие этого более темным участкам передаваемого изображения должен соответствовать видеосигнал с более высоким потенциалом, чем светлым участкам, как показано на рис. 1.27. Верхней штриховой линией на рисунке отмечен некоторый потенциал — *уровень «черного»*, при котором кинескоп полностью закрыт и экран не светится. Нижней штриховой линии соответствует потенциал, называемый *уровнем «белого»*, при котором свечение экрана кинескопа практически максимально. Между названными уровнями возможно размещение потенциалов видеосигнала (1), который будет передаваться без заметных искажений.

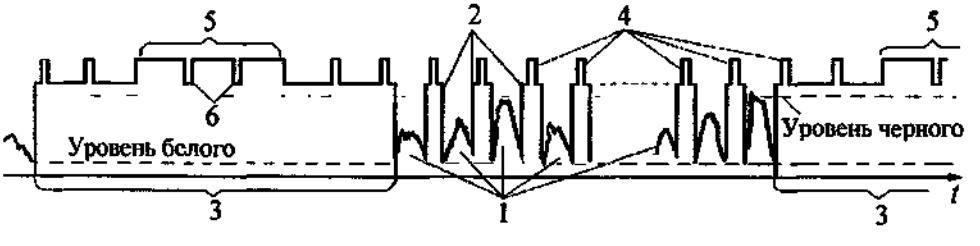



**Рис. 1.27.** Упрощенная временна́я диаграмма полного телевизионного сигнала

Все вспомогательные сигналы располагаются в области потенциалов выше уровня «черного». Здесь размещены гасящие строчные (2) и кадровые (3) импульсы, на «пьедесталах» которых находятся соответственно строчные (4) и кадровые (5) синхроимпульсы. В телевизионных приемниках указанные импульсы отделяют от остального сигнала и используют для запуска разверток и гашения лучей.

Отличительными признаками, по которым производится отделение вспомогательных импульсов от основных, являются их амплитуда, частота следования и длительность. Представленная картина усложняется тем, что строчные синхроимпульсы передают и во время действия кадровых импульсов. Используют также и другие вспомогательные сигналы: так называемые *импульсы «врезки»* (см. рис. 1.27, позиция 6) и *уравнивающие импульсы*, которые для упрощения на временно́й диаграмме не показаны.

После модуляции и усиления полученный телевизионный радиосигнал либо излучается в свободное пространство, либо направляется по специальным линиям связи к телеприемникам. Звуковой канал передающей системы в принципе аналогичен устройству симплексной системы связи, рассмотренной ранее. В современных аналоговых телевизионных системах традиционно применяется амплитудная модуляция несущего колебания сигналами изображения и частотная модуляция звуковым сигналом.

На рис. 1.26 приведена упрощенная телевизионная система с так называемой *построчной* (*прогрессивной*) *разверткой*, частота смены кадров в которой 25 Гц. Однако экран телеприемников такой системы с данной частотой смены кадров сильно мерцает и утомляет глаза человека, поэтому в реальных устройствах используют *чересстрочную* развертку. При этом луч сначала прочерчивает все нечетные строки, а затем — все четные строки изображения. В результате частота смены строк в кадре удваивается и эффект мерцания экрана существенно уменьшается.

В телевизионном приемном устройстве (рис. 1.28) осуществляется обратный процесс преобразования принятого антенной радиосигнала в яркость изображения на экране телевизионной трубки (в современных телевизорах на жидкокристаллическом или плазменном экране).

Модулированное электромагнитное колебание, принятое и преобразованное приемной антенной в телевизионный сигнал, поступает в *селектор каналов с преобразователем частоты*, с помощью которого телезритель подключает нужный канал передающей станции. В селекторе каналов телевизионного приемника также производятся усиление общим усилителем радиочастоты и преобразование несущих частот сигналов изображения и звука. Затем сигналы изображения и звука вместе усиливаются в *УПЧ изображения*. Сигнал изображения детектируется в *видеодетекторе*. Здесь же происходит разделение сигналов изображения и звука. Затем сигнал изображения усиливается в *видеоусилителе* и подается на *кинескоп* (*телевизионную трубку*). Промежуточная частота звукового сигнала усиливается в *УПЧ звука* и после детектирования в *детекторе звука* и усиления *в усилителе звука* подается в *громкоговоритель* (Гр).

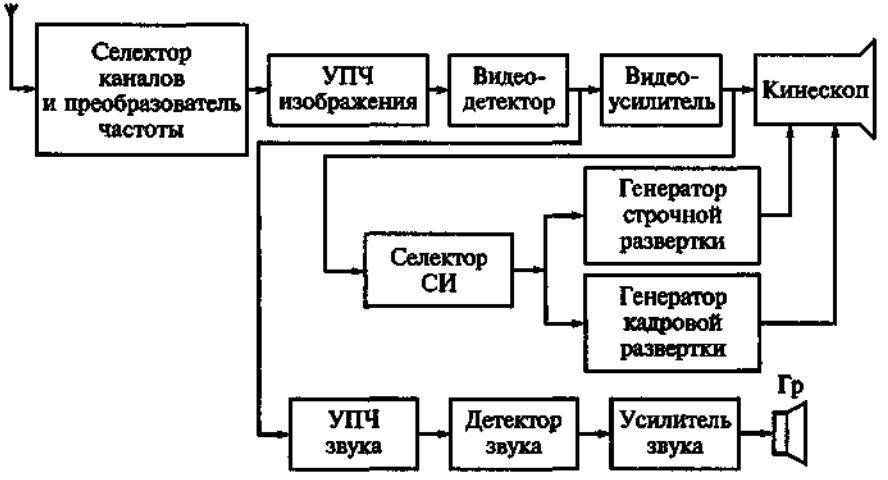



**Рис. 1.28.** Структурная схема телевизионного приемного устройства

Синхронность разверток электронных лучей приемника и передатчика Рассматриваемой телевизионной системы обеспечивается синхроимпульсами строчной и кадровой разверток, которые выделяются селектором синхроимпульсов (селектором СИ) из полученного на выходе видеоусилителя видеосигнала. Эти импульсы подаются на генераторы строчной и кадровой разверток, которые управляют перемещением электронного луча кинескопа по строкам и кадрам соответственно.

Основные параметры и характеристики систем черно-белого телевидения следующие: развертка — чересстрочная; длительность развертки одной строки вместе с обратным ходом луча 64 мкс; длительность развертки одного кадра вместе с временем обратного хода луча 0,04 с; число строк в кадре $Z = 625$; частота следования строк $F_с = F_к \cdot Z = 25 \cdot 625 = 15\,625$ Гц; частота смены кадров $F_к = 25$ Гц; частота смены полей 50 Гц; число элементов разложения в одном кадре $n \approx 521\,000$. Ширина полосы частот сигнала телевизионного изображения составляет около 6,5 МГц. Ширина полосы частот видеосигнала совместно с полосой звуковых частот составляет 8 МГц.

Системы цветного телевидения. Сейчас применение систем черно-белого телевизионного вещания ограничено. Современное вещательное телевидение основано на передаче цветного изображения.

В настоящее время существуют три основные совместимые системы цветного телевидения:

- совместная советско-французская система SECAM (от франц. *Systeme en Couleur avec Memoire* — цветная система с запоминанием; предложена французским инженером А. де Франсом в 1958 г.);
- западно-германская PAL (от англ. *Phase Alternation Line* — строки с переменной фазой; предложена и разработана инженером В. Брухом в 1962 г.);
- американская NTSC (от англ. *National Television System Commitee* — Национальный комитет телевизионных систем; разработана в 1953 г.).

Цветное телевидение в Российской Федерации действует по системе SECAM, совместимой с черно-белой системой и отличающейся поочередной (последовательной) передачей двух цветоразностных сигналов при непрерывной передаче сигнала яркости. Принята во Франции, Российской Федерации, некоторых странах Европы и Африки. В телевизионной системе SECAM для сохранения четкости изображения (передачи мелких деталей) полоса видеосигнала яркости должна быть около 6,5 МГц. Поскольку на экране заметна окраска только крупных деталей изображения, полоса частот цветоразностных сигналов может быть у́же (примерно 700 кГц).

В последнее время телевизионные приемники в нашей стране способны принимать передачи цветного телевидения и по системе SECAM, и по системе PAL. Так же как и SECAM, система PAL в странах Европы и в России работает в 625 строках с частотой 25 кадров, или 50 полей в секунду.

Телевещание с использованием стандарта PAL началось в 1967 г. в Германии и Великобритании, причем несмотря на одинаковое название, системы несколько различались. Так осталось и поныне, только вариантов PAL систем стало еще больше. Для решения проблем с разделением сигнала на составляющие через строку меняется знак амплитуды сигнала *U*. Поэтому колебания яркостного сигнала влияют только на небольшое изменение цветовой насыщенности. Эта методика, по сути, вдвое снижает вертикальное разрешение.

Разработку телевизионной системы NTSC осуществила американская компания *National Television System Committee* (NTSC), аббревиатура которой и дала название стандарту. Это первый формат цветного телевидения, который получил наиболее широкое распространение. Полностью стандарт NTSC был сформулирован 17 декабря 1953 г. в США. Одной из задач, которые ставились при разработке системы NTSC, являлась совместимость с существовавшим на то время форматом черно-белого вещания. Это и определило разрешение в 525 строк с частотой 30 кадров, или 60 полей в секунду. Из-за особенностей большинства телевизионных приемников на самом деле, обычно, человек видит всего около 480 строк.

В системе SECAM передаваемый полный цветной телевизионный сигнал по структуре существенно сложнее, чем сигнал черно-белого телевидения, и несет в себе значительно больше информации. Широковещательная система цветного телевидения SECAM полностью совместима с этой же системой черно-белого телевидения — передачу сигналов цветного изображения осуществляют по тем же каналам и в той же полосе рабочих частот, по которым ведется черно-белое телевизионное вещание. Общие характеристики цветного телевидения практически такие же, как и у черно-белого телевидения.

Для обеспечения принципа совместимости обеих телевизионных систем — цветного и черно-белого — в составе сигнала цветного телевидения должен быть такой сигнал, который на экране черно-белого телевизора давал бы черно-белое изображение. Следовательно, одним из сигналов изображения цветного телевидения должен быть сигнал опорного белого — сигнал яркости. А сколько же цветов надо при этом передавать?

При создании систем цветного телевидения используют физиологические свойства глаза человека. Согласно общепринятой теории трехкомпонентного зрения, три вида колбочек в глазу человека воспринимают синий, зеленый и красный цвет. Остальные цвета представляют собой смесь этих основных цветов: *белый* цвет для глаза человека может быть представлен суммой световых потоков всего трех цветов: красного (*red*) $E_R$, зеленого (*green*) $E_G$ и синего (*blue*) $E_B$. Таким приемом широко пользуются художники, смешивая на палитре краски. Поэтому сигнал яркости $E_Y$ легко кодируется тремя сигналами цветности

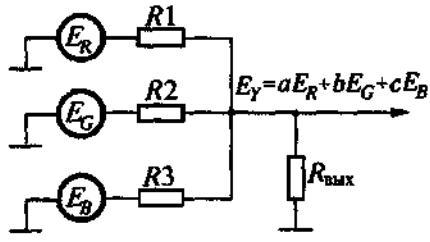



**Рис. 1.29.** Структурная схема формирователя сигнала яркости

$$E_Y = aE_R + bE_G + cE_B. \qquad (1.18)$$

Значения коэффициентов неодинаковы: $a = 0{,}30$; $b = 0{,}59$; $c = 0{,}11$. Они и определяются чувствительностью глаза к различным цветам. Сигнал яркости должен иметь такую же ширину спектра, как в черно-белом телевидении. Сигналы цветности могут быть сравнительно узкополосными, так как глаз человека воспринимает окрашенными в различные цвета только относительно крупные детали. Мелкие детали глаз видит черно-белыми.

На рис. 1.29 приведена структурная схема формирователя (электрическая матрица), используемого для получения сигнала яркости из трех основных цветов. Так как операция получения напряжения сигнала яркости $E_Y$ из цветовых сигналов является линейной и коэффициенты $a$, $b$, $c$ меньше единицы, то электрическая матрица содержит резисторы $R1$, $R2$, $R3$ и представляет собой по существу делители напряжений с общим выходным сопротивлением $R_{вых}$. Для получения информации о цвете изображения используют два цветоразностных сигнала, к искажениям тонов которых глаз менее чувствителен: $E_R - E_Y$ и $E_B - E_Y$. При передаче трех сигналов: $E_Y$, $E_R - E_Y$, $E_B - E_Y$ — яркости и двух цветоразностных — в телевизионном приемнике возможно получение всех требуемых сигналов: $E_Y$, $E_R$, $E_G$, $E_B$. Для этого необходимо выполнение следующих операций:

$$E_R = E_Y + (E_R - E_Y);$$

$$E_G = [E_Y - (aE_R + cE_B)]/b; \qquad (1.19)$$

$$E_B = E_Y + (E_B - E_Y).$$

Формирование сигнала яркости $E_Y$ из трех составляющих $(E_R - E_Y)$, $(E_G - E_Y)$ и $(E_B - E_Y)$ осуществляют электрическими матрицами. Операцию вычитания выполняют элементарным прибавлением сигнала с противоположной фазой.

Для обеспечения совместимости систем цветного и черно-белого телевидения необходимо также, чтобы частота кадров была одинаковой и равной 25 Гц, полоса частот видеосигнала составляла примерно 6,5 МГц, а число строк — 625.

В системе SECAM для передачи цветоразностных сигналов используют две поднесущие частоты, сдвинутые относительно несущей частоты изображения $f_0$ на интервалы $f_{0B}$ = 4,25 МГц и на $f_{0R}$ = 4,406 МГц соответственно для передачи цветоразностных сигналов ($E_B - E_Y$) и ($E_R - E_Y$). Поднесущие частоты лежат в области малых составляющих спектра сигнала яркости и модулируются цветоразностными сигналами. При этом применяют частотную модуляцию с частичным подавлением верхних боковых полос. Цветоразностные сигналы передают поочередно — через строку. В течение передачи одной строки передают один цветоразностный сигнал, в течение передачи следующей строки — другой. Канал связи при этом «уплотняется», однако ширина спектра остается неизменной.

Упрощенная структурная схема цветного телевизионного передатчика изображена на рис. 1.30. Световой поток от изображения, пройдя через объектив О, падает на дихроичные (цветоразделительные) зеркала (светофильтры) $ДЗ_1$ и $ДЗ_2$, разделяющие световой поток по спектру и выделяющие красную, зеленую и синюю составляющие, которые с помощью зеркал $З_1$ и $З_2$ подают на соответствующие телевизионные передающие трубки $ТП_1$, $ТП_2$ и $ТП_3$ (структурно трубки объединены в одну). На выходе трубок формируются видеосигналы $E_R$, $E_G$ и $E_B$. Видеосигналы через соответствующие усилители ▷ поступают на электрическую матрицу, на выходе которой появляются сигналы $E_Y$, $E_R - E_Y$ и $E_B - E_Y$.

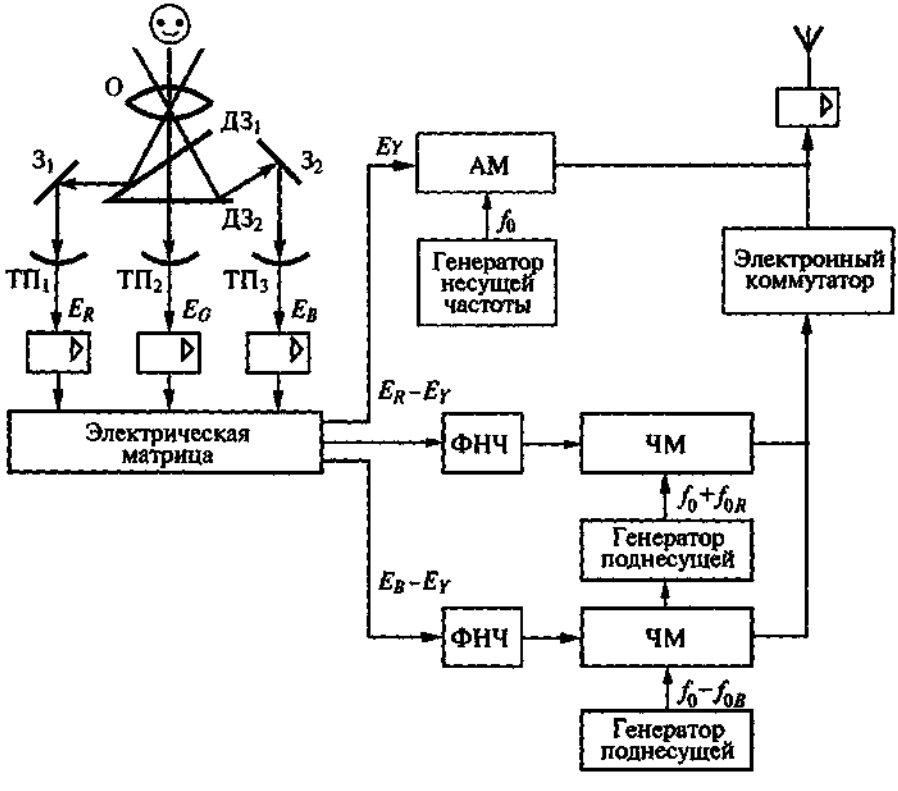



**Рис. 1.30.** Упрощенная структурная схема цветного телевизионного передатчика

Сигнал $E_Y$ поступает на вход амплитудного модулятора (АМ), на другой вход которого подается колебание несущей частоты изображения $f_0$ от генератора несущей частоты.

Амплитудный модулятор формирует амплитудно-модулированный сигнал, несущий информацию об яркости отдельных элементов изображения. Цветоразностные сигналы $E_R - E_Y$ и $E_B - E_Y$ после ограничения их спектра — фильтрации в *фильтрах низкой частоты* (ФНЧ) — поступают на частотные модуляторы (ЧМ), на которые подают также несущие колебания $f_0 + f_{0R}$ и $f_0 - f_{0B}$. Электронный коммутатор поочередно пропускает частотно-модулированные сигналы. Эти сигналы вместе с амплитудно-модулированным сигналом яркости, после усиления в усилителе мощности ▷, поступают в антенну телевизионного передатчика, которая излучает полученные электромагнитные волны. В упрощенной схеме цветного телевизионного передатчика имеются еще генератор синхроимпульсов и генераторы развертки, которые на рис. 1.26 для упрощения не показаны. Передача звука производится так же, как и в черно-белом телевидении.

На рис. 1.31 приведена упрощенная структурная схема цветного телевизионного приемника. Схема содержит антенну, входную цепь, усилитель высокой частоты (УВЧ), смеситель и гетеродин (последние четыре элемента входят в блок селектора каналов). С выхода смесителя сигнал поступает на усилитель промежуточной частоты (УПЧ), затем детектируется видеодетектором и усиливается видеоусилителем. Практически все эти элементы приемника, а также канал звука и развертки аналогичны соответствующим элементам черно-белого телевизионного приемника.

Основное отличие цветного телевизионного приемника — наличие блока цветности (на рис. 1.31 он для упрощения не выделен), формирующего три цветоразностных сигнала $(E_R - E_Y)$, $(E_G - E_Y)$ и $(E_B - E_Y)$. Фильтр $E_Y$ с полосой частот 6,5 МГц, включенный на выходе видеоусилителя, выделяет сигнал яркости.

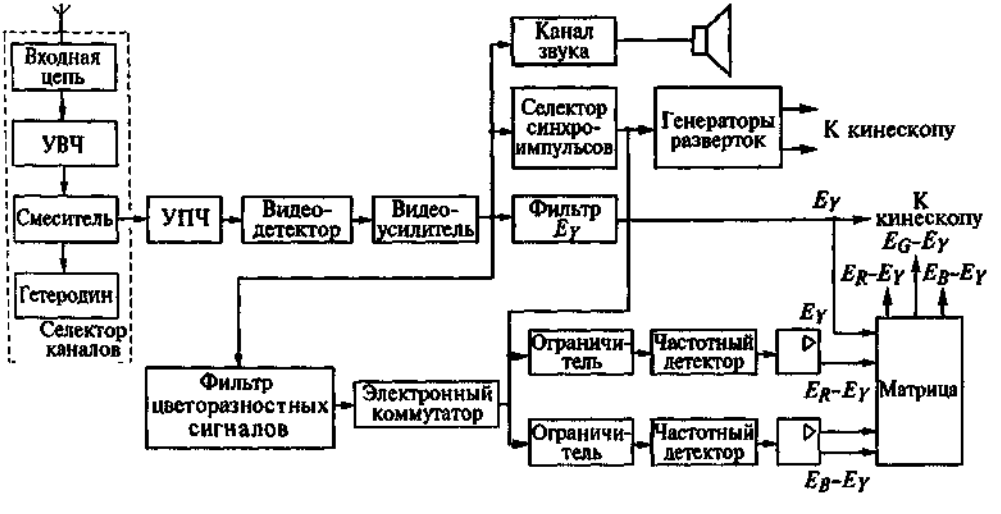



**Рис. 1.31.** Упрощенная структурная схема цветного телевизионного приемника

Сигнал яркости $E_Y$ поступает и на вход фильтра цветоразностных сигналов, который выделяет частотно-модулированные цветоразностные сигналы с несущими частотами $f_{0R}$ и $f_{0B}$. С помощью электронного коммутатора, управляемого синхроимпульсами селектора синхроимпульсов, эти сигналы поступают на ограничители (для устранения паразитной амплитудной модуляции), частотные детекторы и усилители. На выходе усилителей получают цветоразностные видеосигналы ($E_R - E_Y$) и ($E_G - E_Y$). Эти сигналы и сигнал яркости $E_Y$ подаются на матрицу, на выходе которой получают три цветоразностных сигнала ($E_R - E_Y$), ($E_G - E_Y$) и ($E_B - E_Y$). Вместе с сигналом $E_Y$ они поступают на цветной кинескоп и позволяют на нем получить цветное изображение.

Системы цифрового телевидения. Структурная схема современной цифровой телевизионной системы приведена на рис. 1.32.

В качестве *источника телевизионного* (ТВ) *сигнала* (как правило, цветного) служит передающая камера или видеомагнитофон, напряжение с выхода которых поступает на *АЦП изображения*. В цифровой телевизионной системе используется компонентное (раздельное) кодирование, при котором в цифровую форму отдельно преобразуют сигнал яркости и цветоразностные сигналы. Преобразованные в цифровую форму сигнал яркости и цветоразностные сигналы подают на *кодер изображения*. Схема кодирования достаточно сложна и в ней заложен ряд новейших принципов цифрового преобразования сигналов.

В передатчике вместе с сигналами изображения формируют и сигналы звукового сопровождения. Звуковой сигнал с *источника звука* через *АЦП звука* подают на *кодер звука*. Кодированные сигналы изображения и звука объединяют в общий поток информации в *мультиплексоре*. Далее объединенные сигналы изображения и звука поступают на *кодер канала и модулятор*, где осуществляют помехоустойчивое кодирование и модуляцию несущего колебания.

Сформированный радиосигнал цифровой телевизионной системы передается по соответствующему *каналу* (линии) *связи* и поступает в телевизионный приемник. Здесь, в *демодуляторе* и *декодере канала*, производится демодуляция




**Рис. 1.32.** Структурная схема системы цифрового телевидения

(детектирование) несущего колебания и декодирование помехоустойчивого кода общего сигнала изображения и звука. Затем в *демультиплексоре* из общего потока информации выделяются цифровые яркостные и цветоразностные сигналы изображения и сигналы звука, которые потом раздельно обрабатываются в *декодере изображения* и *декодере звука*. После преобразования сигнала яркости и цветоразностных сигналов в *цифроаналоговом преобразователе* (ЦАП, от англ. *Digital-to-Analog Converter, DAC*) *изображения звука* в аналоговую форму, они поступают на монитор (экран), на котором воспроизводится цветное изображение. Декодированный сигнал звука также преобразуется в *ЦАП звука* в аналоговую форму и поступает на громкоговоритель телевизионного приемника.

В последние годы требуемую полосу частот цифрового телевизионного сигнала удалось сузить в 8 раз, используя оптимальное цифровое кодирование телевизионного сигнала, аналогичное кодированию информации в глазу человека. Это позволяет, например, по одному каналу спутниковой связи передавать одновременно две программы цифрового телевидения.

Системы телевидения высокой четкости. Существенное расширение возможностей экрана принесет человеку широкое внедрение цветного телевидения высокой четкости (ТВЧ, от англ. *High Definition TV* — HDTV). В телеприемниках ТВЧ установлено новое соотношение высоты и ширины экрана. Оно соответствует широкоэкранному кино, т. е. 9 : 16 (в обычном телевизоре — 3 : 4). Главное же достоинство ТВЧ состоит в том, что изображение в кадре формируется очень большим количеством строк (более 1 500). Это позволяет получить невиданную доселе яркость и четкость изображения, сравнимого по качеству со слайдом или цветной 35-мм пленкой. Изображение на экране представляется в своем естественном виде с тончайшими цветовыми нюансами. Даже эксперты не всегда могут отличить изображение на экране телевизора высокой четкости от изображения за чисто вымытым окном. Установлено, что изображение ТВЧ по количеству содержащейся видеоинформации превосходит применяемые сейчас системы практически вдвое, а по яркости красок и богатству деталей — во много раз. Звук в ТВЧ системах — стереофонический.

Тенденции развития телевидения. В ближайшие годы телевизионные системы новых поколений будут обладать более широкими возможностями по сравнению с существующими. Прежде всего, это связано с перспективами внедрения массовых *интерактивных* (двусторонних) телевизионных систем, превращающих телевизионные приемники в своеобразные дуплексные устройства, обеспечивающие передачу информации от зрителя по обратному каналу с помощью маломощных цифровых узкополосных радиопередатчиков. Безусловно практически все телевидение будет цветным, цифровым и высокочетким. В последние годы все более широкое применение в мире (и в России) находят *кабельные, спутниковые* и *сотовые системы телевизионного вещания*.

Главная особенность современных телевизоров — многофункциональность. К этим телевизорам можно подключать видеомагнитофон, персональный компьютер, различные специальные приставки. Созданы телевизоры с плоским экраном, которые можно повесить на стену, как картину, и проекционные телевизоры с большим

экраном площадью сотни квадратных метров для коллективного просмотра передач. Разработаны телевизоры со стереоскопическим изображением, карманные телевизоры со сверхмалым экраном, а также телеприемники, на экранах которых будут приниматься одновременно до шести телеканалов.

Цифровая техника позволяет ввести в запоминающее устройство телевизора программу на несколько дней, чтобы утром телевизор разбудил, включил интересующую зрителя программу, автоматически записал на видеомагнитофон нужную передачу. Через системы «Видеотекст», «Телетекст» и прочие уже можно запросить из информационного центра и увидеть на экране необходимую информацию: расписание самолетов, синоптическую карту с прогнозом погоды, сведения о новинках, поступающих в продажу.

## *Радиотехнические системы обнаружения и измерения*

Радиотехнические системы обнаружения и измерения выделяют полезную информацию из принятых сигналов. Это имеет место в системах радиолокации, радионавигации и радиотелеметрии. К радиотехническим системам обнаружения и измерения относятся также так называемые *пассивные радиосистемы*, когда радиопередатчик в системе отсутствует, а информация извлекается радиоприемным устройством из сигналов, поступающих от каких-либо естественных источников электромагнитных колебаний. Приемники сигналов радиотепловых источников (инфракрасных, или ИК-источников), называемые радиометрами, используются, в частности, в пассивной локации.

### Радиолокационные системы

*Радиолокация* (от лат. *locatio* — расположение, размещение и означает определение местоположения объекта по сигналам, излучаемым самим объектом — пассивная локация — или отраженным от него сигналом, излучаемым самой радиолокационной станцией — РЛС — активная локация) — область науки и техники, предметом которой является наблюдение различных объектов (целей) радиотехническими методами: их обнаружение, определение пространственных координат и направление движения, измерение дальности и скорости движения, разрешение, распознавание и др.

Первые РЛС (они получили название RADAR — от англ. *Radio Detection And Ranging*, т. е. *прибор для радиопеленгации и измерения*) были разработаны в 30-е годы XX в. в наиболее развитых странах (США, СССР, Великобритании, Германии). В годы Второй мировой войны англичане успешно использовали радиолокацию для борьбы с немецкими подводными лодками. Первая РЛС непрерывного излучения была создана в 1932 г. (Э. Тейлор, Л. Юнг, Л. Хиленд, США). Работы по созданию отечественных РЛС начались в Советском Союзе в 30-е годы прошлого столетия. В 1934 г. были разработаны первые советские РЛС непрерывного излучения (Борис Константинович Шембель, 1900–1987, и независимо Юрий Константинович Коровин, 1907–1988). Фактическими основателями школы советских специалистов по радиолокации были профессора


П.К. Ощепков


Б.К. Шембель


Ю.Б. Кобзарев


А.А. Расплетин

Павел Кондратьевич Ощепков (1908–1992), Б.К. Шембель, академик Юрий Борисович Кобзарев (1905–1992), Ю.К. Коровин и другие ученые. Неоценимый вклад в развитие отечественной радиолокации внес до сих пор мало известный академик Александр Андреевич Расплетин (1908–1967).

*Обнаружением* называют процесс принятия решения о наличии в радиолокационном луче целей с допустимой вероятностью ошибочного решения. При определении местоположения целей оценивают их координаты и параметры движения, в том числе скорость.

Итак, определение местоположения целей разделяется на две задачи:

- определение дальности (дальнометрия);
- определение условных координат (радиопеленгация).

Под *разрешением* понимают возможность раздельно обнаруживать и измерять координаты одной цели при наличии других, близкорасположенных.

*Распознавание* — получение радиолокационных характеристик различных объектов, выбор информативных устойчивых признаков и принятие решения о принадлежности этих признаков к тому или иному классу.

Технические средства получения информации о радиолокационных целях и называются *радиолокационными станциями* или *системами*. Носителями радиолокационной информации служат приходящие от целей радиолокационные сигналы. Они образуются в результате вторичного излучения, т. е. либо переизлучения первичного излучения специальной аппаратурой или поверхностью цели, либо собственного электромагнитного излучения целей. Соответственно различают методы активной радиолокации, радиолокации с активным ответом и пассивной радиолокации. В первых двух случаях РЛС излучает в направлении на цель зондирующий сигнал, в последнем — облучения целей не требуется. В англоязычной литературе пассивные РЛС называют *primary radar* — первичными радиолокаторами.

Основной целью радиолокации является установление связи между параметрами передающей (приемной) системы и характеристиками отраженного и рассеянного радиолокационной целью излучения с учетом их взаимного расположения в пространстве. Для решения такой задачи при проектировании РЛС используется фундаментальное соотношение, которое носит название *основного*

*уравнения радиолокации* и служит для оценки предельной дальности $R_{max}$ (в локации принято дальность обозначать не $D$, а $R$) обнаружения локатором радиолокационной цели (предполагается, что приемник и передатчик совмещены в пространстве и «работают» на одну антенну):

$$R_{max} = \sqrt[4]{\frac{P_c}{P_{min}} \frac{\lambda^2 \sigma}{64\pi^3} G_0^2 F^2}, \qquad (1.20)$$

где $P_c$ — мощность радиопередатчика; $P_{min}$ — минимальная принимаемая мощность; $G_0$ — коэффициент направленного действия (КНД) антенны в направлении максимума излучения; $F$ — нормированная функция направленности по напряженности поля антенны; $\sigma$ — ЭПР (эффективная площадь рассеяния) цели.

Рассмотрим работу импульсной РЛС (рис. 1.33) (первые импульсные РЛС разработаны в 1935 г. П.К. Ощепковым, СССР, и независимо Р. Ватсон-Ваттом — *R. Watson-Watt*; 1892–1973; Великобритания).

На рис. 1.33, *а* показана структурная схема такой РЛС, способной обнаруживать объекты и измерять дальность их расположения от пункта обнаружения. Основными устройствами импульсной РЛС являются: генератор импульсов, передатчик, состоящий из генератора несущей частоты и модулятора, антенный переключатель (АП), остронаправленная антенна, приемник и измеритель. Принцип действия импульсной РЛС поясняется с помощью упрощенных временны́х диаграмм, показанных на рис. 1.33, *б*.

Генератор импульсов вырабатывает достаточно короткие (доли или единицы микросекунд) импульсы (*1* на рис. 1.33, *б*), которые определяют частоту посылок радиосигналов РЛС. Эти импульсы поступают на передатчик и измеритель. В передатчике с помощью модулятора из несущего колебания формируют высокочастотные импульсы *2* (осуществляется импульсная модуляция), называемые *радиоимпульсами*, которые излучают в окружающее пространство. Антенный переключатель подключает антенну к передатчику во время излучения радиоимпульсов и к приемнику — в интервалах между ними.

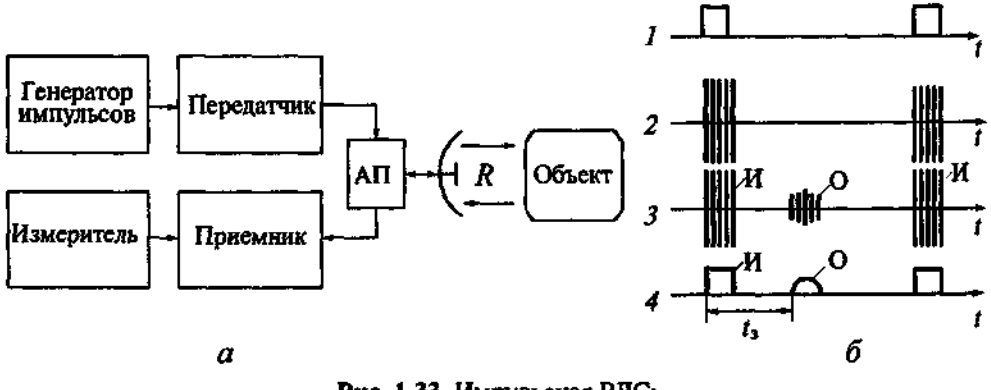



**Рис. 1.33.** Импульсная РЛС:
*а* — структурная схема; *б* — упрощенные временны́е диаграммы

Отраженные от объекта и уловленные антенной РЛС радиоимпульсы *3* попадают в приемник. Отраженные радиоимпульсы располагаются в интервалах между излученными импульсами (соответственно О и И на рис. 1.33, *б*), небольшая часть мощности которых через антенный переключатель также проникает в приемник. После усиления и детектирования в приемнике отраженные импульсы *4* поступают в измеритель. Сравнение в измерителе отраженного импульса с его излученной копией, поступающей с генератора импульсов, позволяет получить информацию об объекте. В частности, дальность до обнаруженного объекта определяется по времени задержки излученного сигнала $t_з$ в соответствии с известной формулой

$$R = ct_з/2. \qquad (1.21)$$

Работают РЛС обычно в диапазонах метровых, дециметровых, сантиметровых и миллиметровых волн, так как в этом случае удается создать узкие (игольчатые) диаграммы направленности при приемлемых габаритных размерах антенн.

В настоящее время принцип действия множества радиолокаторов основан на *эффекте Доплера* (1842 г. К. Доплером установлена зависимость частоты звуковых и световых колебаний от взаимного движения источника и наблюдателя; *K. Doppler*; 1803–1853).

Пассивные РЛС. Известно, что в реальных земных условиях все тела излучают собственное тепловое или радиоизлучение, интенсивность которого больше в ИК-диапазоне и видимом оптическом диапазоне и существенно меньше в радиодиапазоне волн. Тем не менее установлено, что в радиодиапазоне на коротких сантиметровых и миллиметровых волнах оно оказывается весьма заметным и может нести важную полезную информацию. Прием такого излучения может выполняться скрытно от обнаруживаемого объекта.

На рис. 1.34 показана простейшая функциональная схема системы пассивной локации объектов на фоне неба или подстилающей земной поверхности. На вход высокочувствительного радиометра 1 по тракту 2 поступает принятое антенной 3 собственное ИК-излучение объекта 4. В радиометре полученная информация об объекте регистрируется и обрабатывается.

Нелинейные РЛС. Существенное увеличение числа радиолокационных задач стимулирует поиск нетрадиционных методов построения локаторов. Один из таких методов основан на использовании *нелинейного рассеяния электромагнитных волн*. Под нелинейным рассеянием электромагнитных волн в радиолокации понимают явление обогащения спектра сигнала, переотраженного




**Рис. 1.34.** Функциональная схема пассивной локации объектов

обнаруживаемой целью, по сравнении со спектром сигнала облучающего электромагнитного поля. Такой эффект возникает за счет нелинейных свойств отдельных отражающих элементов цели.

Специалистами в области радиотехники давно замечено, что некачественно выполненные электрические соединения и разъемы радиотехнических устройств, расположенные вблизи мощного передатчика РЛС, при их облучении электромагнитным полем могут создавать сигналы на частотах, отличных от частоты излучения. Эти свойства нелинейности электрических соединений были всесторонне изучены и применены на практике. Лабораторные испытания показали, что значительная часть плотных механических соединений металла с металлом и тщательно выполненные пайки практически обладают свойствами пассивных сопротивлений. Поэтому при протекании через них переменного тока не возникают ни гармоники, ни комбинационные частоты. Однако если между металлами нет плотного молекулярного контакта и имеющийся воздушный зазор составляет очень небольшую часть длины волны облучающих их колебаний, то образуется значительная нелинейная проводимость, на концах которой возникает разность потенциалов вплоть до 1 В. При этом прямая ветвь вольт-амперной характеристики стального перехода подобна аналогичной характеристике обычного полупроводникового диода.

Для контакта металл–металл с протекающим в нем переменным током характерно преобладание генерации нечетных гармоник излучения передатчика локатора, причем наиболее ярко выражена третья гармоника, в отличие от полупроводников, где преобладает генерация второй гармоники. Зазор, необходимый для получения нелинейной проводимости между металлами, должен быть около 100 Å, поэтому в большинстве сложных металлических объектов имеется очень много «генераторов гармоник», каждый из которых образован металлическими частями, поворачивающимися, скользящими или неподвижными относительно друг друга. Это могут быть шарнирные крепления дверей, листовые рессоры, стеклоочистители, инструментальные ящики, разводные гаечные ключи, монеты и т. д.

На сегодняшний день известны два варианта построения нелинейных РЛС с использованием передатчика, работающего:

- на одной частоте, и приемника гармоник этой частоты;
- на двух частотах ($f_1$ и $f_2$), и приемника, настроенного на сильный сигнал одной из комбинационных (разностной или суммарной между $f_1$ и $f_2$) частот.

В последнем случае нелинейный контакт двух материалов выполняет роль находящегося на расстоянии нелинейного смесителя частот, вырабатывающего ряд комбинационных частот. Первый вариант проще в реализации. При отработке систем связи такие РЛС используют для локализации источников *интермодуляционных искажений* — ИМИ; *intermodulation distortions —IMD* («эффект ржавого болта»). Присущая нелинейному радиолокатору защищенность от помех естественного происхождения определяет возможность его применения в сугубо военных целях для выделения объектов искусственного происхождения (например танков, бронетранспортеров) на фоне земных покровов.

Уникальные свойства такой РЛС наделяют ее потенциально важной ролью во многих применениях, где не требуется большая дальность (например, в обнаружителях подслушивающих устройств).

Кратко коснемся акустоэлектронных и оптических систем извлечения информации. Развитие *акустоэлектронных систем* извлечения информации, работающих по принципу РЛС, потребовало разработки мощных импульсных ультразвуковых генераторов и соответствующих систем обработки отраженных от объектов акустических сигналов сложной формы. По аналогии с РЛС (радарами) такие системы назвали *сонарами* (от англ. SONAR — *SOund Navigation And Ranging* — гидролокатор, эхолот). Установлено, что современные сонары позволяют «видеть» и исследовать внутренние органы человека, заглянуть в глубь Земли на расстояние до 5 км, находить в морской воде рыбные косяки и подводные лодки на глубине до 10 км.

С появлением мощных импульсных оптических излучателей направленного действия (лазеров) начали интенсивно развиваться *оптические системы извлечения информации*. По аналогии с радарами такие системы стали называть *лидарами* (лазерные локаторы ИК-диапазона). Современные лидары позволяют определять расстояние от Земли до Луны с точностью до нескольких метров, наблюдать искривление земной поверхности при приливах, определять координаты спутников и летающих объектов, состав атмосферы и наличие в ней загрязняющих примесей.

### Радионавигационные системы

*Радионавигация* — область науки и техники, занимающаяся теорией и практическим применением радиотехнических систем для определения местоположения космических объектов, самолетов, судов, других транспортных средств и людей на водной и земной поверхностях, в околоземном или космическом пространстве.

*Радионавигационными системами* называют совокупность нескольких радиотехнических устройств, расположенных в одном или нескольких неподвижных наземных пунктах и на подвижном объекте, взаимодействующих между собой с помощью каналов радиосвязи и предназначенных для определения навигационных величин или навигационных элементов. Системы данного типа предназначены для определения географических координат морских кораблей, самолетов, космических кораблей и других управляемых и пилотируемых объектов, а также их проведения по заданному курсу. Для этих целей используют радиотехническую аппаратуру, установленную как на борту управляемого или пилотируемого объекта, так и в ряде определенных точек поверхности Земли. Обычно на земной поверхности располагают два или более радиопередатчика, координаты которых заранее известны. На рис. 1.35 показан принцип действия простейшей радионавигационной системы морского корабля. На берегу моря находятся два постоянно действующих радиопередатчика — *радиомаяка* — $РМ_1$ и $РМ_2$. Для определения местонахождения корабля достаточно в его управляю-

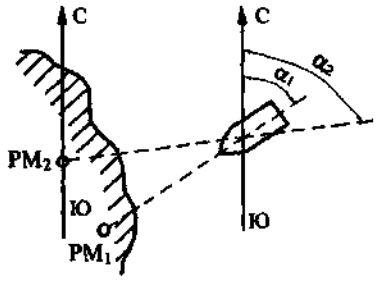



**Рис. 1.35.** К пояснению принципа радионавигации

щем компьютере вычислить угловые координаты $\alpha_1$ и $\alpha_2$, по которым принимают излучаемые с берега моря радиомаяками сигналы.

Спутниковые системы радионавигации. В настоящее время требования к точности определения навигационных параметров повышены, а запросы на услуги радионавигационных систем значительно увеличились. Это привело к бурному развитию систем навигации в виде так называемых *спутниковых радионавигационных систем* (СРНС). Примерами *спутниковых радионавигационных систем*, используемых для получения глобальной оперативной высокоточной навигации сухопутных, морских, воздушных и низкоорбитальных космических объектов, могут служить созданные в 1980–1990 гг. среднеорбитальные СРНС второго поколения — системы GPS (НАВСТАР) в США и ГЛОбальной НАвигационной Спутниковой Системы (по первым буквам) ГЛОНАСС в России. Термин «глобальная оперативная навигация» означает, что подвижный объект, оснащенный *навигационной аппаратурой потребителя* (НАП), может в любом месте приземного пространства в любой момент времени определить (уточнить) параметры своего движения — три координаты и три составляющие вектора скорости.

В СРНС применяются космические радиомаяки (навигационные ИСЗ) и навигационные радиосигналы, содержащие так называемую *эфемеридную информацию* (ЭИ — отражает движение космических объектов в *эфемеридном времени* —равномерной шкале времени) о параметрах движения навигационных ИСЗ. Навигационная аппаратура потребителя на подвижном объекте, положение которого необходимо определить, принимает навигационные радиосигналы от радиовидимого с поверхности Земли навигационного ИСЗ в течение сеанса навигации продолжительностью 5 … 15 мин. На основе принятой ЭИ и в результате измерения приращений дальности до навигационного ИСЗ, НАП определяет (уточняет) две горизонтальные координаты объекта на поверхности земного эллипсоида. В состав СРНС также входят *навигационные космические аппараты* (НКА), расположенные на круговых орбитах с высотой порядка 20 000 км над поверхностью Земли.

Благодаря использованию атомных стандартов частоты на НКА в системе навигационной аппаратуры обеспечивается взаимная синхронизация навигационных радиосигналов, излучаемых орбитальной группировкой НКА. В НАП на подвижном объекте в сеансе навигации применяются радиосигналы не менее чем от четырех радиовидимых НКА, с помощью которых измеряются три разности дальностей и три разности скоростей объекта относительно четырех НКА. Результаты измерений и ЭИ, принятые от каждого НКА, позволяют определить (уточнить) три декартовы координаты и три составляющие вектора скорости подвижного объекта, а также определить смещение *шкалы времени* (ШВ) объекта относительно ШВ системы.

Рассмотрим возможности системы ГЛОНАСС по определению координат подвижных объектов на земной поверхности и погрешностей их измерений. На рис. 1.36 приведена схема расположения среднеорбитальной группировки ИСЗ в системе ГЛОНАСС. Полная орбитальная группировка (ОГ) содержит 24 штатных навигационных спутника на круговых орбитах с наклонением в 64,8° в трех орбитальных плоскостях (вид со стороны полюса Земли) по восемь ИСЗ в каждом из них, составляющих определенные углы с плоскостью экватора. При навигации каждому объекту обеспечивается прием сигналов от четырех навигационных космических аппаратов (один ИСЗ высокий и три низких). Номинальный период обращения НКА равен 11 ч 15 мин 44 с, при этом высота круговой орбиты H = 19 100 км над поверхностью Земли. Основными геометрическими характеристиками орбитальной группировки (ОГ) штатных НКА в СРНС, от которой зависит точность навигации потребителей в СРНС, являются геометрические свойства созвездия НКА, которое «видит» наземный потребитель. В системе ГЛОНАСС каждый штатный НКА в ОГ постоянно излучает шумоподобные непрерывные навигационные радиосигналы в диапазонах частот 1 600 МГц и 1 250 МГц. Для навигационных радиосигналов информация в цифровом виде формируется на борту НКА на основе данных, передаваемых от *наземного комплекса управления* (НКУ) системы на борт НКА с помощью радиотехнических средств.

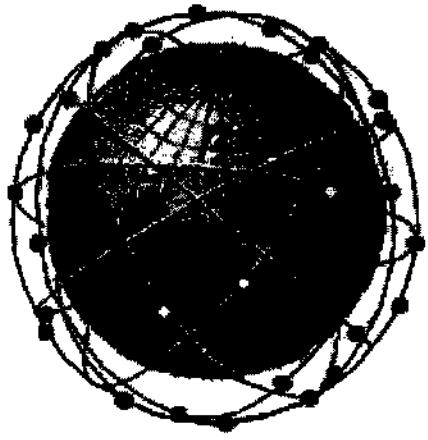


**Рис. 1.36.** Схема расположения спутников в системе ГЛОНАСС

Навигационные сигналы, излучаемые штатными НКА, образуют в околоземном пространстве радионавигационное поле. В системе ГЛОНАСС каждый штатный НКА излучает навигационные сигналы в сторону Земли с помощью передающих антенн, рабочая часть диаграммы направленности которых имеет ширину $2\varphi_0 = 38$ °. Это означает, что при полной ОГ радионавигационное поле на высотах $h_0 = 2\,000$ км «покрывает» диск Земли с запасом, т. е. потребитель в любой точке этого пространства «освещается» радиолучами не менее чем от четырех штатных НКА, образующих по отношению к нему удовлетворительное по геометрическому фактору созвездие для оперативного автономного определения координат и вектора скорости. Дальность от приемной антенны, размещенной на поверхности Земли, до околозенитного (угол возвышения $\beta = 90$ °) НКА составит $D = H = 19\,100$ км, до пригоризонтного ($\beta = 5$ °) НКА — $D = 24\,000$ км.

Создание СНРС GPS и ГЛОНАСС знаменует значительный прогресс в глобальной навигации приземных подвижных объектов — сухопутных, морских, воздушных и низкоорбитальных космических. Применяются такие системы и для выполнения других специальных задач.

Корреляционно-экстремальные системы навигации. Совершенно новым, бурно развивающимся направлением науки и техники является *навигация по физическим полям с использованием методов радиолокации.* Системы навигации, использующие подобный метод, относятся к автономным. *Автономные системы навигации* отличаются тем, что информация о положении объекта (часто это летательный аппарат) относительно требуемой траектории движения извлекается из параметров заранее известного физического поля после обработки выходных сигналов бортовых датчиков. Автономные системы навигации обычно реализуют корреляционно-экстремальный (здесь *корреляция* — связь) принцип работы, основанный на сравнении изображений земной поверхности или совокупности ориентиров текущего изображения с эталонным, полученным заранее, и поэтому носят название *корреляционно-экстремальных систем навигации* (КЭСН). Упомянутые изображения формируются с помощью различных искусственных и естественных физических полей: магнитных и гравитационных полей Земли; изображения поля рельефа участка поверхности Земли в оптическом, тепловом, инфракрасном диапазонах и радиодиапазонах волн; поля радиолокационного контраста.

*Магнитные и гравитационные поля* создаются соответственно магнитными породами и массой Земли, рельеф которой как навигационное поле обладает высокой стабильностью и хорошими картографическими параметрами. Использование этих полей в КЭСН основано на взаимно однозначном соответствии распределения параметров поля конкретной географической местности Земли аналогично тому, как каждый человек обладает неповторимыми отпечатками пальцев. Измерение параметров этих полей производится активным зондированием с учетом барометрических данных. *Оптическое поле земных покровов* формируется в видимом диапазоне излучений и характеризуется яркостными и спектральными характеристиками отражения. *Тепловые поля Земли в инфракрасном, сантиметровом и миллиметровом диапазонах волн* зависят от термодинамической температуры и физико-химических свойств ее поверхностного слоя. Датчиками теплового поля служат радиометры. *Поле радиолокационного контраста* характеризуется значениями удельных ЭПР земных покровов. Поле достаточно эффективно и при больших высотах над относительно ровной местностью, когда рельефометрические системы не обеспечивают заданной точности. Карту удельных ЭПР интерпретируют как радиолокационное изображение местности, полученное РЛС с высоким разрешением. При применении широкого луча диаграммы направленности реальному радиолокационному изображению соответствует карта удельных ЭПР поверхности с размытыми контурами. Использование поля в радионавигации — одно из перспективных направлений.

Упрощенно принцип действия КЭСН можно пояснить следующим образом. В бортовую память летательного аппарата вводятся эталонные изображения (их может быть 2 … 20) различных участков поверхности Земли на предполагаемой траектории полета. Источником измерительной информации является бортовая система, регистрирующая реальные изображения под летательным аппаратом. В бортовом коррелометре над участком коррекции происходит сравнение текуще-

го изображения с эталонным, выбранным из бортового устройства памяти в предполагаемых координатах местоположения. Результатом сопоставления является определение наиболее близкого к наблюдаемому изображению эталона и осуществление по нему точной привязки текущих координат летательного аппарата к местности.

Отличительной особенностью КЭСН является наличие мощной и развитой памяти для хранения эталонной информации, которая обычно представлена в цифровом виде. Для этого исходная информация о поле преобразуется в цифровые карты. Поэтому эталоном в общем случае является *детальная цифровая радиолокационная карта* местности.

### Системы радиотелеуправления и радиотелеметрии

*Телеуправлением* называют раздел телемеханики, разрабатывающий методы и средства передачи команд для управления объектами на расстоянии (дистанционного управления объектами). Телеуправление широко применяется в военном деле для управления полетом ракет, беспилотных самолетов, космических объектов и др. Телеуправление объектом по радиоканалу называют *радиотелеуправлением*. В этом случае управление объектом осуществляется по командам управления радиотехнической системы или используется возможность наведения объекта по радиолучу. Для телеуправления используют как радиоканалы, так и проводные и оптические линии связи, при этом ведут контроль за выполнением всех команд.

Системы радиотелеуправления обеспечивают движение летательных аппаратов по заданной траектории в автоматическом режиме. Такие системы радиотелеуправления широко применяют для решения сложных проблем дистанционного управления полетом космических объектов. Пример коррекции параметров околоземной орбиты *космического корабля* (КК), осуществляемой с помощью системы радиотелеуправления, показан на рис. 1.37.

На траекторию движения КК в пространстве в основном действует сила взаимного тяготения его массы и массы Земли. Очевидно, что требуемые значения параметров орбиты КК не могут быть реализованы абсолютно точно, поэтому реальная орбита всегда имеет некоторое отклонение от расчетной. Измерение параметров реальной орбиты, выработку команд на коррекцию ее траектории, а также контроль над работой многочисленных блоков и устройств КК выполняет система радиотелеуправления. Она объединяет в своем составе *Центр управления полетом* (ЦУП) и наземные или морские *измерительные пункты*. Все технические звенья и устройства, входящие в систему радиотелеуправления, связаны линиями радиосвязи, обеспечиваю-

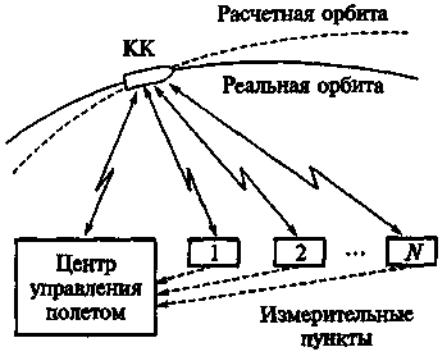



**Рис. 1.37.** Структурная схема системы радиотелеуправления

щими их взаимодействие как между собой, так и с устройствами КК. Данные о реальной орбите КК, полученные всеми измерительными пунктами, передаются по линиям радиосвязи в ЦУП. Там они обрабатываются на мощном компьютере совместно с данными других измерений и сопоставляются с расчетными параметрами. Если возникли отклонения от расчетной орбиты, то вырабатываются радиокоманды по коррекции движения КК. По этим командам происходит включение двигателей бортовой установки, корректирующей направление движения КК в космическом пространстве.

В последние годы из широкого спектра систем радиотелеуправления, нацеленных в основном на решение различных задач военного назначения, выделился новый класс — мультимедийные системы, выполняющие автоматически под контролем компьютеров сложные функции передачи, приема, обработки и хранения различной информации.

*Радиотелеметрия* — область науки и техники, позволяющая с помощью электромагнитных волн получать на расстоянии данные о работе и состоянии людей и механизмов и о протекании физических или технологических процессов. Система радиотелеметрии должна воспринимать изменение какого-либо параметра, преобразовывать его в электрический сигнал, модулировать им несущую частоту и передавать на определенное расстояние радиоприемному устройству для распознавания и анализа сигналов. Системы радиотелеметрии применяются для контроля за протеканием процессов на движущихся и труднодоступных объектах, например при испытаниях самолетов, космических ракет и кораблей, исследованиях атмосферы с помощью радиозондов, передаче сведений о погоде автоматическими метеорологическими станциями, а также для контроля работы автоматизированных установок.

### Радиоастрономические системы

*Радиоастрономия* — наука, изучающая с помощью электромагнитных волн космические объекты — звезды, планеты, реликтовое радиоизлучение Вселенной (оставшееся с момента *Большого взрыва* — около 15...20 млрд лет назад). Различают активную и пассивную радиоастрономию.

Системы активной радиоастрономии содержат те же части, что и системы активной радиолокации (см. рис. 1.33). Объектами изучения являются относительно близко расположенные планеты Солнечной системы. Устройства пассивной радиоастрономии (радиотелескопы) аналогичны устройствам пассивной радиолокации (см. рис. 1.34). С их помощью изучают Солнце, объекты, расположенные далеко за пределами Солнечной системы, а также реликтовое радиоизлучение. Радиоастрономические исследования дают ценные сведения о природе космических тел и о происходящих в них процессах.

### Мультимедийные системы

*Мультимедийными системами* называют объединенные радиотехнические средства передачи и обработки информации, быстродействующие компьютеры с мощным программным обеспечением и широкополосные и высокоскоростные

линии связи. Мультимедийная система способна объединять в единый телеуправляемый комплекс (пакет) графику, звук и видеоизображения, обрабатывать их, передавать и принимать в реальном масштабе времени (*real time*). В общем случае технология организации обмена информацией между такими системами осуществляется через сети Internet.

Системы мультимедийного направления в последние годы бурно развиваются и открывают новые возможности развития в глобальном масштабе мировой культуры, искусства, медицины, биологии, обмена информацией и др. Повышение скорости передач информации в радиорелейных и волоконно-оптических сетях увеличивает оперативность функционирования таких систем. На этой основе создают перспективные виды радиотехнических систем так называемой *виртуальной реальности*, открывающие новые возможности систем связи в городах. Одна из них — *система виртуального* (т. е. без физического присутствия человека) осмотра различных объектов и помещений, предназначенная для обслуживания потребителей информации. Основу такой системы составляют скоростные компьютеры с необходимым программным обеспечением вместе с проводными, волоконно-оптическими, радиорелейными линиями связи и линиями с выходом на ИСЗ.

Становится возможным в реальном масштабе времени непосредственно из квартир осматривать выставленные на продажу различные объекты и здания, получать характеристики изделий торговых предприятий, осматривать товары в магазинах, изучать цены, посещать библиотеки, музеи и выставки в разных странах, а также проводить видеоконференции, выполнять вычисления на удаленных компьютерах, осуществлять дистанционное обучение в других институтах и странах.

### *Современные системы подвижной радиосвязи*

Одной из наиболее быстро развивающихся отраслей связи сегодня является подвижная связь, использующая как наземные, так и спутниковые системы радиосвязи. Особенно быстрыми темпами как в мире, так и у нас в России идет развитие сетей сотовой радиосвязи. К 2008 г. количество людей, пользующихся мобильными телефонами, превысило 850 млн. По числу абонентов системы мобильной связи уже можно судить об уровне и качестве жизни в данной стране. Пока доля населения, пользующегося сотовой связью в России, составляет около 7...10 %, в Финляндии для сравнения — 75 %. Однако темпы роста абонентов мобильной связи в России (почти 200 % в год) вселяют оптимизм.

По классификации МСЭ системы подвижной связи относятся к системам беспроводного доступа абонентских линий. Характерным признаком любых систем беспроводного доступа абонентских линий является наличие радиоканала на абонентском участке.

На первом этапе развития радиотехники в начале XX в. радиосвязь развивалась как морская подвижная связь. В те годы этот вид связи являлся единственно возможным для организации связи судов между собой и с берегом. Фирмой «Маркони» в Великобритании, а затем и на предприятиях других стран (России, США, Франции и Германии) было организовано производство судовых искро-

вых радиостанций. В частности, еще до 1904 г. более пятидесяти судов военно-морского флота России было оснащено судовыми радиостанциями.

Заметим, что нормальное функционирование системы связи предусматривает обмен информацией в любых службах электросвязи (*электросвязью* называется передача сообщений посредством электрических сигналов), который должен осуществляться по определенным, заранее оговоренным правилам (*стандартам*). Сейчас эти правила разрабатываются рядом международных организаций электросвязи. Широкое внедрение средств судовой подвижной связи, существенно повышающей безопасность плавания, обусловило необходимость принятия международных правил радиообмена и стандартов на средства морской радиосвязи. Такие правила и стандарты были приняты на первой Международной конференции по радиосвязи в Берлине в 1903 г.

Потребности в средствах наземной подвижной связи для оперативного управления действиями полиции привели в 1921 г. к созданию в США первой диспетчерской системы телеграфной подвижной связи. По сути это была система пейджинговой связи, так как имела однонаправленное действие и служила для передачи распоряжений дежурным бригадам полиции. На начальном этапе развития систем наземной подвижной связи в них использовали телеграфные режимы работы, а позже — телефонные режимы с применением для передачи сообщений *амплитудной модуляции*. В 1940 г. в США в диапазоне ОВЧ создана первая система подвижной связи с использованием *частотной модуляции* (ЧМ; см. гл. 2). В 1948 г. в США создана первая полностью автоматическая радиотелефонная система подвижной связи. В СССР серийный выпуск первых систем подвижной радиосвязи был налажен в 1952 г. (заметим, что фактически история сотовой связи насчитывает уже более 40 лет, и хотя в настоящий момент в Российской Федерации распространены в основном западные разработки, первый сотовый телефон был сконструирован в начале 70-х гг. XX в. Воронежским научно-исследовательским институтом связи — ВНИИС).

Эффективность систем наземной подвижной связи для управления в службах безопасности (полиции, пожарной службе, скорой помощи и т.п.), для управления работой транспорта и в других областях приводит к быстрому прогрессу в этой области.

Системы подвижной (часто говорят — мобильной, относя это к сотовым системам связи) радиосвязи (СПР) обеспечивают одновременно связью большое число мобильных абонентов, местоположение которых на определенной территории произвольно. Поэтому практически все СПР построены по методу многостанционного (в англоязычной и др. литературе часто множественного) доступа.

Принципы организации многостанционного доступа напоминают принципы уплотнения каналов. В теории связи под *многостанционным доступом* (МД) понимают возможность обращения к одной *базовой приемопередающей* (БПС; от англ. — *Base Transceiver Station — BTS*) или спутниковому ретранслятору нескольких *мобильных станций* (МС; абонентский терминал; радиотелефон; мобильный телефон; сотовый телефон; англ. *Mobile Station* — MS), при которой последние могут одновременно передавать и получать через нее информацию.

Эффективность методов МД в основном оценивается по пропускной способности (емкости), быстродействию, используемому частотному ресурсу и некоторым другим показателям систем связи. В данном случае под емкостью любой системы связи понимают число различных каналов — радиосвязных, телефонных, телевизионных, передачи цифровых данных и прочее, либо, в более общем виде, для цифровых систем, число бит в секунду которое можно передать через данную сеть.

Проблема выбора наилучшего метода МД заключается в нахождении базиса (ансамбля) ортогональных сигналов, при которых обеспечиваются оптимальные параметры и характеристики системы подвижной радиосвязи. В радиотехнике и теории передачи информации формирование базисов ортогональных сигналов основано на разделении сигналов по частоте, времени и форме. В соответствии со способами формирования базисов ортогональных сигналов различают три основных метода организации МД.

Многостанционный доступ с частотным разделением каналов (МДЧР; англ. — *Frequency Division Multiple Access — FDMA*; аналог ЧРК) является наиболее простым по организации работы, при котором каждая подвижная станция работает в некоторой полосе частот на отведенном участке спектрального диапазона. Между рабочими полосами соседних каналов предусмотрены небольшие защитные частотные интервалы, позволяющие с требуемой точностью разделить принимаемые сигналы различных подвижных станций. Однако в любой стране используемый частотный спектр — уникальный стратегический запас, и это не восполняемый государственный ресурс.

Многостанционный доступ с временны́м разделением каналов (МДВР; англ. — *Time Division Multiple Access — TDMA*; аналог ВРК) получил применение в системах подвижной связи из-за ограниченности специально выделенного странам и регионам частотного спектра. При таком доступе ортогональность сигналов в подвижных станциях достигается выделением каждой из них для излучения или приема сигналов определенного, периодически повторяемого временно́го интервала — *TDMA*-кадра. Длительность кадра в основном определяется сетевым трафиком. Интервалы излучения сигналов БПС и подвижных станций взаимно синхронизированы, что исключает их временно́е перекрытие. Первая спецификация технологии TDMA была разработана специалистами американской Ассоциации телекоммуникационной промышленности (*Telecommunications Industry Association* — TIA) в 1988 г. в соответствии с «Техническими требованиями пользователей» (UPR) — документом, изданным Ассоциацией промышленности сотовой связи — (CTIA), в котором определено техническое содержание систем подвижной связи 90-х гг. XX в. Эта спецификация в том же году была опубликована как стандарт IS-54 на многостанционный доступ с временны́м разделением каналов систем связи. Для повышения пропускной способности сети связи TDMA, как правило, используется совместно с частотным разделением каналов.

Многостанционный доступ с кодовым разделением каналов (МДКР; англ. — *Code Division Multiple Access — CDMA*) основан на

использовании *широкополосных* или *шумоподобных* (*pseudonoise*) сигналов (обоим вариантам термина соответствует аббревиатура ШПС). В системах мобильной радиосвязи стандарта CDMA используются все преимущества и частотного и временно́го разделения каналов. Во-первых, сигнал имеет большую длительность, распределен во времени, и поэтому пиковая излучаемая мощность много меньше, чем при FDMA и TDMA, хотя средняя мощность одинакова. Во-вторых, при сигналах большой длительности нет крутых и мощных фронтов импульсов излучения. Третье преимущество — системы стандарта CDMA позволяют вводить множество кодовых комбинаций, обеспечивая почти идеальную скрытность переговоров и помехозащищенность. Однако реализация кодового метода многостанционного доступа к каналу радиосвязи сопряжена с определенными техническими трудностями, связанными как с проблемой синхронизации в работе всех абонентских станций, так и выравниванием по мощности сигналов, принимаемых базовой станцией с целью исключения подавления слабого сигнала сильным.

Отметим, что методы разделения *одномерных сигналов* уже рассматривались. В спутниковых и других радиотехнических информационных системах используются еще и различные методы *многостанционного доступа с пространственным разделением* (МДПР; *Space-Division Multiple Access* — SDMA) по направлению прихода радиоволн (в частности, применяют двулучевую приемную антенну, к которой подключены два приемника с одинаковыми полосами частот, что позволяет осуществлять одновременный доступ к спутнику из двух разных точек на Земле) и их *пространственной поляризации* (*polarization-division multiple access* — PDMA).

Существующие сейчас системы подвижной связи можно разделить на пять больших групп:

- системы сотовой подвижной связи (ССПС);
- профессиональные системы подвижной связи (ПСПС);
- системы персонального радиовызова (СПРВ), или пейджинговые (от англ. *paging* — письменное сообщение) системы;
- системы подвижной спутниковой связи (СПСС);
- системы беспроводных телефонов (СБТ).

Все перечисленные системы подвижной связи построены на основе сотовой концепции и работают по определенным *протоколам*.

### Системы сотовой подвижной (мобильной) связи

В 1946 г. исследовательская лаборатория *Bell Laboratories* (компания *AT&T* г. Сент-Луис, штат Миссури, США) создала первую сеть мобильной связи. Это была простейшая шестиканальная (т. е. с шестью несущими частотами) система связи с одной базовой приемопередающей станцией для передачи и приема абонентских сообщений. Эта сеть связи строилась по так называемому принципу *неизбежности* или *фатальности*: на самый высокий небоскреб в городе подвесили (установили) антенну, к которой подсоединили передатчик большой мощности. Масса первого радиотелефона составляла 30 кг, и для работы он требовал

наличия у подвижного абонента аккумулятора большой емкости и генератора постоянного тока, поэтому «мобильники» устанавливались в автомобилях. Переключение абонента между каналами связи, в поисках свободного, осуществлялось вручную. Радиопередатчик позволял пассажирам или водителю связаться с АТС и таким образом совершить звонок. При этом телефонное общение было сложным (симплексным) — нельзя было и слушать и говорить одновременно. Так, чтобы донести свое сообщение до собеседника, нужно было нажать и удерживать кнопку телефонной трубки, а чтобы услышать ответ, кнопку надо было отпустить. Чтобы позвонить на радиотелефон, приходилось сначала звонить на телефонную станцию и затем сообщать номер оператору. Всего такая «первобытная» система связи поддерживала 23 пользователя одновременно и предназначалась для бизнесменов, переезжающих из г. Нью-Йорка в г. Бостон.

Но инновационная идея *Bell Laboratories* не прижилась — слишком дорого выходило пользование услугами мобильной связи. Впрочем, зерно было посеяно. Поскольку данной системе связи был отведен ограниченный частотный ресурс, то повышение количества обслуживаемых абонентов требовало пропорционального увеличения числа несущих частот базовой станции. А для связи был выделен диапазон с фиксированными частотными каналами. Если в одно время используются близкие по частоте каналы связи, то общаться с помощью телефонов практически невозможно.

Первая задача успешно решалась по мере бурного развития элементной базы, в частности создания биполярных транзисторов. Проблему эффективности использования ограниченного частотного ресурса удалось решить путем разработки сотовой концепции системы связи. Идея сотового принципа организации сетей подвижной связи была выдвинута в 1947 г. сотрудником лаборатории *Bell Laboratories* Д. Рингом (*D. Ring*). Концепция сотовой связи оказалась простой. Вся обслуживаемая зона (территория) связи разбивается на *соты* — ячейки (в идеале — правильные шестиугольники; топология такой сети напоминает пчелиные соты — от англ. *cell* — откуда и пошло сегодняшнее название сотовых телефонов) с повторным использованием частот (англ. *frequency reuse*) в каждой из них (рис. 1.38). Это значительно повышало эффективность частотного диапазона, что в свою очередь увеличивало емкость системы. В центре каждой ячейки устанавливается маломощная базовая приемопередающая станция с одной или некоторым определенным набором несущих частот (каналов связи), достаточным для установления абонентской связи согласно предполагаемому трафику. Базовые приемопередающие станции с помощью проводной, радиоканальной связи или волоконно-оптической линии связи подключаются к выходу сотового терминала, который соединен с телефонной сетью общего пользования.

Через двадцать лет данная идея нашла свое воплощение в сотовых сетях подвижной радиосвязи общего пользования. Внедрение подвижных сетей радиосвязи начинается с 70-х гг. в XX в., вначале в США, а позже в западноевропейских странах, в Японии и в других регионах мира. Благодаря их созданию новые услуги подвижной радиосвязи стали доступными для сотен миллионов людей многих стран мира.

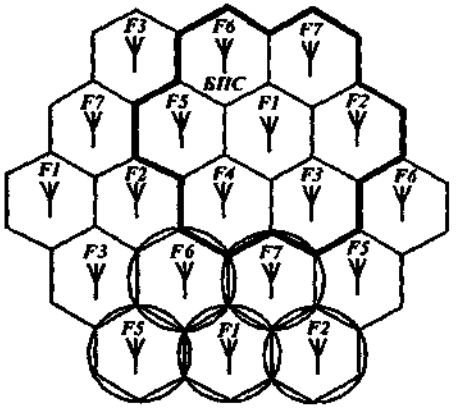



**Рис. 1.38.** Построение сотовой системы подвижной связи

Отметим, что физически в сотовых сетях связи радиопокрытие какой-либо территории осуществляют ячейками, антенны БПС которых имеют круговые диаграммы направленности. И тем не менее реально связь осуществляют фактически по сотовой модели. Дело в том, что пересечение соседних окружностей происходит по хордам, которые в идеале и образуют шестигранные ячейки — соты (см. круги и шестигранники на рис. 1.38). В связи с тем, что любая ячейка имеет небольшой радиус действия, допустим 1...5 км, одна маломощная базовая станция будет уже обслуживать меньшую территорию, и поэтому ее мощность (как и мощность мобильной станции) может быть существенно снижена. Реально мощность каждой базовой станции может быть уменьшена в десятки и сотни раз, однако их суммарная мощность, естественно, велика и соизмерима с мощностью одной крупной БПС, которая обслуживала бы ту же территорию. Заметим, что наряду с информативными сигналами БПС излучает так называемые *пилот-сигналы* — специальные немодулированные или иные колебания. Измеряя и сравнивая пилот-сигналы от разных БПС, МС выбирает наибольший.

БПС с круговой диаграммой направленности антенн осуществляет передачу сигнала одинаковой мощности практически по кругу, что для абонентских станций в соседних сотах эквивалентно приему помех со всех направлений. В этом случае особенно мешающее действие приему сигналов оказывают взаимные помехи по совпадающим частотным каналам — *соканальные помехи.* Для избежания воздействия соканальных помех соты с одинаковым набором несущих частот перемежают буферными сотами с другим набором частот. Группа сот в зоне обслуживания с различными наборами частот называется *кластером*, а число частот в наборе — размерностью кластера. На рис. 1.38 жирными линиями выделена сотовая структура с размерностью кластера $n = 7$.

Чтобы снизить общий уровень интерференционных помех от соседних сот и абонентских устройств, а также помех от посторонних источников ЭМИ, на базовых станциях используется многосекционная направленная антенна, позволяющая делить общее пространство радиоперекрытия на отдельные сектора. Антенна БПС с секторной ДН излучает практически всю энергию передаваемого сигнала в заданном направлении, а уровень боковых излучений сокращается до минимума. Секторное построение антенн БПС позволяет многократно применять набор частот при одновременном снижении уровня соканальных помех.

В зависимости от числа действующих в ячейке абонентов, нагрузки и электромагнитной обстановки на местности используются антенны различной конфигурации и размеров. Наибольшую емкость обеспечивает сотовая модель сис-

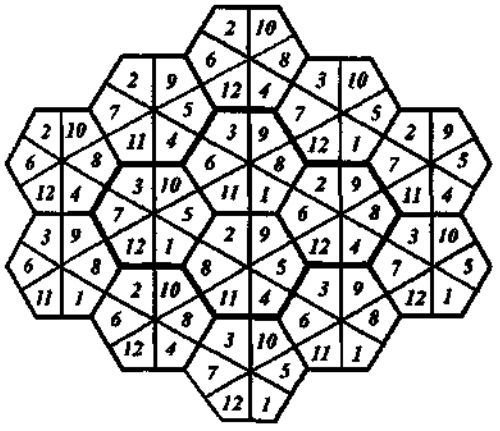


**Рис. 1.39.** Модель ССПС с двенадцатью группами частот

темы радиосвязи, содержащая четыре БПС с шестью 60-градусными антеннами (рис. 1.39). Из структурной схемы системы данной модели следует, что каждая частота используется дважды в зоне, состоящей из четырех БПС (четыре соты выделены жирной линией). Благодаря такой модели построения каждая из четырех БПС в пределах зон действия шести 60-градусных антенн в одной ячейке может работать на двенадцати группах частот ($n$ = 12). Все сотовые системы связи с повторным использованием набора частот разрабатывались с учетом важного требования — координаты местоположения мобильного абонента заранее неизвестны и непредсказуемы в пределах заданной зоны обслуживания данной сети. Благодаря высокоточной автоматической регулировке коэффициента усиления выходных усилителей мощности передатчиков БПС, эффективность секторного перекрытия близка к 100%.

Одной из основных проблем при разработке систем сотовой связи является обеспечение непрерывной связи во время передвижения абонента в зоне обслуживания. Для ее решения сотовая концепция включает в себя принцип «эстафетной передачи» (*hand off* — сопровождение или переброску; *handover* — хэндовер) переговорных сигналов из ячейки в ячейку, вследствие чего абонент может вести разговор, свободно пересекая границы сот, автоматически переключаясь с одной БПС на другую. Современные хэндоверы бывают двух типов:

- внешний — когда меняется БПС, через которую идет связь с сетью;
- внутренний — когда во время разговора меняется канал приема/передачи.

Обычно внедрение сотовой сети связи начинается с развертывания небольшого числа крупных сот с радиусом действия 1...35 км, получивших название *макросот*. Когда нагрузка в ячейке достигает уровня, при котором существующего числа каналов недостаточно (вероятность непредоставления канала абоненту становится более 5 %), эта сота разделяется на более мелкие с пониженной мощностью передатчиков БПС и МС. При этом макросотовая структура постепенно трансформируется в сеть с более мелкими сотами (*микросоты*) с большим их числом и радиусом действия до 1 000 м, а пропускная способность сети на территории региональной ячейки возрастает в число раз, равное числу вновь созданных сот. Такой способ преобразования сотовых сетей связи называют *расщеплением*. В этом случае мощность радиопередатчиков базовых станций уменьшается еще больше. Этот способ разделения повторяется, пока сеть не достигнет расчетного значения пропускной способности. Старое оборудование при расщеплении сохраняется полностью, меняется только мощность источни-

ков излучения. Микросоты предназначаются для трафика, отражающего медленно передвигающихся на небольшие расстояния или стоящих абонентов, находящихся на улицах, в помещениях, вокзалах, аэропортах.

Принципы построения микросотовых и макросотовых сетей существенно отличаются. Создание небольших сот приводит к сложной проблеме, когда абонент в быстро движущемся транспорте в течение одного сеанса связи проходит через несколько ячеек. Это вызывает рост числа переключений между БПС, а значит, необходимость разработки быстродействующих алгоритмов переключения при «эстафетной передаче» абонента. В этом случае непрерывность связи обеспечивается способностью системы МС передавать связь тем БПС, в зонах которых он оказывается в данный момент. Центр коммутации системы на основе непрерывных измерений сигналов БПС, ближайших к движущемуся мобильному абоненту, определяет момент его пересечения границы двух сот. После этого центр переключает разговорный канал из первой ячейки во вторую за столь короткое время, что сохраняется непрерывность разговора. Однако это требует сложного алгоритма определения номера той ячейки из нескольких соседних, куда въезжает движущийся транспорт (входит человек), и сложных схемотехнических решений. Этот же алгоритм освобождает канал в первой ячейке и осуществляет поиск канала и установление по нему связи во второй, соседней.

Второе отличие связано с трудностями прогнозирования условий распространения радиоволн на небольших обслуживаемых системой связи территориях. Для этого требуются электронные карты местности, топография структур улиц, строений и т. д. Если в какой-либо ячейке или группе сот трафик начинает существенно превышать расчетное значение, ее разделяют на ряд более мелких ячеек — *пикосот* — с радиусом обслуживания 10...100 м и пониженной мощностью передатчиков БПС. При этом пропускная способность сети увеличивается в число раз, равное числу вновь образованных пикосот. Как правило, при микро- и пикосотовой структурах построения сети надобность в применении эстафетной передачи абонента и многократном использовании частот отпадают.

Бурное развитие современных радиоэлектронных и «связных» технологий позволило осваивать новую концепцию построения ССПС, связанную с использованием в БПС *интеллектуальных антенных систем*, автоматически перестраивающих свои диаграммы направленности на мобильные станции. На практике это стало возможно с внедрением ФАР и адаптивных (интеллектуальных) антенн, разработанных специально для цифровых систем подвижной связи. Наиболее эффективными оказались адаптивные ФАР, реализующие максимальный коэффициент усиления антенны в направлении ведущего переговоры мобильного абонента и обеспечивающие минимальный уровень соканальных помех в приемнике. Интеллектуальная ФАР состоит из ряда элементарных излучателей, объединенных микропроцессором с амплитудными и фазовыми анализаторами принимаемых радиосигналов. По результатам анализа амплитудных и фазовых соотношений сигналов, поступающих на элементарные излучатели от мобильной станции, сигнальный процессор определяет направление оптимального приема и формирует требуемую диаграмму направленности решетки.

Первоначально развитие получили аналоговые системы (стандарты) сотовой связи: так называемое *первое поколение*, или 1G (от англ. *First Generation*). К ним относятся североамериканский стандарт AMPS, скандинавский стандарт NMT-450 (первая сеть, внедренная в Российской Федерации; Москва и Санкт-Петербург — 1991 г.) и ряд др. Следующим этапом развития ССПС стало создание цифровых систем второго поколения (2G): в США — D-AMPS и общеевропейский стандарт GSM.

Знаменательной вехой в развитии систем сотовой подвижной связи является 1989 г., когда фирмой «Qualcomm» (США) была завершена разработка новой цифровой системы второго поколения, использующей технологию CDMA. Эта технология в несколько раз повышала эффективность использования спектра в сотовой связи и позволяла создавать сети весьма большой емкости. В странах Западной Европы, в которых распределение полос частот между разными службами существенно отличается от стран американского континента, сети на основе этой технологии не создавались. В них происходило интенсивное развитие сотовых сетей стандарта GSM. В России в 1997 г. на основе технологии CDMA начали создаваться сети абонентского доступа.

В настоящее время в России в основном применяются зарубежные ССПС четырех стандартов (цифры обозначают диапазон рабочих частот):

- аналоговые NMT-450 и NMT-900 (*Nordic Mobile Telephone* — Скандинавская система подвижной телефонной связи, диапазоны 450 и 900 МГц);
- аналоговая AMPS (*Advanced Mobile Phone System* — Перспективная подвижная телефонная система, диапазон 800 МГц);
- цифровые GSM-900, GSM-1800 и два его варианта — DCS-1800 (*Digital Cellular System*) и PCS-1900 (Personal Communication Service); GSM — от названия группы *Groupe Spécial Mobile* — Глобальная система подвижной связи; в 1991 г. аббревиатура GSM приобрела иную трактовку — *Global Standart for Mobile Communications* — Глобальный стандарт для подвижной связи;
- цифровая D-AMPS (*D — digital* — цифровая — IS-54; диапазоны 900, 1800 и 1900 МГц) и сеть CDMA фирмы Qualcomm (диапазоны 800 и 1 900 МГц).

Система CDMA по сравнению с GSM обеспечивает более высокое качество связи, меньшие энергетические затраты, но сложна в построении.

Первые две цифровые системы подвижной радиосвязи базируются на комбинировании метода многостанционного доступа с частотным разделением каналов с методом многостанционного доступа с временны́м разделением каналов при частотном дуплексном разносе прямых и обратных каналов радиосвязи, последняя система (CDMA) — на многостанционном доступе с кодовым разделением каналов.

Стандарт NMT-450 нашел широкое распространение в Северной и Центральной Европе (Швеции, Финляндии, Бельгии, Российской Федерации, Австрии, Венгрии, Турции и других странах), однако в последние годы его дальнейшее внедрение практически заморожено.

Все эти системы подвижной связи используют модели сот с радиусом действия от 0,1 до 35 км.

### Сотовые системы подвижной связи стандарта GSM

Система GSM относится ко второму поколению сетей сотовой связи, в которой использовано комбинирование методов многостанционного доступа с частотным и временны́м (это основной метод) разделением каналов (кстати, об этом редко упоминается), и представляет собой цифровую систему связи с программным управлением. В ней использованы многоуровневая модель ВОС, пакетная система сигнализации и принципы построения интеллектуальной сети, и в частности отделение функций собственно коммутации вызовов от предоставления услуг. Элементы системы способны контролировать все основные характеристики сигнала в процессе его передачи, а также устранять обнаруженные неисправности и выполнять множество функций по обслуживанию сети (модифицирование своей конфигурации, слежение за местом расположения объекта, обеспечение функции эстафетной передачи и защиты передаваемой информации, оценивание мощности несущей к помехе).

Стандарт GSM обладает рядом специфических услуг сотовой связи. К ним относятся:

- использование SIM-карты (модуль подлинности абонента) для доступа к каналу и услугам связи;
- закрытый для подслушивания радиоинтерфейс;
- шифрование передаваемых сообщений;
- аутентификация абонента и идентификация абонентского оборудования по криптографическим алгоритмам;
- автоматический роуминг (*Roaming* — «блуждание»; *автоматическое подключение* абонентов к *местной сети связи* GSM при его перемещении в другую зону обслуживания; обычно при перемещении в другую страну);
- применение «Службы коротких сообщений» (*Service of short messages* — SMS) — передача с телефона на телефон коротких текстовых сообщений.

Также система GSM предоставляет пользователям следующий набор услуг: вызов спецслужб (скорой помощи, полиции, пожарной службы по номеру 112 на европейском континенте). Система сотовой связи стандарта GSM работает в диапазонах 890...915 МГц для передатчиков мобильных станций (линия передачи «вверх», т. е. к базовым станциям) и 935...960 МГц для передатчиков базовых станций (линия передачи «вниз» — т. е. к мобильным станциям).

Ширина полосы канала связи составляет 0,2 МГц, что позволяет обеспечивать 124 канала связи в отведенном частотном ресурсе. Дуплексный разнос частот передачи и приема одного канала связи равен 45 МГц. Максимальная дальность радиосвязи или радиус ячейки сотовой структуры составляет 35 км, минимальная — 50 ...75 м.

Архитектура сотовой сети связи состоит из трех основных компонентов (рис. 1.40) — сотовых телефонов, базовых станций и сетевой подсистемы. Сеть имеет в своем составе *административный центр* (*Administration Center — ADC*), в котором распложены административно-управленческие структуры. *Центр управления сетью* (*Network Management Center — NMC*) обеспечивает

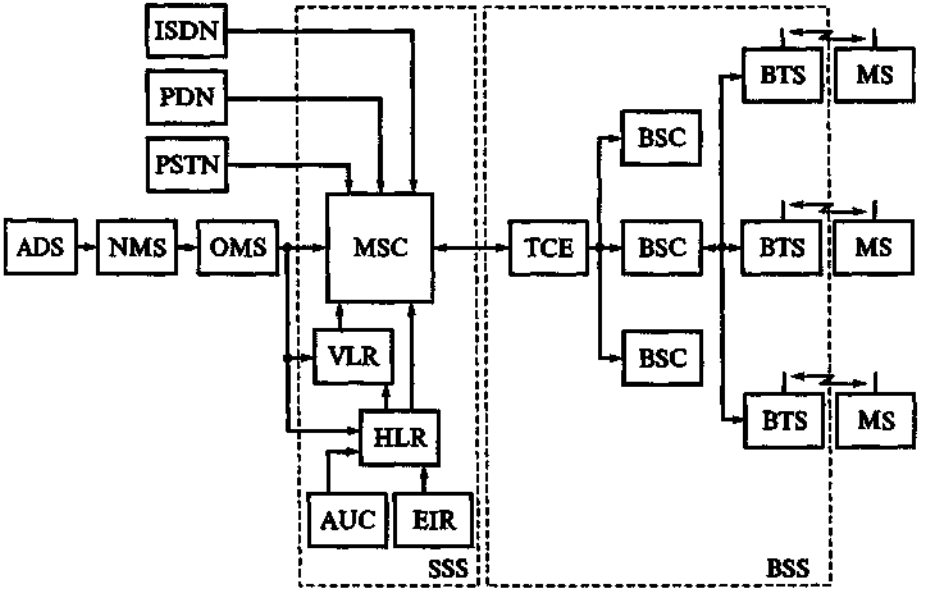



**Рис. 1.40.** Архитектура и основные компоненты сети GSM

оптимальное иерархическое управление системой, производит эксплуатацию и техническое обслуживание, а также управление трафиком во всей сети. Кроме того, NMC контролирует работу устройств автоматического управления и отражает на дисплее состояние всей сети во всех регионах для операторов этого центра. Операторы NMC в экстремальных ситуациях задействуют процедуру «приоритетного доступа» для оперативных служб.

*Центр эксплуатации и технического обслуживания* (*Operations and Meintenance Center — OMC*) — второй основной узел сети, который осуществляет контроль качества работы системы и управление ее элементами. OMC производит обработку аварийных сигналов, оповещающих обслуживающий персонал, и регистрирует сведения о неисправностях и аварийных ситуациях в других устройствах сети. В функции OMC также входят: управление поступающим трафиком; сбор статистических данных о нагрузке в узлах сети, запись их в компьютер управления и вывод на дисплей для анализа операторами. OMC управляет перепрограммированием пакетов обеспечения базы данных сети. Функциональное сопряжение различных элементов системы осуществляет ряд стандартизованных интерфейсов.

*Центр коммутации подвижной связи* (*MSC — Mobile Services Switching Center*) является основной частью *подсистемы коммутации* (*Switching SubSystem — SSS*), входящей в центральный терминал сети. По существу MSC представляет собой интерфейс между фиксированными главными сетями: PSTN (ТСОП), PDN, ISDN. Данный интерфейс обеспечивает все виды соединений, связанных с мобильными станциями, и обеспечивает маршрутизацию и управление вызовами мобильных абонентов. На MSC возложена также функция ком-

мутации радиоканалов, к которым относится «эстафетная передача» при перемещении абонента из одной ячейки в другую. MSC составляет статистические данные, необходимые для контроля работы сети связи, формирует систему расчетов (*биллинг*) по состоявшимся вызовам и переговорам, поддерживает процедуры безопасности доступа к радиоканалам. Одной из важных функций MSC является регистрация местоположения подвижных абонентов и передача управления соседнему MSC при переходе абонента в другую зону обслуживания. Процедура регистрации местоположения мобильной станции обеспечивает вызовы перемещающимся абонентам от других подвижных абонентов или от абонентов телефонной сети.

Центр коммутации подвижной связи отслеживает местоположение мобильных станций, используя *регистр положения* (*Home Location Register — HLR*) и *регистр перемещения* (*Visited Location Register — VLR*). Регистр HLR представляет собой банк данных об обслуживаемых абонентах и содержит международный идентификационный номер и адрес мобильного абонента (*International Mobile Subscriber Identity — IMSI*), который используют в *центре аутентификации* (*Authentication Center — AUC*) для удостоверения подлинности абонента. Кроме того, в регистре HLR хранится та часть информации о местоположении мобильной станции, которая находится в данной зоне обслуживания, обеспечивая ее своевременный вызов. В нем ведется регистрация роуминга, включая данные о *временном идентификационном номере мобильного абонента* (*Temporary Mobile Subscriber Identity — TMSI*) и соответствующем VLR. Необходимо отметить, что эстафетная передача мобильного абонента из ячейки в соседнюю, обслуживаемых одним MSC (т. е. в его зоне обслуживания), осуществляет один из его *контроллеров базовых станций* (*Base Station Controller — BSC*).

В сети подвижной связи GSM соты группируются в географические зоны (LA), которым присваивается свой идентификационный номер (LAC). Каждый VLR содержит данные об абонентах в нескольких LA. Когда подвижный абонент перемещается из одной LA в другую, данные о его местоположении автоматически обновляются в VLR. Если старая и новая LA находятся под управлением различных VLR, то данные на старом VLR стираются после их копирования в новый VLR. Текущий адрес VLR абонента, содержащийся в HLR, также обновляется. В целом VLR представляет собой временный банк данных о мобильном абоненте, находящемся в зоне его регистрации.

Это исключает постоянные запросы данных о мобильном абоненте в устройстве HLR и сокращает время на обслуживание вызовов.

Абоненту сети присваивают *стандартный модуль подлинности* (*Subscriber Identity — Module — SIM*, или SIM-карта), содержащий *алгоритм аутентификации* (*Authentication Algorithm*), *ключ аутентификации* (*Individual Subscriber Authentication Key – Ki*) и IMSI. В результате проверки этой информации разрешают доступ абонента в сеть. *Регистр идентификации оборудования* (*Equipment Identity Register — EIR*) содержит центральную базу данных, подтверждающих подлинность *международного идентификационного номера оборудования мобильной станции* (*International Mobile Station Equipment Identity — IMEI*).

**SIM-карта.** Наличие SIM-карты в телефоне значительно упрощает жизнь пользователям сетей GSM, поскольку с ее помощью достигается независимость аппаратов от конкретного оператора сотовой связи. Модуль оформлен в виде банковской карточки и содержит в памяти все необходимые данные, связанные с полномочиями абонента и предоставляемыми ему услугами связи. С абонентской станцией поступают так же, как с банкоматом: пока в станции нет карты — услуги связи не предоставляются. SIM-карта позволяет абоненту пользоваться любой станцией стандарта GSM, например, установленной в такси, поезде или телефонной будке. Вынув модуль SIM из одного телефонного аппарата и вставив его в другой, абонент может продолжать пользоваться всеми теми услугами, на которые он подписался. Использование SIM-карт позволило исключить «двойников» на сетях подвижной связи стандарта GSM. SIM-карта содержит также криптографические ключи и алгоритмы шифрования, используемые для организации шифрования данных для обеспечения конфиденциальности связи.

Эти достаточно сложные в реализации процедуры направлены на борьбу с несанкционированным доступом к услугам ССПС (*fraud* — фрод, буквально обман, мошенничество) и прослушиванием разговоров пользователей. Несмотря на это, существует немало прецедентов взлома сетей GSM. Разделы спецификации стандарта GSM, описывающие меры криптографической защиты, являются секретными. Однако есть мнение, что именно открытость в этом вопросе поможет успешно бороться с мошенничеством и прослушиванием.

База данных EIR содержит три списка, где номера IMEI абонентов помечены следующим образом:

- *белый список* — номера, закрепленные за санкционированными подвижными станциями;
- *черный список* — номера подвижных станций, которые украдены или им отказано в обслуживании сетью;
- *серый список* — номера подвижных станций, имеющих нерешенные с сетью проблемы.

*Оборудование базовой станции* (*Base Station System — BSS*) состоит из трех основных узлов: *транскодера – преобразователя аналогового сигнала в цифровой* (*Transcribe Code Element — TCE*), BTS и BSC. Транскодер осуществляет преобразование сигналов речи передающего канала и данных MSC (скорость передачи 64 кбит/с ИКМ-линией) к виду, определяемому соответствующим протоколом стандарта GSM. Согласно требованиям стандарта скорость передачи сигналов в цифровой форме должна составлять 13 кбит/с (*полноскоростной* канал). Если требуется в заданной полосе передавать по каналу несколько речевых сообщений в цифровой форме, то надо снизить скорость передачи. Это установлено стандартом, и в перспективе в системе GSM будут использовать «*полускоростной*» речевой канал со скоростью передачи 6,5 кбит/с для ИКМ-линии.

В протоколах сети GSM предусмотрена передача данных MSC и речи со скоростью 64 кбит/с. Это позволяет использовать в каждом канале четырехкратное временно́е уплотнение данных цифровых сигналов. Поскольку один полноскоростной канал ведет передачу со скоростью 13 кбит/с, то в транскоде-

ре и MSC к передаваемому потоку ведется добавление дополнительных неинформационных битов (*стаффингование*, от англ. *stuffing* — наполнять, укомплектовывать) до скорости передачи 16 кбит/с. Таким образом, формируется 30-канальная ИКМ-линия, позволяющая передавать 120 речевых каналов. В дополнение к этим каналам организуется еще два служебных канала для передачи сигнальной информации и пакетов специальных данных.

Абонентские MS служат для организации связи абонентов сети с PSTN. Стандартом предусмотрено пять моделей MS: модель 1-го класса имеет выходную мощность 20 Вт и предназначена для установки на мобильном транспорте. Выходной мощностью 0,8 Вт обладает карманная модель 5-го класса. В оборудование MS системы введено устройство адаптивной регулировки мощности передатчика, обеспечивающее оптимальное качество связи при изменении расстояния до BTS. Все включенные MS постоянно работают в режиме «дежурного приема» (*stand-by*) на канале вызова. Для вызова абонента его закодированный опознавательный сигнал включается одновременно на всех BTS зоны обслуживания. Получив свой вызывной сигнал, MS подтверждает факт его принятия на ответной частоте канала вызова. После установления этой процедуры ЦКС подключает на связь переговорный канал той базовой приемо-передающей станции (ячейки), в зоне которой обнаружена мобильная станция. Если вызов осуществляется подвижным абонентом, то его MS автоматически находит и вводит в связь свободный канал ближней базовой приемопередающей станции.

Важным для MS является эфирный интерфейс — радиоинтерфейс обмена между MS и BTS, поскольку на одной частоте могут одновременно «разговаривать» восемь пар абонентов. В сети GSM каналы связи делят на *физические* и *логические*. Передачу речи и данных в физических каналах организуют кадры длительностью 4,615 мс, состоящие из восьми слотов (от англ. *Slot* — разъем). Каждый слот соответствует своему каналу речи, т. е. восемь каналов речи передаются в одном частотном канале при полноскоростном кодировании речи (при полускоростном, используемом для повышения емкости сети, но с потерей качества передаваемой речи, — шестнадцать каналов). Информационный кадр может быть кадром канала трафика или канала управления. При этом кадры группируются в мультикадры, те в свою очередь — в суперкадры, а из суперкадров складывается гиперкадр длительностью 3 ч 28 мин 53,76 с. Необходимость большого периода гиперкадра объясняется требованиями шифрования данных. В аппаратуре системы используют *эквалайзеры* (от англ. *equalizer* — корректоры), обеспечивающие выравнивание импульсных сигналов, амплитуда которых меняется вследствие *интерференционных замираний*. Подобные явления связаны с многолучевым распространением волн в городских застройках.

Служба SMS напоминает широко известную службу пейджинга — персонального радиовызова. Во многом именно эта технология связи «убила рынок пейджинговой связи». При передаче SMS используется пропускная способность каналов сигнализации. Сообщения могут передаваться и приниматься подвижной станцией. В рамках этой услуги связи абоненты могут обмениваться буквенно-цифровыми и текстовыми сообщениями в объеме до 160 знаков латини-

цей и до 70 знаков кириллицей. Передачу коротких сообщений можно использовать в чрезвычайных ситуациях или при перегрузке каналов речевой связи. Тарифы на передачу SMS обычно значительно ниже тарифов речевой связи.

Большинство стран приняли стандарт GSM диапазона частот 900 МГц и развивают сети в диапазоне 1800 МГц. Это DCS-1800 в Европе, и 1900 МГц — PCS-1900 в США. В настоящее время в Европе, США и России широко применяют стандарт второго поколения на основе систем CDMA (система IS-95). Их пропускная способность превышает в пределах той же полосы частот существующую пропускную способность сетей подвижной связи более чем в 15 раз. Сейчас идет практическое внедрение систем подвижной связи третьего поколения IMT-2000 (3G). Эту сеть называют FOMA (*Freedom of Mobile Multimedia Access* — свободный доступ к мобильным мультимедийным ресурсам). Она была спроектирована для организации новой мобильной коммуникационной системы, которая смогла бы быть запущена в любом уголке планеты и позволяла абонентам качественно общаться, слушать музыку, проводить видеоконференции и многое другое. Отличительными чертами систем 3G являются:

- доступность услуг связи в любом месте и в любое время, «связь всегда и везде» (*anywhere, anytime*);
- существенное увеличение номенклатуры услуг, в первую очередь, услуг мультимедиа и беспроводного доступа в *Internet*;
- мобильный доступ ко всем ресурсам единого мирового информационного пространства, интеграция услуг сетей фиксированной и мобильной связи;
- гибкий маркетинг.

В большинстве цифровых ССПС используют фазовые или частотные методы манипуляции, как наиболее эффективные по потребляемой мощности и полосе рабочих частот. ССПС является системой массового обслуживания со случайным потоком вызовов (описывается распределением Пуассона), случайной продолжительностью обслуживания (подчиняется экспоненциальному распределению) и фиксированным числом каналов связи. Было бы нерационально ограничивать число абонентов числом каналов, так как вероятность того, что все абоненты захотят воспользоваться связью одновременно, крайне низка. Поэтому ССПС строят из расчета среднего трафика, рассчитываемого как произведение средней частоты вызовов на среднюю продолжительность обслуживания одного вызова. Если трафик оказался выше, то абонента ждет режим *system busy* (система занята, или перегружена; при очень больших загрузках пропускная способность сети может стать нулевой и эту ситуацию называют *коллапсом сети*).

С тех пор, как системы сотовой связи стали массированно внедрять во всем мире, прошло немного лет. Первые абонентские мобильные станции имели значительные размеры и были похожи скорее на небольшие переносные радиостанции, чем на малогабаритные телефонные трубки. Затем, по мере развития, уменьшались их размеры и масса, улучшался дизайн и ряд других показателей.

Ожидается, что на базе этих разработок вскоре будет создана глобальная системы подвижной связи четвертого поколения 4G, которые позволят передавать данные в сотовых сетях со скоростью выше 100 Мбит/сек.

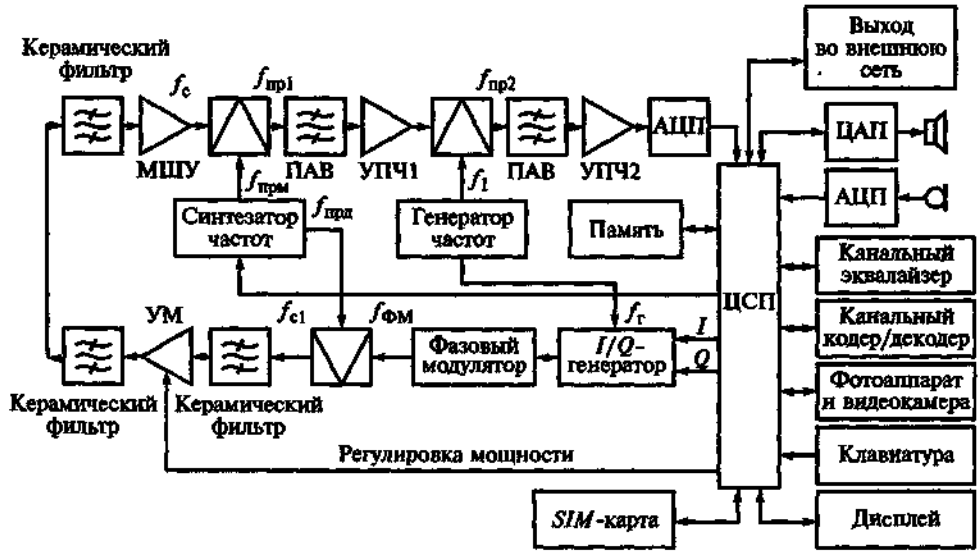



**Рис. 1.41.** Структурная схема цифрового сотового радиотелефона

На рис. 1.41 представлена структурная схема современного цифрового сотового радиотелефона сотовой системы стандарта GSM. Приемное устройство радиотелефона представляет собой соединенные последовательно ненаправленную, достаточно широкополосную антенну и супергетеродинный приемник с двойным преобразованием частоты радиосигнала. Принятый антенной переговорный радиосигнал $f_c$ через высокочастотный полосовой *керамический фильтр* (фильтр со стабильными частотными характеристиками) и *малошумящий усилитель* (МШУ) поступает на один вход первого смесителя △ приемника. На другой его вход подается напряжение гетеродина $f_{прм}$ с синтезатора частот (многочастотного генератора со ступенчатым переключением частот).

Сигнал первой промежуточной частоты $f_{пр1}$ выделяется полосовым фильтром на ПАВ (англ. *SAW filter*), усиливается *усилителем первой промежуточной частоты* (УПЧ1) и поступает на первый вход второго смесителя приемника. На второй вход смесителя подается напряжение гетеродина $f_г$ (вспомогательный генератор) с *генератора частот*. Полученный в результате преобразования полезный сигнал второй промежуточной частоты $f_{пр2}$ отфильтровывается полосовым фильтром на ПАВ, усиливается усилителем УПЧ2 и поступает на АЦП. В АЦП аналоговый сигнал преобразуется в цифровой код, с которым оперирует *цифровой сигнальный процессор* (ЦСП; *Digital Signal Processor — DSP*).

Как правило, в таких радиотелефонах кроме цифровой структуры имеется и аналоговая часть. Антенна одновременно является и передающей, и приемной. Обычно она представляет собой так называемую *низкопрофильную антенну* (см. далее). Аналоговая часть радиотелефона включает в себя высокочастотные и низкочастотные передающее и приемное устройства, которые выполнены по классической для любой системы радиосвязи схеме. Передающее устройство мобильного радиотелефона формирует информационный радиосигнал с доста-

точно сложным законом модуляции. В режиме передачи, созданный в ЦСП, цифровой переговорный сигнал поступает на аналоговую часть радиопередатчика. Модулирующий сигнал формируется в *I/Q-генераторе*, на который подается колебание генератора частот. С выхода *I/Q*-генератора полученный сигнал поступает на *фазовый модулятор*, с которого колебание $f_{ФМ}$ подается на смеситель ▽ передатчика. На второй вход смесителя приходит напряжение частоты $f_{пред}$ с синтезатора частот. Преобразованный сигнал $f_{c1}$ через полосовой керамический фильтр подают на регулируемый *усилитель мощности* (УМ), который управляется сигнальным процессором. Регулировка излучаемой мощности телефона осуществляется по специальным командам БПС, через которую реализуется связь с мобильным абонентом. Усиленный до необходимого уровня мощности сигнал частоты $f_{c1}$ через полосовой керамический фильтр поступает в антенну, излучающую его в окружающее пространство.

Цифровая часть схемы радиотелефона формирует и обрабатывает передаваемые и принимаемые информационные и служебные сигналы. Она включает *цифровой сигнальный процессор*, *память* (оперативную, постоянную и другие виды памяти), SIM-карту, АЦП, ЦАП, *канальный эквалайзер* (выравниватель амплитуд сигналов, в данном случае импульсных), *канальный кодер/декодер*, *клавиатуру*, *дисплей*, *фотоаппарат*, *видеокамеру* и *выход во внешнюю сеть*. Логическая часть телефона выполняет операции кодирования/декодирования, сжатия и восстановления сигнала; обрабатывает информацию, вводимую пользователем с клавиатуры, и осуществляет ряд других задач.

Последние разработки цифровых радиотелефонов существенно расширили сервисные возможности. *Абонентские терминалы* весьма разнообразны как по своему конструктивному исполнению, так и по сервисным возможностям, ими предоставляемыми. Среди существенных сервисных возможностей отметим:

- наличие кнопки временного отключения микрофона от сети;
- наличие оперативной памяти для повторного вызова последнего абонента, в том числе и для многократного вызова (автодозвона) занятого абонента;
- наличие долговременной памяти номеров приоритетных абонентов;
- постановку собеседника на удержание с включением фоновой музыки;
- автоматическое определение номера (схема АОН) вызывающего абонента с отображением на дисплее и звуковым его воспроизведением;
- защиту от АОН вызываемого абонента (анти-АОН);
- запоминание номеров вызывающих абонентов и времени каждого вызова;
- индикацию во время разговора второго вызова и его номера;
- наличие персональных кодов-паролей;
- наличие автоответчика и встроенного диктофона для записи сообщений;
- наличие дистанционного управления телефоном;
- возможность подключения телефона к компьютеру и внешней (например, *Internet* через технологии Wi-Fi) сети;
- возможность принимать и пересылать другим абонентам *SMS-ки*;
- возможность получать данные о погоде, биржевую информацию;
- наличие встроенного цифрового фотоаппарата, видеокамеры и т. д.

**Подвижная связь в городах**

В современных системах сотовой связи используются радиоволны дециметрового диапазона, которые испытывают сильные отражения от окружающих объектов и подстилающей поверхности. Это приводит к многолучевому распространению радиосигнала. Сложение в точке приема радиоволн, пришедших разными путями и имеющих соответственно разные фазы, но сравнительно одинаковые мощности, вызывает усиление результирующего сигнала до 10 дБ или, что чаще, ослабление до 30 дБ. Искажения результирующего сигнала обусловливают межсимвольную интерференцию. Колебания среднего уровня сигнала приводят к замираниям. Они бывают быстрыми и медленными. Опасность представляют первые. Для борьбы с быстрыми замираниями используют разнесенный прием и медленные скачки по частоте (*Slow Frequency Hopping*).

При осуществлении подвижной связи в городах имеют место проблемы, связанные с распространением радиоволн. Проблемы возникают в условиях города при связи БПС с движущимся абонентом, когда сравнительно короткие, но переменные по длине линии связи быстро превращаются из открытых трасс в закрытые. В этом случае к приемной антенне приходят несколько сигналов с разным запаздыванием по времени за счет неоднократных переотражений волн элементами зданий. При этом уровень принимаемого сигнала испытывает глубокие, до 15...40 дБ, быстрые замирания, зависящие от плотности застройки города зданиями. Впоследствии стало ясно, что для электромагнитных полей в условиях городов характерны пространственные интерференционные явления, образуемые множеством волн с различными амплитудами и фазами из-за дифракции на препятствиях и многократных отражений от них. Было установлено, что период пространственных флуктуаций сигнала по порядку значений близок к длине волны излучения. Задача о распространении сигналов в городах оказалась многопараметрической, поскольку уровни принимаемых сигналов зависели от рельефа местности, высот антенн передатчика и приемника, плотности застройки, высоты крыш зданий, ширины и направления улиц, наличия отдельно стоящих деревьев и лесопарковых насаждений и уличного транспорта. В настоящее время установлены основные закономерности распределения электромагнитных полей в городах. При высоте приемной антенны на уровне крыш зданий напряженность поля убывает обратно пропорционально квадрату расстояния. При высоте антенн в 3 м над земной поверхностью поле убывает пропорционально $1/R^m$, где $m$ = 2,9...3, для крупного города с небольшой этажностью зданий и $m$ = 2,7...2,8 для небольших городов. Ослабление поля в тени зданий составляет в 50 % случаев 18...20 дБ в диапазоне частот 470...670 МГц.

Моделирование местности и зданий города позволяет определять лишь средние значения уровней сигналов и не может служить для оценок статистики полей ЭМИ в городе. В последние годы были изучены экспериментально закономерности распространения на коротких расстояниях по улицам городов, а также внутри помещений и зданий, что позволило создать сотовые системы с автомобилями, пешеходами и связь внутри учреждений на базе радиотелефонов.

### Профессиональные (транкинговые) системы подвижной связи

Профессиональные (предназначены для корпоративных групп абонентов — бригад скорой помощи, МЧС, пожарных, ФСБ, милиции и т. д.) системы подвижной связи с так называемым *свободным и равным доступом мобильных станций к общему частотному диапазону* позволяют абонентам работать на любом переговорном канале сети.

В мировых стандартах профессиональных систем подвижной связи метод свободного и равного доступа мобильных абонентов ко всем каналам сети связи называют *транкингом* (от англ. *trunk* — ствол, магистраль). При этом любой свободный переговорный канал может быть временно закреплен за мобильным абонентом для конкретного сеанса связи в зависимости от трафика сети. Для этого в мобильные станции встроены специальные микропроцессоры, позволяющие им сканировать (т. е. искать) запрограммированные частоты сети, передавать при каждом выходе в эфир собственный код, код входа в систему и номер вызываемого абонента. Транкинговые системы подвижной связи получили широкое распространение. До середины 60-х гг. XX в. развивались так называемые *производственные системы подвижной связи* (*Private Mobile Radio* — PMR), создаваемые отдельными организациями для удовлетворения своих потребностей в подвижной связи на ограниченных территориях.

С конца 60-х гг. XX в. начинается интенсивное развитие сетей транкинговой связи как производственных, так и систем подвижной связи общего пользования (*Public Access Mobile Radio* — PAMR). Системы PAMR создаются операторами сетей подвижной связи на коммерческой основе и разворачиваются на обширных территориях. Абонентам этих сетей предоставляется возможность связи не только с абонентами данной сети, но и с абонентами ТСОП. В конце XX в. становится необходимым создание глобальных сетей PAMR, которые охватили бы большие регионы, включающие ряд стран. Абоненты этих сетей должны иметь связь независимо от своего местонахождения и возможность выхода на ТСОП. Это особенно необходимо для служб безопасности (милиции, таможенных служб), так как позволяет им предпринимать согласованные действия по пресечению деятельности преступных группировок и т. п. Особенностями транкинговых систем являются: весьма незначительное время установления связи между абонентами; возможности осуществления группового вызова абонентов, установления непосредственной связи между терминалами абонентов без использования базовых приемо-передающих станций сети связи и т. д.

До 1995 г. создавались аналоговые транкинговые системы, в которых передавались сигналы телефонии и применялась частотная модуляция. Ширина полосы одного канала составляла 25...30 кГц. Значительной вехой в развитии систем транкинговой связи явилась разработка спецификации MPT-1327, которой руководствовались многие фирмы при выпуске оборудования. В последнее десятилетие XX в. в США и Европе были разработаны цифровые системы транкинговой связи (TETRA — *Trans European Trunked Radio*; IDEN — *Integrated Digital Enhanced Netwok*; EDACS — *Enhanced Digital Access System* и др.).

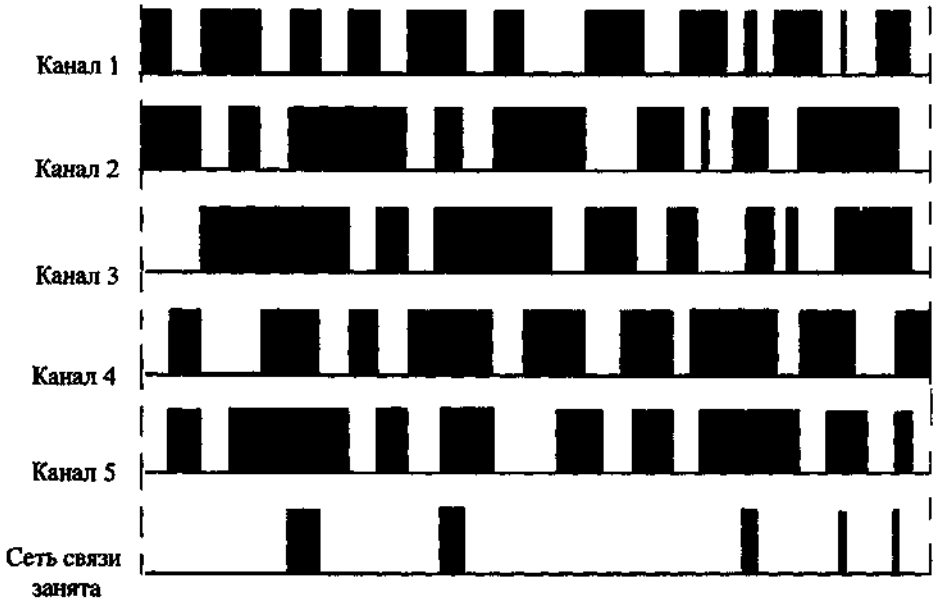



**Рис. 1.42.** Диаграммы часового трафика пятиканальной транкинговой системы связи

Стандарт на систему профессиональной связи TETRA был разработан в 1992 г. Для этой системы выделено несколько полос частот в диапазоне частот ниже 1 ГГц, одна из которых (380...400 МГц) предназначена для создания сетей TETRA европейских служб безопасности. В системе абонентам предоставляется услуга роуминга, и сегодня уже началось внедрение этой системы в ряде стран Западной Европы.

В цифровой системы транкинговой связи TETRA в каждом частотном канале шириной 25 кГц передают сигналы четырех абонентов. Таким образом, по спектральной эффективности эта система в 4 раза превосходит обычные системы с частотной модуляцией. Помимо передачи речи в цифровой форме возможна передача данных со скоростью 7,2 ... 28 Кбит/с, допустимы несколько уровней приоритета вызовов, групповые и срочные вызовы, передача пакетных данных, возможность непосредственной связи между абонентами, минуя базовую станцию (БС), и т. д.

Рассмотрим упрощенные диаграммы типичного часового трафика работы пятиканальной транкинговой системы профессиональной подвижной связи со средней продолжительностью одного сеанса переговоров абонентов 3...5 мин (рис. 1.42). Темные участки на рис. 1.42 отражают ситуации, когда каналы связи заняты переговорами, а светлые — когда они свободны. Если бы абонент фиксировано закреплялся за отдельным каналом связи, то вероятность немедленного доступа составила бы менее 50 %, в то время как при транкинговом методе подключения к любому свободному каналу вероятность такого доступа резко возрастает (практически до 80...90 %).

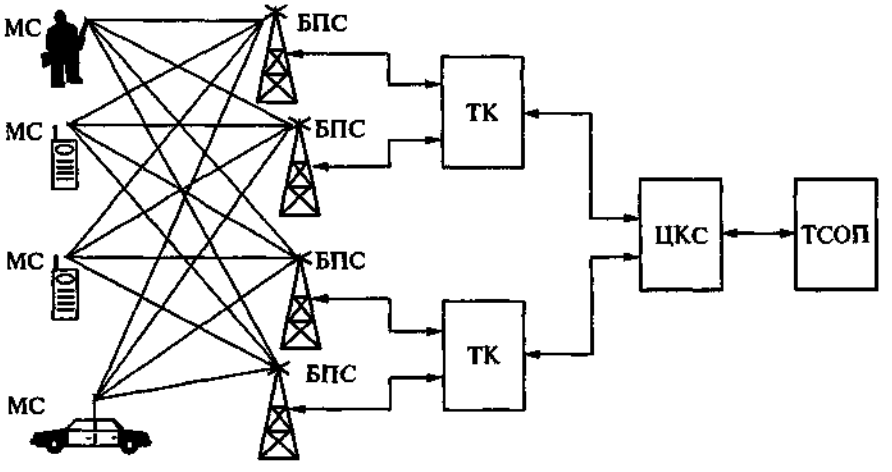



**Рис. 1.43.** Упрощенная структура транкинговой системы подвижной связи

Упрощенная структура транкинговой системы подвижной связи показана на рис. 1.43.

В транкинговых системах связи реализуются два основных метода выделения канала конкретному абоненту.

При первом методе выделения канала абоненту поиск свободного канала и подачу сигнала вызова производит мобильная абонентская станция, которая осуществляет сканирующий последовательный *автоматический поиск вызывного канала* (АПВК). В этом случае перед установлением связи, когда мобильная станция ведет сканирующий поиск свободного канала с применением устройства АПВК, на каждом определенном канале производится попытка вхождения в связь с базовой станцией с установкой тактовой и циклической синхронизации. Вследствие этого длительность цикла установления канала связи возрастает на несколько порядков по сравнению с длительностью при фиксированном закреплении каналов за определенными мобильными абонентами. Поэтому использование профессиональных транкинговых систем связи с АПВК эффективно при работе с 10...15 частотными каналами.

При втором методе построения транкинговой системы поиск свободного канала связи возлагается на подсистему управления базовой станции ПСПС. В этом случае для поиска свободного связного канала используется специальный канал управления базовой станции, через который обеспечивается контроль работы всей сети, включая процедуры установления, обеспечения и прекращения связи.

Важным аспектом для сети подвижной связи считается возможность, а скорее необходимость оперативного переключения канала связи на другую несущую частоту при появлении помех высокого уровня. Кроме того, для повышения надежности системы подвижной связи предусмотрено автоматическое переключение на исправный канал при выходе из строя работающего канала связи или снижения его выходной мощности.

### Системы персонального радиовызова

По принципу действия СПРВ — однонаправленная (симплексная) система связи, предназначенная для вызова подвижного абонента на двустороннюю (отложенную на некоторое время) связь по принципу «в любое время, в любое место зон обслуживания, любому человеку».

Начало развития современных СПРВ общего пользования можно отнести к 1956 г., когда была создана первая пейджинговая система *Multiton*. В системе, которая для передачи сообщений использовала специально выделенный радиоканал, абонент имел малогабаритный приемник — пейджер, способный из общего потока сообщений, передаваемых по радиоканалу, выделить адресованный ему сигнал. При приеме этого сигнала в зависимости от принятой кодовой комбинации издавался звук определенного тона, услышав который абонент мог, нажав на кнопку, прослушать посланное ему речевое сообщение. Позже, из-за необходимости улучшить эффективность использования радиоканала, отказались от передачи речевого сообщения. Вызов абонента включал тоновый звуковой сигнал пейджера, который извещал его о необходимости совершить определенные действия. Обычно ширина полосы канала пейджинговой системы составляла 25 кГц, и для передачи сигналов использовалась частотная модуляция. Для систем выделялись каналы в диапазоне частот 50...900 МГц.

Важной вехой в развитии пейджинговой связи явилась разработка в 1976 г. протокола POCSAG, принятого в качестве международного. В 1982 г. впервые были разработаны пейджеры с дисплеем, на котором абонент мог увидеть посланное ему буквенно-цифровое сообщение. В системах ПРВ, использующих этот код, информацию можно передавать со скоростью 512, 1 200 и 2 400 бит/с.

В упрощенном виде основные технические средства СПРВ можно представить в виде следующих крупных элементов (рис. 1. 44):

- средства сбора сообщений от отправителей в пейджинг-центре;
- средства операторского зала (или пейджинг-центра), включая АРМ операторов и пейджинг-терминал;
- средства передачи сообщений на пейджеры, включая различные линии связи, передающие устройства и ретрансляторы;
- сервисные подсистемы СПРВ, к которым можно отнести группу подсистем сервисного обслуживания абонентов, а также группу подсистем, предназначенных для внутреннего (внутрисистемного в СПРВ) применения.




**Рис. 1.44.** Основные технические средства СПРВ

Работу пейджинговой сети связи упрощенно можно описать следующим образом. Мобильные абоненты системы связи всегда имеют при себе малогабаритный приемник персонального вызова (пейджер), находящийся в режиме дежурного приема. Вся обслуживаемая территория охватывается сетью маломощных передатчиков в соответствии с известной сотовой моделью их размещения (см. рис. 1.38). Эти передатчики с помощью проводной или радиоканальной связи подключены к выходу пейджингового терминала, который, в свою очередь, связан с телефонной сетью общего пользования.

Пейджинговый терминал преобразует поступившее сообщение в цифровой формат соответствующего кода СПРВ, переводит его в буферную память компьютера и ставит в очередь к ранее поступившим сообщениям. Далее закодированное сообщение через сеть всех радиопередатчиков пейджинговой системы излучается («выстреливается») в эфир. Включенные на прием абонентские пейджеры непрерывно анализируют адреса поступающих вызовов. При совпадении поступившего адреса с его собственным сообщение принимается, записывается в буфер памяти и высвечивается на дисплее пейджера. При этом о принятом сообщении абонент извещается звуковым и световым сигналами или вибрацией корпуса приемника.

Наряду с пейджерами разработаны и начинают внедрять малогабаритные двусторонние пейджеры — *твейджеры*, или *трансиверы* (приемопередатчики), передающие подтверждение приема сообщения и краткую ответную информацию на базовую станцию. В такие абонентские приемники вводится ряд дополнительных сервисных услуг.

В 1992 г. создается общеевропейская система ERMES (*European Radio Message System*), работающая в полосе частот 169,4...169,8 МГц. Эта система обеспечивает общеевропейский роуминг и высокую скорость передачи сигналов (6,25 Кбит/с). Она позволяет создавать сети очень высокой емкости для передачи разных видов сообщений, включая текстовые. Сети ERMES не получили широкого распространение ввиду сложности оборудования.

В1993 г. для СПРВ был разработан протокол FLEX, обладающий повышенной помехоустойчивостью и имеющий набор возможных скоростей передачи сообщений (1,6; 3,2 и 6,4 Кбит/с). Основное достоинство этого протокола состоит в его гибкости — он обеспечивает высокую степень согласования с существующими системами ПРВ, в которых применяется протокол POCSAG. Кроме того, пейджеры FLEX за счет синхронного режима работы имеют увеличенный в 4-5 раз срок службы батарей пейджеров по сравнению с пейджерами POCSAG.

### Системы подвижной спутниковой связи

Перспективным направлением развития подвижной связи общего пользования является создание спутниковых систем. Системы подвижной спутниковой связи начали развиваться в последние два десятилетия XX в. Одной из первых подобных систем явилась созданная в 1967 г. в США опытная система «TATS». Данные системы спутниковой связи предназначены для организации переговоров между абонентами телефонных сетей общего пользования и мобильными

станциями, устанавливаемыми на подвижных объектах (автомобилях, кораблях, самолетах и т. д.), а также осуществления персональной подвижной связи на базе сотовых сетей. При использовании персональной спутниковой связи обеспечивают соединение перемещающегося в пространстве абонента по его неизменному (подобно телефонному), закодированному номеру.

В основу организации спутниковой системы радиосвязи заложена достаточно простая идея. На искусственном спутнике Земли, или, проще, спутнике (иногда называемым сателлитом), располагается активный ретранслятор СПСС. Спутник находится на заданной орбите и движется над Землей длительное время, получая электропитание от солнечных батарей, установленных на его платформах, или от малогабаритных ядерных электростанций.

На спутнике-ретрансляторе расположена антенная система и приемопередающая аппаратура, осуществляющие прием, преобразование, обработку (например, усиление, изменение частоты несущей и пр.) и передачу радиосигналов в направлении земных станций (ЗС) — станций радиосвязи, расположенных на земной поверхности и предназначенных для обеспечения собственно связи. Отметим, что в наземных системах радиосвязи аналогичные станции называются наземными. На рис. 1.45 приведена упрощенная структура построения современной системы подвижной спутниковой связи, работающей непосредственно с телефонной сетью общего пользования

Системы мобильной спутниковой связи классифицируют по двум признакам: типу используемых орбит и различию в зонах обслуживания и размещения ЗС. В состав любой сети спутниковой связи входят земные станции (земной и абонентские сегменты) трех видов:

- абонентские станции (АС) — авиационная, морская, сухопутная, переносная, персональная;
- земные стационарные станции сопряжения (ЗСС);
- станции управления сетью (СУС).

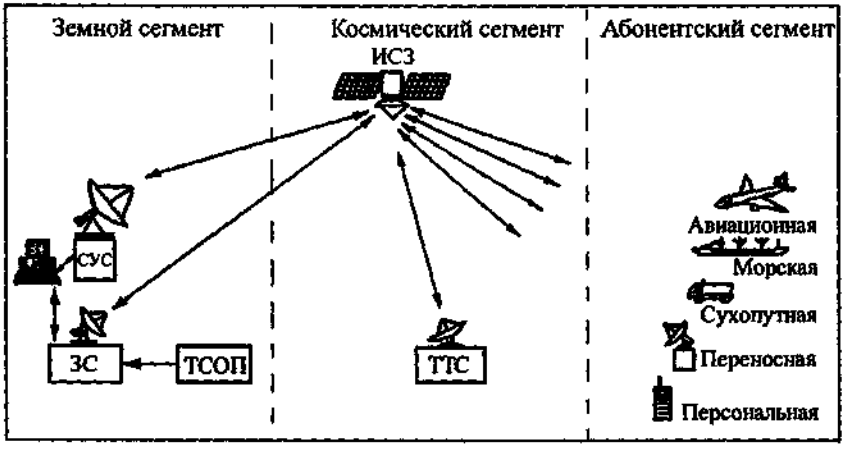



**Рис. 1.45.** Упрощенная структура построения СПСС

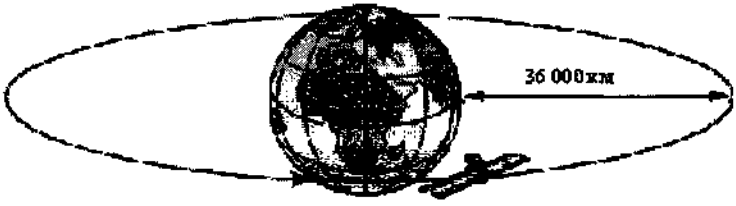



**Рис. 1.46.** Геостационарная орбита спутника

Очень часто на схемах земные станции ЗСС и СУС объединяют и обозначают как СУС. Кроме того, собственно космический (спутниковый) сегмент содержит устройства, условно названные *терминалом телеуправления спутником* (ТТС), обеспечивающим эксплуатацию, телеуправление и контроль за работой систем спутника связи.

По типу используемых орбит различают СПСС со спутниками, расположенными на геостационарных (*Geosynchronous*; орбита, рассчитанная таким образом, чтобы спутник постоянно находился над одной и той же точкой земной поверхности; для этого он должен перемещаться со скоростью вращения Земли, т. е. его период обращения равен 24 ч; высота 36 000 км; рис. 1.46), высокоэллиптических промежуточных и низких земных орбитах (*Low Earth Orbit — LEO*). Последние называются системами спутниковой связи на низкоорбитальных спутниках (высота орбит ИСЗ 200...700 км).

Системы подвижной радиосвязи на низкоорбитальных спутниках позволяют создать на поверхности Земли плотность потока мощности электромагнитных колебаний, достаточную для работы с легкими абонентскими станциями размером с портативную телефонную трубку, и дополняют сотовые системы радиосвязи. Наиболее распространенной международной системой подвижной спутниковой связи является глобальная сеть связи *Inmarsat-M*, предназначенная для обслуживания подвижных абонентских станций. Сеть Inmarsat-M обеспечивает связь практически с любой точкой мира, позволяет подключить компьютерную сеть Internet, факс и ряд других устройств передачи цифровых данных. Космический сегмент системы связи базируется на геостационарных спутниках, расположенных над Атлантическим, Тихим и Индийским океанами.

В настоящее время спутниковая связь все более переводится в плоскость *персонального обслуживания* подвижных абонентов. Энергетический баланс линий спутниковой связи до последнего времени не позволял уменьшить абонентскую станцию до размеров сотового телефона. Однако применение спутников, находящихся на негеостационарных орбитах, в том числе низкоорбитальных, позволяет, в сравнении с геостационарными спутниками, существенно уменьшить задержку в канале связи, что весьма существенно для передачи речевых сообщений, снизить энергетику линии, что позволяет значительно уменьшить габаритные размеры и массу абонентского терминала, а также использовать абонентские терминалы с ненаправленными антеннами. Это создает пре-

имущества перед геостационарными и высокоорбитальными спутниками и позволяет разрабатывать СПСС с персональными радиотелефонами типа сотового, снабженными ненаправленными антеннами. При этом существенно уменьшаются затухание сигнала на трассах Земля – спутник и спутник – Земля и его запаздывание в каналах связи. Для сравнения отметим, что время задержки сигнала у геостационарных систем спутниковой связи составляет около 300 мс (это особенно заметно по разговорам корреспондентов на телевизионном экране, когда они ведут репортаж через спутниковую систему связи), а у низкоорбитальных — не более 200 мс. Такое уменьшение запаздывания сигналов способствует двухскачковому (двукратному) методу передачи сигналов через спутники.

СПСС с низкоорбитальными спутниками обеспечивают достаточно широкие функции в обслуживании абонентов. Прежде всего, они позволяют организовать телефонную персональную связь с подвижным абонентом, находящимся вне зоны действия телефонных сетей (сотовых и прочих). Кроме того, они широко внедрены в морских службах спасения для радиоопределения местоположения объекта, пейджинга, электронной почты и т. д.

Проект современной спутниковой системы связи основан на широком международном сотрудничестве, в котором участвуют и российские компании. В проекте орбитальной группировки практически каждой спутниковой системы связи используется до 70 спутников-ретрансляторов, расположенных на 4 – 8 орбитах (см. аналог на рис. 1.36). Любой спутник орбитальной группировки своими лучами формирует несколько наземных сот связи. В совокупности один ретранслятор создает на Земле подспутниковую зону диаметром примерно 4 500 км. Полная орбитальная группировка формирует практически сплошную спутниковую зону связи, покрывающую всю поверхность Земли.

Из отечественных сетей космической связи наиболее перспективной является система *Сигнал*. Космический сегмент системы связи Сигнал включает «созвездие» из 45…55 спутников-ретрансляторов, находящихся на орбитах высотой 700...1 500 км. Спутники расположены небольшими группами (3…5 штук) в определенных плоскостях неба так, что при движении по заданным орбитам они узкими диаграммами направленности своих антенн совокупно формируют сотовую структуру заданной зоны обслуживания.

Помимо упомянутых систем в ряде стран разрабатывают другие проекты систем спутниковой подвижной связи общего пользования, а также специализированные системы спутниковой подвижной связи, предназначенные для контроля над состоянием и местоположением транспортных средств, обеспечения связи в чрезвычайных ситуациях, осуществления экологического и промышленного мониторинга и т. п. Некоторые из них уже реализованы.

### Системы беспроводных телефонов

Системы беспроводных телефонов (русск. жаргон — *радиоудлинители*), или системы беспроводного доступа, занимают «промежуточное» положение между сотовыми и проводными сетями. Для них характерны высокая плотность и незначительная скорость перемещения абонентов (до 10 км/ч), а также воз-

можность использования на обслуживаемой территории множества базовых станций без опасности создания взаимных помех. Кроме того, системы беспроводной связи работают в нелицензируемом диапазоне частот, а следовательно, при построении нет необходимости в частотном планировании и координации с другими сетями, работающими в том же диапазоне. Системы беспроводных телефонов стали составлять конкуренцию сотовым сетям, особенно при обслуживании абонентов на дальностях 25...1000 м. В 1975 г. американская фирма «Моторола» выпустила первый беспроводной телефонный аппарат (CT — *Cordless Telephone*), позволявший абоненту свободно передвигаться с радиотелефонной трубкой в радиусе около 100 метров от базы, подключенной проводом к ТСОП. Связь радиотелефонной трубки с платформой осуществлялась по радиоканалу в диапазоне 40 ... 80 МГц с помощью частотной модуляции.

Принцип построения СБТ во многом аналогичен построению сотовых сетей связи. Внедрение систем беспроводных телефонов позволяет реализовать концепцию персональной связи, обеспечивающую связь «всегда и везде» при использовании малогабаритных абонентских станций (телефонных трубок в руке; англ. — *hand-held*). В первый период развития беспроводные телефоны использовались для персонального обслуживания на небольших территориях — в офисах фирм, квартирах, бассейнах и т. д. Связь осуществлялась в диапазонах 25...50 МГц с помощью частотной модуляции (как в системах радиовещания на метровых волнах). В системе связи применялось частотное разделение каналов, число которых не превышало десяти, а дальность обслуживания абонентов была менее 300 м. В 1985 г. в Европе был разработан аналоговый стандарт CT1 (произносится «си-ти-1») на 40-канальные дуплексные системы беспроводных телефонов, работающих в диапазоне 900 МГц и использующих метод МДЧР. Переговоры абонентов велись через индивидуальную базовую станцию с использованием персонального идентификационного кода. Затем в Европе был внедрен расширенный аналоговый стандарт беспроводных телефонов CT1+, имеющий 80 дуплексных каналов радиосвязи.

В начале 90-х гг. XX в. в Великобритании был введен в действие цифровой стандарт беспроводных телефонов CT2 диапазона 864...868 МГц, обеспечивающий конфиденциальность переговоров и лучшее качество приема сообщений. В стандарте CT2 наряду МДЧР (40 каналов с полосой частот по 100 кГц на канал) впервые было применено *временно́е дуплексное разделение* (ВДР) *режимов приема и передачи* сигналов в каждом канале (рис. 1.47). При временно́м дуплексном разделении в канале один интервал используют для передачи пакета сообщений от мобильной станции к базовой станции ($A \rightarrow B$), а следующий — для приема пакета сообщений для абонента от базовой станции ($B \rightarrow A$) и т. д.

В 1992 г. рядом зарубежных компаний была разработана общеевропейская цифровая система беспроводных телефонов связи типа *DECT* (*Digital European Cordless Telecommunication*). В системах связи стандарта DECT используют 10 несущих частот, а также многостанционный доступ с временны́м разделением каналов в сочетании с временны́м дуплексным разделением режимов приема и передачи сообщений, т. е. МДВР/ВДР.

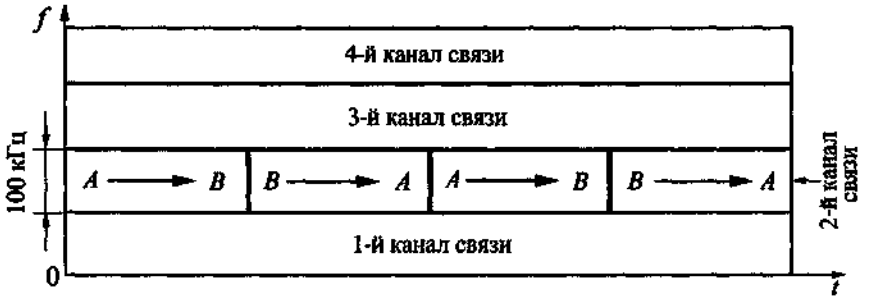



**Рис. 1.47.** Передача сообщений путем временно́го дуплексного разделения

Идея создания беспроводного доступа (CAP — *Cordless Terminal Mobility Access Profile*) состояла в том, что в микросотовых системах общего пользования должен использоваться единый терминал, способный работать дома как обычный радиотелефон, а офисе — как подключаемый к базовой станции телефон беспроводного доступа.

Современные системы беспроводной связи сети стандарта DECT обладают достаточно высокой скоростью транспортировки сообщений — до 552 Кбит/с. При этом в абонентских устройствах используются маломощные передатчики (до 10 мВт), а речевые кодеки обеспечивают качество передачи речи, близкое к качеству проводной связи.

Системы с микросотовой архитектурой на базе DECT способны обеспечить очень высокую емкость сети с плотностью абонентов до 100 тыс. чел./км$^2$. Подобные сети абонентского доступа наиболее эффективны в городах, однако они могут применяться и в сельской местности, особенно там, где концентрация абонентов достаточно высока (например, в коттеджных поселках). Не следует сбрасывать со счетов и такие достоинства DECT, как возможность оперативного развертывания сетей абонентского доступа, а также отсутствие платы за использование частотного ресурса (поскольку оборудование DECT работает в нелицензируемом диапазоне частот).

Домашние мини-АТС. Удобство беспроводных телефонных устройств определило высокий спрос на них частного потребителя. Основной объем поставок DECT-оборудования в России и мире приходится на системы для дома и малого офиса (в мире свыше 95 % от общего числа систем, установленных в 1999 г.). Данный факт лишний раз подчеркивает ориентацию DECT на рынок потребителей с невысокими доходами.

В потребительском сегменте рынка наиболее популярны DECT-системы конфигурации «одна базовая станция плюс одна-две трубки». Чрезвычайно привлекательна перспектива трансформации беспроводного телефона DECT в «домашнюю АТС» путем простого добавления дополнительных трубок. Как правило, базовый блок поддерживает 5 – 7 трубок, причем, в соответствии с требованиями DECT, эти же трубки способны работать и в более крупных системах — офисных или системах местной беспроводной связи.

### Системы местной беспроводной связи

В настоящее время серьезное внимание телекоммуникационных компаний обращено на разработку и внедрение в коммерческое использование *систем фиксированного радиодоступа к телефонным сетям общего пользования*, или *систем радиотелефонной связи*, или *систем местной беспроводной связи* (*Wireless Local Loop — WLL*). Системы местной беспроводной связи (связь «*последней мили*») находят широкое применение при обслуживании стационарных абонентов в труднодоступных сельских и малонаселенных районах. Наличие определенных проблем, связанных с количеством обслуживаемых абонентов и их удаленностью от телефонных сетей общего пользования, прокладка кабельных или оптоволоконных линий связи становится экономически невыгодной по сравнению с применением радиоканальных сетей, соединяющих стационарных абонентов с ТСОП.

Абонентская сеть связи, построенная на медном кабеле или на волоконно-оптических линия, представляет собой довольно громоздкое хозяйство, требующее, как правило, длительного поэтапного внедрения и значительных капитальных затрат при установке оборудования. Невысокий процент использования каждой абонентской пары на местных сетях делает неэффективными и непривлекательными для инвесторов крупные капиталовложения и снижает окупаемость кабельных систем. Расширение сети требует больших инженерных работ на кабельных трассах. Прокладка и организация линий связи становится сложной проблемой, особенно в старых городах, и требует повышенных капитальных затрат в сельской местности.

Системы фиксированного доступа WLL используют как временные или постоянные сети связи, не требуют прокладки дорогостоящих медных или волоконно-оптических кабелей, проведения сложных инженерных работ и вводятся в строй в считанные месяцы. Гибкая технология WLL позволяет обеспечивать потребности густонаселенных районов городов с исторически ценной застройкой, быстро растущих пригородов и дачных поселков, малых городов и редконаселенной сельской местности без развитой инфраструктуры электросвязи. Стоимость системы WLL, использующей радиоканалы, не зависит от длины кабеля, состояния грунтов, наличия водных поверхностей и заболоченных участков в пределах зоны обслуживания.

Внедрение в беспроводный доступ цифровой технологии позволило значительно повысить пропускную способность и коэффициент использования спектра частот, а также предоставить новые услуги. В результате сегодня цифровой беспроводный абонентский шлейф в состоянии конкурировать с медным кабелем в отношении надежности, габаритных размеров оборудования, качества речи и темпов реализации. Так, в США в 1995 г. более половины всей введенной за год номерной емкости было осуществлено с помощью систем WLL.

Для абонентского фиксированного радиодоступа можно использовать различные современные технологии в области радиосвязи. Наиболее известными являются две технологии, стандартизованные Европейским институтом стан-

дартизации в области связи (ENSI): системы стандартов CT2 и DECT. До настоящего времени оба стандарта использовали в основном для бесшнуровой радиотелефонии в офисах с относительно большим числом абонентов. В последнее время наибольшее применение для беспроводного фиксированного доступа находит стандарт DECT.

По сравнению с обычной кабельной сетью беспроводная телефонная система WLL обладает следующими преимуществами:

- на порядок более высокие темпы ввода в эксплуатацию, меньшая трудоемкость установочных работ;
- малый срок окупаемости (3 – 4 года) и в 1,5...2 раза меньше капитальные затраты;
- оптимальное решение для пересеченной и сильно изрезанной коммуникациями, дорогами и водоемами местности, фактически абсолютная независимость от типов грунтов;
- простота и гибкость при расширении сети, возможность сравнительно легкой трансформации в сеть мобильной связи;
- число отказов WLL составляет 6...10 % от числа отказов телефонной кабельной сети;
- в несколько раз более низкая стоимость 10-летнего жизненного цикла по сравнению с обычными телефонными сетями.

### Системы воздушной подвижной связи

Профессиональные системы воздушной подвижной связи начали разрабатывать еще в 20-х гг. XX в. (первая передача сообщения с самолета на Землю была осуществлена в 1910 г. в США), однако первая коммерческая национальная система воздушной подвижной связи общего пользования *Airfone* была создана в США лишь в 1980 г. Данная система предоставляла возможность пассажирам самолетов устанавливать и поддерживать через размещенные на территории страны БПС связь с любым абонентом сети ТСОП. Уже тогда прямо во время полета пассажиры могли решать проблемы, связанные с заказом такси, гостиниц, билетов на все виды транспорта, вести деловые переговоры, посылать факсы. В США все пассажирские самолеты, летающие на внутренних линиях, оснащены системой *Airfone*.

В 1992 г. были выделены полосы частот 1 670...1 675 МГц (Земля–самолет) и 1 800...1 805 МГц (самолет–Земля) для системы радиосвязи TFTS (*Terrestrial Flight Telecommunications System*), разработанной в ETSI. Система поддерживает 164 радиоканала шириной в 30 кГц и обеспечивает международный роуминг. Разработан план размещения БПС на территории всех европейских стран.

В настоящее время в четырех странах Западной Европы (Великобритании, Франции, Италии и Швеции) установлены шесть наземных станций радиосвязи для проведения опытной эксплуатации системы TFTS. Сейчас более 350 самолетов Швеции, Франции и Великобритании оснащены оборудованием системы радиосвязи TFTS. Полное развертывание этой системы должно закончится в конце первого десятилетия XXI в.

### Глобальная сеть *Internet*


Т. Бернерс-Ли

Компьютерные сети по территориальной распространенности могут быть локальными, региональными и глобальными. *Локальные* — это сети, перекрывающие территорию не более 10 км$^2$; *региональные* — расположенные на территории города или области; *глобальные* — на территории государства или группы государств, например всемирная сеть *Internet*. Первая локальная компьютерная сеть типа *Internet* создана в 1969 г. в США (главным идеологом сети *Internet* считается американец греческой национальности Майкл Дертуозос — *Michael Dertuozos*; 1937–2001). Изобретателем современной сети Internet (разработчиком *Internet*-протоколов) является британец Тим Бернерс-Ли (*Tim Berners-Lee*; р. 1955 г.). Условно технология организации обмена информацией между различными техническими системами и сетями связи называют *WWW-технологией* (*WWW — World Wide Web — всемирная паутина*), а сеть, организуемая из систем, линий связи, спутниковых каналов и некоторых дополнительных сервисных устройств (серверов) и компьютеров, — *сетью Internet*.

В литературе встречаются многочисленные определения понятия «*Internet*», некоторые из них, собственно, и не являются определениями, поскольку не раскрывают содержание термина и его особенности. Более подходит следующее определение: «*Internet* — объединение сетей связи, соединенных между собой по протоколам *TCP/IP* (протоколам *Internet*) так, что в любой момент времени каждый компьютер, включенный в Internet, может передать каждому другому компьютеру пакет данных *IP* и получить ответ за доли секунды». Но и это определение страдает недостатком, так как не учитывает наличие в *Internet* многих других служб, которые не только передают данные. Можно предложить такое определение: «*Internet* — совокупность различных телематических служб (*telematic service* — службы электросвязи) и служб данных, которая базируется на взаимно соединенных сетях связи и построена на основе стандартов, получивших название протоколов *TCP/IP* или протоколов *Internet*». Эти стандарты разрабатывает Инженерная проблемная группа *Internet* (*IETF*), являющаяся подкомитетом руководящего органа *Internet* — «Общества *Internet*» (*ISOC*). Кроме *IETF* в состав *ISOC* входит Центр сетевой информации *Internet* (*INIC*), осуществляющий регистрацию пользователей и оказывающий им информационные услуги, подкомитет Исследовательский центр *Internet* (*IRET*) и Координатор работы *Internet* (*IANS*), отвечающий за единую систему адресации.

Стандартизация *Internet* развивалась постепенно по мере развития самой глобальной сети, начиная примерно с 1969 г. Основой современного *Internet* являются два главных протокола: протокол *IP* — межсетевой, обеспечивающий только доставку информации по заданному адресу, и протокол *TCP*, обеспечивающий надежную доставку информации, сохранение порядка следования пакетов, качества передачи информации. Разработка стандартов *Internet* относится к

80-м гг., в конце которых и начале 90-х гг. XX в. начался бурный рост *Internet*, обусловленный рядом причин, в том числе коммерцизацией сети.

Бытует мнение, что у *Internet* нет единого хозяина, что это нечто подобное общественной организации. На самом деле это далеко не так. Например, *IETF* принадлежит ряду компаний, в число которых входят *U.S. Robotics Asend, Telecommunications, Microsoft*, и др. В настоящее время число ее пользователей в мире превысило 500 млн, из них почти 10 млн абонентов находятся в России.

### Беспроводные сети технологии Wi-Fi

Основной технологией построения беспроводной сети, позволяющей выход в Internet, на данный момент является технология *Wi-Fi* (читается «вай-фай», сокращение от англ. *Wireless Fidelity*, что переводится как высокая точность беспроводной передачи данных).

Технология Wi-Fi (иногда говорят «беспроводной *Internet*») обозначает высокочастотную беспроводную локальную сеть (*WLAN*), которая обеспечивает высокоскоростную связь с *Internet*. Такая беспроводная сеть позволяет передавать и получать информацию через Internet при помощи радиосигнала. По сути, этот сигнал почти ничем не отличается от радиосигнала, с которым работает сотовый телефон. Беспроводную сеть Wi-Fi можно подключить к глобальной информационной сети, тем самым обеспечить беспроводной Internet в любой точке помещения.

Wi-Fi-сети обладают рядом существенных преимуществ по сравнению с кабельными сетями:

- масштабируемость (легко изменять размер и топологию сети);
- отсутствие монтажно-строительных работ при построении Wi-Fi-сети.

Существуют несколько стандартов, в которых работают сети Wi-Fi. Самые популярные из них — IEEE 802.11 a, b и g. Они работают на частотах 2,4 ГГц или 5,5 ГГц и обеспечивают скорость соединения 11 Мбит/с и 54 Мбит/с. Прямая и обратная совместимость стандартов позволяет устройствам разных стандартов успешно работать друг с другом.

Wi-Fi может использоваться для доступа в Internet на расстоянии в несколько километров. Как правило, одна точка доступа может обеспечить радиус действия до 100...200 м. Помимо домашних, кабинетных и офисных сетей, Wi-Fi получил широкое распространение в сфере организации публичного доступа в Internet. Хот-спот (от англ. *hot spot* — «горячее пятно»; точка беспроводного доступа в Internet в публичном заведении) в университете, конференц-зале, ресторане, кафе, кинотеатре, гостинице, аэропорту, вокзале, школе, бизнес-центре, Internet-кафе или просто в небольшом городском парке дает любому возможность подключиться к сети при помощи ноутбука, карманного компьютера или смартфона (*smartphone*; по сути гибрид мобильного телефона и карманного компьютера), оснащенного Wi-Fi-адаптером. Итак, имея его, ноутбук и точку доступа к *Internet*, можно навсегда забыть о многих неудобствах. Возможность проверить почту, заказать гостиницу через Internet или воспользоваться *Internet* -телефонией всегда будет в точках доступа: в аэропорту, гостинице или кафе.

Как подключиться к Wi-Fi? Прежде всего, необходим ноутбук со встроенным или дополнительно приобретенным Wi-Fi-адаптером и хот-спот. Большинство ноутбуков оснащены встроенным беспроводным адаптером; для остальных можно использовать PCMCIA-адаптер беспроводной сети. Вы устанавливаете на свой ноутбук небольшую программу и вставляете адаптер в разъем. При входе в зону действия хот-спота адаптер автоматически найдет точку доступа, и это отразится индикацией беспроводного соединения в углу экрана.

Обеспечение безопасности информации в сетях Wi-Fi. Пока (на начало 2009 г.) трудно рассчитывать на то, что беспроводная сеть надежно защитит вашу информацию, или наоборот надо опасаться того, что данные будут открыты для всеобщего доступа. С точки зрения безопасности уязвима не сама технология Wi-Fi, а операционная система *Windows*, для которой с завидной регулярностью выходят заплатки, закрывающие обнаруженные (тщательно вычисленные хакерами) слабые стороны.

Беспроводная сеть Wi-Fi была спроектирована для обмена файлами между компьютерами (главным образом ноутбуками). А защита файлов осуществляется поверх сети. К сожалению, с приходом новой технологии появляются новые угрозы безопасности и конфиденциальности информации. Wi-Fi задействует протоколы защиты WEP и WPA. Теоретически хакер может войти в Wi-Fi-зону и попытаться прослушивать то, что у вас идет по беспроводной сети, как и любой радиообмен. Но в отличие от радио, данные по Wi-Fi поставляются закрытым способом, чему способствует WEP-шифрование. WEP (*Wired Equivalent Privacy* — защита эквивалентная секретности) — характеристика стандарта 802.11, которая используется для обеспечения безопасности передачи данных. Она идентична протоколу безопасности в кабельных локальных сетях без применения дополнительных методов шифрования. Согласно стандарту 802.11 шифрование данных WEP используется для предотвращения несанкционированного доступа к данным при использовании беспроводных сетевых устройств и/или для предотвращения перехвата трафика беспроводных локальных сетей. Более современный протокол защиты WPA проверяет пользователей сети через сервер и задействует 128-битные ключи шифрования и динамические ключи сессии для обеспечения защиты беспроводной сети.

Фаервол/Firewall — барьер, отгораживающий вашу собственность от злоумышленников, который может быть аппаратным или программным. Для среднестатистического пользователя Wi-Fi правильно сконфигурированного программного фаервола должно быть вполне достаточно. Большой популярностью пользуются программы *Norton Firewall* и *Zone Alarm*. Нужно помнить, что фаерволы не являются 100 %-ной гарантией защиты, поскольку могут быть написаны новые вирусы, заражающие эту программу.

Для получения доступа к сети Wi-Fi нужен собственный ключ протокола WEP (или протокола WPA). Эти протоколы позволяют администратору беспроводной сети определять для каждого пользователя набор ключей, основанный на «строке ключей», которая обрабатывается алгоритмом шифрования. Любой пользователь, не имеющий требуемого ключа, не может получить доступ в сеть.

При включении WEP все станции (как клиентские, так и точки доступа) получают свой ключ, который применяется для шифрования данных, прежде чем последние будут поданы на передатчик. Если станция получает пакет, не зашифрованный соответствующим ключом, он исключается из трафика. Этот метод служит для защиты от несанкционированного доступа и перехвата данных. Подключение Wi-Fi по выделенной линии скоро перестанет доминировать, и хот-споты захватят города, как некогда это сделали сотовые вышки.

### *IP-телефония (Internet Phone; Интернет-телефония)*

В настоящее время телефонные сети общего пользования, достигнув пределов «совершенства», фактически начинают тормозить развитие технологии передачи сообщений. Коммутация телефонных каналов, в отличие от коммутации пакетов, более не в состоянии удовлетворять растущие потребности рынка связи, в том числе в новых и дополнительных услугах и снижении удельных затрат на расширение сетей. Хотя в последнее десятилетие благодаря компьютерно-телефонной интеграции отмечен определенный рост в развитии ТСОП, цена такой интеграции в сети с коммутацией каналов высока. За последние 10 ... 15 лет телекоммуникационная отрасль претерпевает радикальную перестройку. В начале 90-х гг. XX в. техническая революция перенесла человека из эры традиционной связи на основе коммутации каналов в эпоху более совершенной технологии передачи речи — пакетной телефонии. Сегодня пользователи уже осознают экономические и технологические преимущества новой технологии.

Концепция передачи голоса по сети с помощью персонального компьютера зародилась в университете штата Иллинойс (США). В 1993 г. Чарли Кляйн выпустил в свет Maven, первую программу для передачи голоса по сети с помощью персонального компьютера. Одновременно одним из самых популярных мультимедийных приложений в сети стала CU-SeeMe, программа видеоконференций для *Macintosh*, разработанная в Корнельском университете. Израильская компания «VocalTec» предложила первую версию программы *InternetPhone*, разработанную для владельцев мультимедийных PC, работающих под *Windows*.

IP-телефония — новейшая технология, которая используется в сети *Internet* для передачи речевых сигналов. IP-телефония — это предоставление услуг телефонной связи по сетям передачи данных, в частности сетям *Internet*. Что же такое IP-телефония как сетевая концепция? Это интеграция передачи данных и сетей телефонии при ведущем положении услуг передачи данных. Таким образом, как сетевая концепция IP-телефония предусматривает наличие сети передачи данных, где дополнительной услугой обеспечивается телефония.

В обычном телефонном звонке подключение между обоими собеседниками устанавливается через телефонную станцию исключительно с целью разговора. Голосовые сигналы передаются по определенным телефонным линиям, через выделенное подключение. Общий принцип действия телефонных серверов IP-телефонии таков: с одной стороны, сервер связан с телефонными линиями и может соединиться с любым телефоном мира. С другой стороны, сервер связан с Internet и может связаться с любым компьютером в мире. Сервер принимает

стандартный телефонный сигнал (голосовые сигналы, т. е. слова, которые мы произносим), оцифровывает и значительно сжимает его, преобразует в цифровые пакеты данных и отправляет в *Internet* с адресом назначения. При этом используется протокол *Internet* (TCP/IP). Для пакетов, приходящих из сети *Internet* на телефонный сервер и уходящих в телефонную линию, операция происходит в обратном порядке. Для адресата пакеты данных перегруппировываются и декодируются в голосовые сигналы оригинала. Обе составляющие операции (вход сигнала в телефонную сеть и его выход из телефонной сети) происходят практически одновременно, что позволяет обеспечить полнодуплексный разговор. На основе этих базовых операций можно построить много различных конфигураций. Допустим, звонок телефон–компьютер или компьютер–телефон может обеспечивать один телефонный сервер. Для организации компьютерной связи телефон(факс) – телефон(факс) нужно два сервера.

Заметим, что телефонная сеть была создана так, чтобы гарантировать высокое качество услуги даже при больших нагрузках. IP-телефония, напротив, не гарантирует качества, причем при больших нагрузках оно значительно падает.

Существуют два базовых типа телефонных запросов IP-телефонии:

- с компьютера на компьютер;
- с компьютера на телефон.

Качество связи IP-телефонии можно оценить такими характеристиками:

- уровень искажения голоса;
- частота «пропадания» голосовых пакетов;
- время задержки (между произнесением фразы первого абонента и моментом, когда она будет услышана вторым абонентом).

Анализ показывает, что для получения приемлемого качества речи задержки в сети должны быть минимальными, а дополнительная полоса пропускания, которая требуется для передачи речи в существующих корпоративных сетях, ничтожно мала по сравнению с «коммуникационными трубопроводами», необходимыми для транспортировки видео, графических файлов и потоков данных, порождаемых киберпутешественниками и видеоконференциями.

Интеграция телефонии и передачи данных — концепция не новая, она легла в основу другой технологии — ISDN, которая выполняет похожие задачи и, таким образом, является потенциальным конкурентом на телекоммуникационном рынке. В то же время технология ISDN принципиально отличается от IP-телефонии тем, что обеспечивает интеграцию сетей телефонии и передачи данных при ведущем положении услуг телефонии, поскольку использует более соответствующий телефонной связи принцип коммутации каналов.

Исходя из этого различия двух технологий, а также уровня развития сети и следует делать выбор той или другой концепции. Технология ISDN внедряется, когда необходимо передать данные по существующей телефонной сети. Если эта сеть аналоговая, то этапами внедрения ISDN будут: аналоговая телефонная сеть — цифровизация первичной и вторичной сетей — интегрированная цифровая сеть (ИЦС) или IDN — ISDN. Цепочка внедрения услуг IP-телефонии

длиннее: аналоговая телефонная сеть — цифровизация первичной и вторичной сетей — создание наложенной сети передачи данных — IP-телефония.

IP-пакетная телефония приносит с собой новые возможности:

- повышается эффективность использования полосы пропускания каналов за счет эффективных алгоритмов сжатия данных;
- обеспечивается будущая интеграция интерактивных мультимедийных услуг, например видео по запросу;
- организуется управление единой сетью передачи речи, данных и видео;
- используются существующие широко применяемые протоколы;
- выбирается способ телефонной связи, предоставляемый конечным пользователям и позволяющий экономить расходы на междугородные звонки.

Междугородняя (международная) связь осуществляется с помощью телефонных серверов, организация или оператор услуги должны иметь по серверу в тех местах, куда и откуда планируются звонки. Стоимость такой связи на порядок меньше стоимости телефонного звонка по обычным телефонным линиям. Особенно велика эта разница для международных переговоров.

### *Волоконно-оптические линии (системы) связи*

Волоконно-оптическая линия (система) связи — вид сетей передачи сообщений, при котором информация передается по линиям связи, которые представляют собой диэлектрические оптические волноводы. Такие волноводы называют *оптическим волокном*.

На рис. 1.48 показано простейшее оптическое переговорное устройство. Когда к началу 60-х гг. XX в. появились первые пригодные к эксплуатации и надежные лазеры, стало очевидным, что свет предстал в новом качестве — когерентное электромагнитное колебание на несколько порядков увеличило границы применяемого в радиотехнике и технике связи диапазона частот. Оптимистические расчеты едва ли можно опровергнуть: длина волны в 1 мкм соответствуют частоте $3 \cdot 10^{14}$ Гц. Если принять лишь 1 % этого диапазона в качестве полосы сигналов, которыми можно модулировать данное колебание, то получим значение 3 000 ГГц. Это соответствует приблизительно миллиарду телефонных разговоров (полоса телефонного сигнала около 3 000 Гц) или миллиону телевизионных программ (полоса частот телевизионного сигнала составляет 8 МГц), которые можно передать одним световым лучом.

Известно, что самый лучший и дорогой коаксиальный кабель с медными проводниками обладает едва ли одной тысячной долей этой пропускной способности и что в будущем крайне необходимо будет передавать информацию очень большого объема.

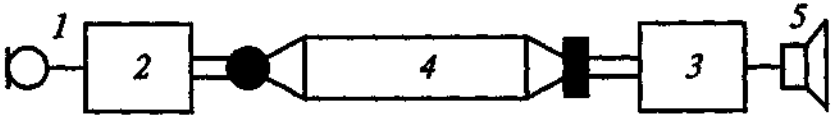



**Рис. 1.48.** Простейшее оптическое переговорное устройство:
*1* — микрофон; *2*, *3* — усилители сигнала; *4* — ВОЛС; *5* — телефон

Число телефонных абонентов в мировой телефонной сети неудержимо повышается, а растущие хозяйственные и промышленные отношения между странами и континентами требуют все больше качественных каналов связи. Когда же в сферу рассмотрения перспективных проектов включили возможность использования *видеотелефона* (а передача даже одного изображения требует почти тысячекратной пропускной способности по сравнению с телефонным сигналом), то стало необходимым считаться с сильно возросшей потребностью в каналах передачи информации.

На рис. 1.49 представлена упрощенная структурная схема волоконно-оптической системы связи. Передаваемые сигналы через *ИКМ-линию* и *преобразователь кода* (ПК) поступают в *оптический передатчик*, основой которого является *электронно-оптический преобразователь* (ЭОП) — полупроводниковый лазер. ИКМ-линия и преобразователь кода создают кодовую импульсную последовательность, а *согласующее устройство* (СУ) согласует ЭОП с *оптическим кабелем* (ОК). В приемнике производится обратное преобразование *оптоэлектронным преобразователем* (ОЭП) оптического сигнала в электрический. В качестве ОЭП используют полупроводниковый фотодиод, в котором под воздействием поступающего через СУ светового потока возбуждается электрический сигнал. С ОЭП этот сигнал через ПК поступает в *импульсно-кодовый демодулятор* (ИКД-линию).

Важной составной частью ВОЛС являются *линейные регенераторы* (ЛР) — магистральные усилители оптического сигнала, включенные в сети ОК. В таких системах, как правило, используется цифровая многоканальная связь с временны́м уплотнением (разделением) каналов и импульсно-кодовой модуляцией оптической несущей.

Особенности построения линейных кодов для цифровых волоконно-оптических систем связи зависят от физических свойств среды распространения передаваемых сигналов. Оптическое волокно, а также источник излучения в передающем и фотоприемник в приемном оптических модулях предъявляют специфические требования к свойствам цифрового сигнала.

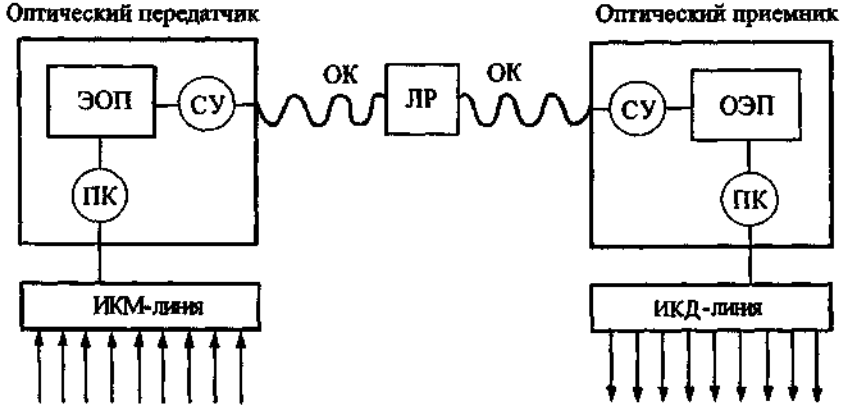



**Рис. 1.49.** Упрощенная структурная схема волоконно-оптической системы связи

Поскольку импульсные посылки излучаемой оптической мощности могут быть только положительными или нулевыми (известно, что интенсивность оптического излучения является по самой природе положительной величиной), невозможно непосредственное использование биполярных кодов, применяемых при передаче по электрическим кабелям связи.

Надо отметить, что человечество не впервые обращается к волнам оптического диапазона для целей связи. С древних времен люди пользовались световой сигнализацией, а в 70-е гг. XIX в. А.Г. Белл демонстрировал оптический беспроволочный телефон. В наше время оптические линии связи используются очень широко. Получены и промышленностью производятся световоды с потерями, меньшими 1 дБ/км (это меньше, чем потери света при прохождении через оконное стекло!), что дает возможность размещать переприемные станции линий дальней связи на расстоянии 30...40 км друг от друга. Таким образом, мы находимся в пределах нового этапа развития, когда на смену электрону приходит фотон, обладающий существенно более широкими возможностями в плане передачи информации.

### *Портативные антенны мобильных телефонов*

Бурное внедрение в последние годы сотовой связи потребовало уменьшения массы и габаритных размеров телефонных трубок и их приемопередающих антенн. Разработку портативных антенн мобильных телефонов определяют два основных аспекта — приемлемая широкополосность (около 10 % относительно значения несущей частоты) и максимальная равномерность электромагнитного излучения по азимутальному углу с высоким коэффициентом усиления. Эти условия противоречат тому факту, что антенна подносится близко к голове пользователя, которая по электрическим параметрам близка к поглощающему диэлектрическому эллипсоиду с размером главной оси, соизмеримой с длиной волны. При телефонных переговорах голова абонента поглощает и рассеивает электромагнитную энергию, излучаемую антенной, так, что нарушается азимутальная (меридиональная) равномерность излучения.

Наиболее простой способ достижения равномерного азимутального излучения — размещение антенны выше головы пользователя. Это требует применения для антенны неизлучающего основания размером не менее 15 см. Антенна в виде полуволнового симметричного вибратора или аналогичного излучателя имеет почти такую же длину. Общий размер основания и излучателя в этом случае составит около 30 см. Такие габаритные размеры недопустимы, так как размер сотового телефона не должен превышать 10 см, что приблизительно соответствует расстоянию между ртом и ухом взрослого человека. Разборная конструкция длиной 30 см представляет серьезные проблемы в постоянно уменьшающихся объемах мобильных радиотелефонов, поэтому на рынке подвижной связи нашли применение относительно малогабаритные антенны.

Большинство излучателей, используемых в мобильных радиотелефонах, представляют собой шлейфовые (штыревые), реже спиральные и конечно современные низкопрофильные антенны.

### Шлейфовые антенны

Работающая в диапазоне 800...900 МГц шлейфовая антенна (рис. 1.50), представляет собой расположенный вертикально симметричный полуволновый вибратор, имеющий круговую диаграмму направленности в азимутальной плоскости. Вибратор состоит из двух проводников разного диаметра. Один из них соединен с центральным проводом питающей коаксиальной линии и должен иметь длину, близкую к $\lambda/4$, чтобы достичь хорошего согласования антенны в рабочей полосе частот. Проводник с бóльшим диаметром и разомкнутым концом (шлейф) имеет вид металлического стакана. Он соединен с внешней оплеткой питающей коаксиальной линии и выполняет функции реактивного сопротивления — согласующего дросселя для высокочастотных токов. Дроссель наиболее эффективен, если отрезок линии, образованный внешним проводником коаксиальной линии и внутренней поверхностью шлейфа, настроен в резонанс на несущую частоту. Резонансная длина металлического шлейфа должна быть немного меньше $\lambda/4$.

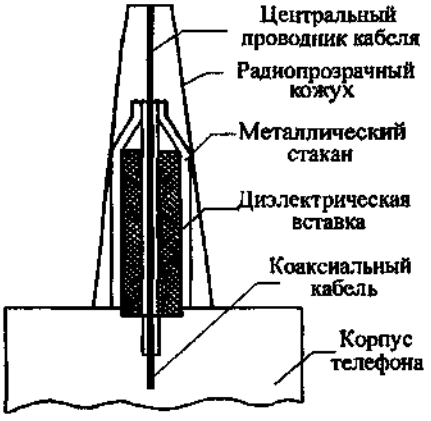



**Рис. 1.50.** Шлейфовая антенна

Работа антенны зависит от геометрических размеров питающей коаксиальной линии и формы металлических поверхностей корпуса телефона. Полуволновый вибратор имеет неоднородности диаметра в плече шлейфа-стакана и точке питания антенны. Ухудшения в работе наблюдаются за пределами диапазона ± 5 % от частоты резонанса, поскольку в этом случае высокочастотные токи потекут по внешней поверхности питающей линии. Протекающие по корпусу радиотелефона высокочастотные токи частично поглощаются рукой абонента, увеличивая дополнительные омические потери и снижая коэффициент усиления антенны.

### Цилиндрические спиральные антенны

Такие антенны широко использовались в ранних конструкциях мобильных радиотелефонов. При соответствующем выборе параметров спиральные антенны (рис. 1.51) имеют обычно приемлемые габаритные размеры и весьма эффективны по равномерности электромагнитного излучения и коэффициенту усиления. Однако их рабочая полоса частот сравнительно невелика. Цилиндрическая спиральная антенна (см. рис. 1.51, *а*) представляет собой свернутый в спираль проводник *1*, который питается через коаксиальную линию *2*. Внутренний провод коаксиальной линии соединяется со спиралью, а внешняя оплетка — с небольшим металлическим диском *3*.

На рис. 1.51, *а* показаны размеры спирали антенны: $a$ — радиус; $s$ — шаг; $L$ — длина по ее оси.

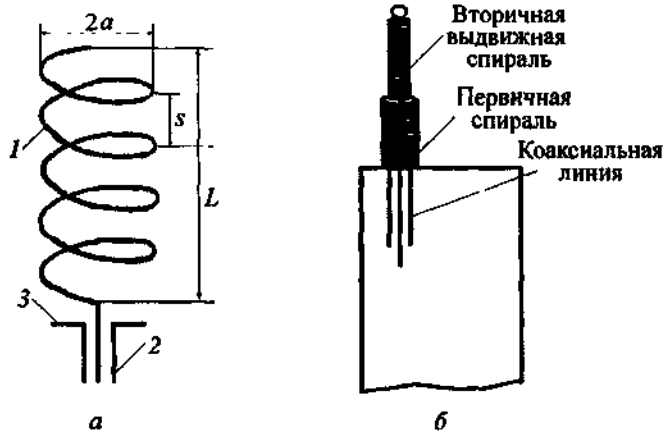



**Рис. 1.51.** Цилиндрическая спиральная антенна:
*а* — спираль; *б* — двойная спираль на корпусе радиотелефона

В спиральных антеннах портативных телефонов используется режим ненаправленного излучения, который реализуется при диаметре спирали $2a$, значительно меньшем длины волны ($\lambda > 12a$). При этом в плоскости, совпадающей с осью спирали, диаграмма направленности имеет форму круговой восьмерки, а в плоскости витков антенна равномерно излучает во всех направлениях. Поскольку антенна (вместе с телефоном) практически всегда расположена вертикально, то используется меридиональное излучение радиоволн.

В режиме излучения электромагнитных колебаний спиральная антенна радиотелефона на резонансной частоте возбуждает значительные уровни высокочастотных токов на корпусе радиотелефона, который становится частью излучающей системы. Возбужденные токи корпуса радиотелефона частично поглощаются рукой абонента, и ее ладонь необходимо рассматривать как диэлектрический материал с омическими потерями, обернутый вокруг корпуса. Часть токов проходят от корпуса в руку и рассеиваются там, создавая дополнительный механизм потерь. Снижение коэффициента усиления спиральной антенны составляет около 3 дБ для человека с сухими руками.

Радиотелефонная спиральная антенна достаточно мала и в значительной степени затеняется головой абонента, находясь от нее на расстояниях 2...4 см. Провалы в диаграмме направленности спиральной антенны по сравнению с ДН симметричного полуволнового вибратора достигают 10–12 дБ. Поэтому единственное достоинство спиральных антенн телефонов — их небольшой размер.

Сравнительно низкоэффективные рабочие характеристики и показатели спиральных антенн в диапазонах рабочих частот 800...900 МГц заставили разработчиков прибегнуть к их усложнению. Применяемые в настоящее время цилиндрические спиральные антенны состоят из двух спиралей (см. рис. 1.51, *б*) — *первичной*, жестко установленной на корпусе (длина 2...2,5 см при четвертьволновой электрической длине), и *вторичной* (длина 10 см при полуволновой электриче-

ской длине). Вторичная спираль размещается внутри корпуса телефона и в выдвинутом состоянии становится основным излучателем. Она запитывается концом первичной спирали, и в ней отсутствуют потери, вызванные рукой абонента, держащей корпус телефона. Для уменьшения вредного влияния рассмотренных типов антенн на человека их иногда встраивают в низ корпуса телефона.

### Низкопрофильные антенны

В последние несколько лет в большинстве мобильных станций систем подвижной радиосвязи стали широко применяться нетрадиционные излучатели нового типа, в частности, низкопрофильные антенны. Наибольшее распространение в мобильных радиотелефонах получили микрополосковые и *F*-образные низкопрофильные антенны (рис. 1.52).

Базовая, или проволочная, модель *F*-образной низкопрофильной антенны показана на рис. 1.52, *а*. Данный вид антенны может быть получен изгибом несимметричного четвертьволнового вибратора в *L*-образную форму и установленного на проводящей плоскости корпуса. Возбуждается антенна в смещенной от основания точке, к которой подключается внутренний проводник коаксиальной питающей линии. Внешняя оплетка коаксиальной линии соединяется с проводящей плоскостью корпуса.

Согласование полных сопротивлений *F*-образной антенны и линии достигается выбором положения точки питания. Частотный диапазон функционирования *F*-образной антенны пропорционален высоте *H*.

Низкопрофильная антенна стала основой для более эффективной плоскостной *F*-образной антенны, в которой излучающий элемент выполняется в виде полоски на двусторонней диэлектрической подложке (рис. 1.52, *б*). Подключение коаксиальной питающей линии в данной конструкции осуществляется так же, как и в проволочной модели. Плоскостная антенна является одним из основных типов внутренних встроенных антенн для перспективных моделей радиотелефона, и она может располагаться практически в любой части его корпуса.

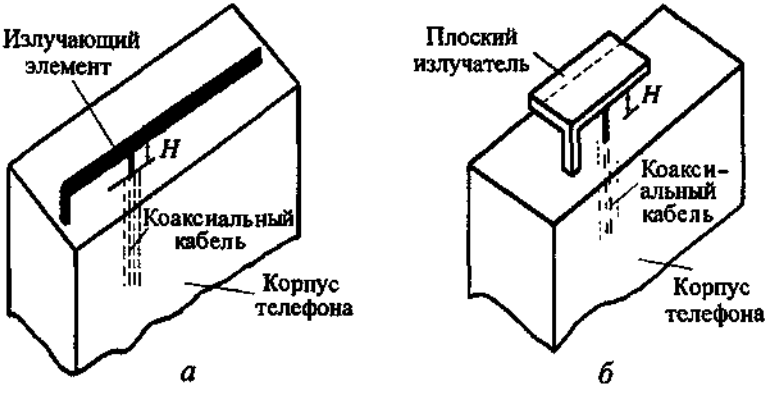



**Рис. 1.52.** Низкопрофильные антенны:
*а* — базовая проволочная; *б* — полосковая плоскостная

***Влияние электромагнитного излучения на организм человека***

Шведские ученые провели исследование, результаты которого опубликовали в «European Journal of Cancer Prevention» («Европейский журнал предупреждения рака») и которое основано на сравнении результатов обследования около 2 000 шведских пациентов, у которых между 2005 и 2008 гг. были обнаружены опухоли мозга, и аналогичной по численности контрольной группы здоровых людей. Исследователи обнаружили, что для пациентов, пользовавшихся мобильными телефонами стандарта NMT, риск возникновения раковой опухоли мозга в том месте, где трубка прикладывалась к голове, возрастал на 30 %, по сравнению с теми, кто не использовал такие телефоны. Для тех, кто использовал стандарт NMT более 10 лет, риск возрастал на 80 %.

Немаловажным фактором, определяющим разработку и внедрение систем подвижной связи, стало обеспечение безопасности человека с точки зрения воздействия электромагнитного излучения. До сих пор никто из ученых и специалистов категорически не утверждает, что вредного воздействия от радиотелефонов нет, равно как никто не может утверждать, что это воздействие в определенной мере вредно. Развитие систем подвижной связи сопряжено с облучением населения небезвредным электромагнитным полем, создаваемым антеннами их мощных передающих устройств.

Проблема защиты человека от ЭМИ состоит в том, что излучающие устройства находятся на в густонаселенных районах городов: антенны базовых станций, установленные на крышах, углах зданий; мобильные станции, находящиеся на автомобилях, — все они облучают и вышедших на связь абонентов и прохожих на улицах. Люди, пользующиеся ручными радиостанциями и телефонами на улицах и общественных местах, также подвергают электромагнитному облучению как себя, так и окружающих. Хотя мощности этих аппаратов невелики, но их излучающие антенны расположены в непосредственной близости от головы человека. Очевидно, что пользователи мобильных станций находятся в *ближней зоне* (поле индукции) излучения портативных антенн, а из-за пониженного (по отношению к свободному пространству) характеристического сопротивления среды в поле индукции уровень *абсорбируемой* (поглощаемой) человеком мощности излучения возрастает.

В результате многочисленных исследований учеными установлено, что реакция организма человека на ЭМИ зависит не только от величины мощности излучения, но и от таких его параметров, как несущая частота, вид модуляции, ширина спектра излучаемого радиосигнала, поляризация радиоволн, время облучения и т. д. Выявить однозначную связь между каким-либо одним параметром излучения и явно выраженным эффектом его воздействия на организм человека пока не удалось. Проявление биологической активности характеристик и параметров электромагнитного поля в значительной степени зависит от взаимного расположения излучающей антенны и облучаемого человека, наличия в районе их размещения отражающих предметов и заземления, что особенно характерно для мобильных станций.

Анализируя вопросы обеспечения электромагнитной безопасности МС, расположенных на подвижных объектах (автомобильные, портативные и ручные радиостанции и радиотелефоны), следует отметить ряд особенностей структуры ЭМИ, воздействующего на пользователя:

- расположение абонента в ближнем поле излучения;
- соизмеримость размеров органов человека с длиной излучаемых волн;
- высокая степень влияния металлических предметов и конструкций на формирование поля ЭМИ;
- отсутствие достаточно надежного экранирования подвижного абонента от непосредственного воздействия излучения антенн.

Зависимость поглощения энергии ЭМИ и ее распределения внутри тела человека определяется не только электрическими свойствами тканей, но и формой и размерами органа, а также соотношением этих размеров с длиной волны излучения. Анализ показал, что для частот 0,03...10 ГГц характерно наличие ряда максимумов поглощения, при которых тело человека как бы впитывает в себя поле ЭМИ и поглощает энергии больше, чем приходится на его поперечное сечение. В этом случае резко проявляются интерференционные картины, приводящие к сильной корреляции как общей поглощенной энергии, так и ее распределения в зависимости от полей конкретных длин волн, размеров и анатомического строения органов и электрических свойств тканей человека. Локальные максимумы поглощения телом человека («горячие точки») ЭМИ имеют место в диапазоне частот 750...2 500 МГц, а максимум, обусловленный резонансом с общим размером тела, лежит в диапазоне частот 50...300 МГц.

При всесторонних биологических исследованиях установлено, что наиболее чувствительны к воздействию ЭМИ центральная нервная система, глаза и гонады (эндокринные железы половых органов). В результате такого воздействия может нарушаться деятельность сердечно-сосудистой, кроветворной, нейроэндокринной, иммунной систем и обменных процессов. При рассмотрении принципов определения предельно допустимых уровней (ПДУ) электромагнитного излучения следует иметь в виду, что в качестве нормируемого параметра принята энергия поля, падающая на единицу поверхности тела человека при рабочей частоте передатчика, соответствующей наиболее выраженной реакции организма.

В большинстве зарубежных стран ПДУ электромагнитного излучения *при воздействии на человека более 6 мин* составляет 1...5 мВт/см$^2$ площади тела. Этот разброс связан как с различиями в методологических подходах к регламентации и критериям оценки вредности фактора, так и с недостаточной изученностью его биологического действия на различные органы человека.

Стандарты США, на которые ориентированы большинство национальных стандартов других стран, а также международные рекомендации исходят из повреждающего теплового действия электромагнитного фактора. Конкретные значения ПДУ для частот выше 10 МГц рассчитаны по показателю удельного поглощения мощности, который не должен превышать 2 мВт/г массы тела. При этом допускается усреднение на массу всего тела или его часть на время воздействия, ширину импульса излучения, период модуляции.

В последние годы в Российской Федерации ПДУ ЭМИ систем подвижной радиосвязи регламентирует нормативный документ — Гигиенические нормативы ГН 2.1.8./2.2.4.019 – 94 «Временные допустимые уровни (ВДУ) воздействия электромагнитных излучений, создаваемых системами сотовой радиосвязи». Согласно этим нормативам, плотность потока энергии (ППЭ ПДУ) на облучение пользователей антеннами радиотелефонов не должна превышать 100 мкВт/см$^2$. Кстати, принципиальным моментом этого документа является то, что в нем введены две категории электромагнитного облучения — профессиональное и непрофессиональное воздействия.

Основной причиной, влияющей на различие стандартов разных стран, следует считать недостаточную изученность биологического воздействия ЭМИ на человека. В частности, это зависимости биологических эффектов от параметров излучения (частота, характер и мощность генерации) и условий воздействия на организм (систематическое, прерывистое, эпизодическое и т. д.).

### *Глобальные системы связи будущего*

Отметим, что задачи в мирового масштаба — обеспечение согласованности параметров оборудования систем связи, разрабатываемого, производимого и эксплуатируемого во всех странах, а также определение перспектив развития электросвязи — решает Международный Союз Электросвязи. В частности, стандартизирующим органом МСЭ является сектор стандартизации электросвязи МСЭ-Т (здесь буква Т означает *телекоммуникации*). МЭС-Т ведает также стандартизацией вопросов включения систем радиосвязи в сети электросвязи общего пользования. Имеется обширный и постоянно пополняющийся набор рекомендаций комитета МСЭ-Т по всем сторонам работы телекоммуникационных систем и сетей. Основной стратегической целью МСЭ является создание глобальной информационной структуры, в результате чего будет реализована глобализация и персонализация связи. Глобализация и персонализация связи — это две тенденции развития телекоммуникационных систем и сетей, предусматривающие совершенствование связи по сетевым и техническим (или технологическим) направлениям.

*Глобализация связи* — это создание Всемирной (глобальной) сети связи, в которую интегрируются (объединяются) *национальные сети связи* (сети связи отдельных государств), а также входящие в них *региональные* и *абонентские сети связи*. Это позволит любому абоненту земного шара в любое время пользоваться необходимыми ему услугами связи.

*Персонализация связи* согласно концепции UPT (*Universal Personal Telecommunication*) выражается в том, что каждый житель Земли с момента рождения получает персональный номер, который регистрируется во Всемирной сети связи, а поиск абонента при адресации ему информации осуществляется автоматически с помощью интеллектуальных функций сети связи.

В настоящее время под эгидой МСЭ ведутся работы по созданию глобальной международной ССПС третьего поколения, получившей наименование *FPLMTS* (*Future Public Land Mobil Telephone System* — будущая сухопутная мо-

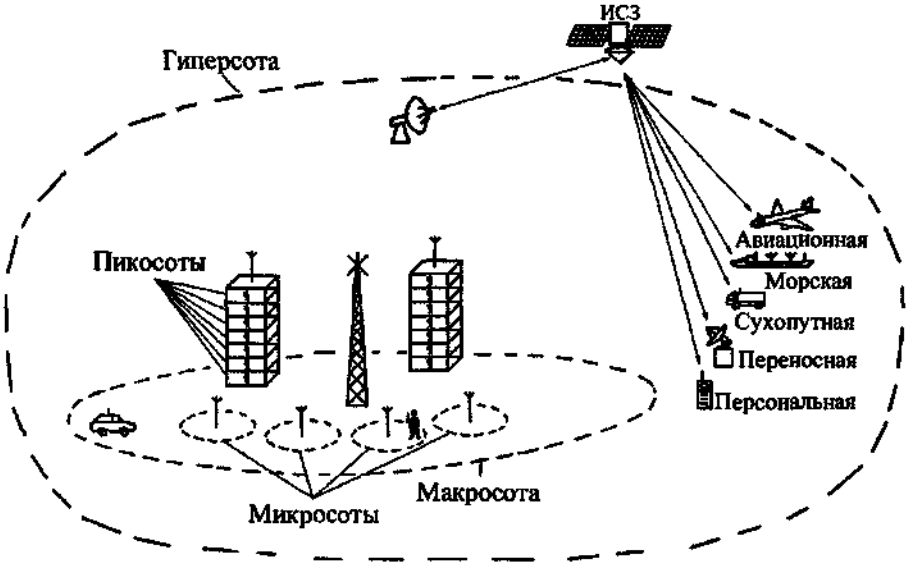



**Рис. 1.53.** Архитектура построения будущей системы мобильной связи

бильная телефонная система общего пользования). В этой системе глобальной радиотелефонной связи предполагается разработка наземного и космического сегментов. Архитектура ее построения может включать ряд различных сценариев. Один из возможных вариантов построения будущей глобальной системы подвижной связи с использованием космических, сотовых и беспроводных сетей представлен на рис. 1.53 (для упрощения на рисунке соты обозначены не шестигранными, а круговыми ячейками).

Внутри помещений зданий и офисов организуются пикосоты с радиусом обслуживания до 100 м с высокой емкостью, определяемой большим количеством абонентов на единицу обслуживаемого объема. В общественных местах (вокзалах, аэропортах, торговых центрах), на улицах для пешеходов создаются микросоты с радиусом действия до 1 км. Предполагается, что для обеспечения связью автомобилистов будут использовать сотовые системы связи, имеющие макросотовую структуру с радиусом обслуживания до 25...30 км. Спутниковый сегмент системы обслуживает абонентов воздушных, морских и речных судов, грузового автомобильного и железнодорожного транспорта с развертыванием *гиперсотовой* структуры радиусом до сотен и тысяч километров. Начало внедрения наземных компонентов системы FPLMTS намечено на 2010 г., ввод спутниковой подсистемы ожидается в 2015 г. Глобальная система связи FPLMTS будет действовать как единое целое, работающее по принципу установления связи между абонентами «где угодно, когда угодно и с кем угодно».

Подводя краткие итоги, отметим, что переход в более высокие частотные диапазоны, появление новых радиотехнологий, ориентация сетей третьего поколения на использование спутниковых сегментов передачи информации — один

из основных аспектов концепции развития систем подвижной радиотелефонной связи общего пользования. Заметим также, что в настоящее время происходит достаточно быстрой процесс *конвергенции* (слияния) связи и информатики в *инфотелекоммуникационные* технологии, переходящие к *инфосфере*, и Российская Федерация должна быть готова к этому.

## *Бытовые радиоэлектронные системы*

В последние годы радиоэлектронная аппаратура и элементы систем радиосвязи прочно вошли в быт людей. Коротко остановимся на некоторых из этих радиоэлектронных устройств.

### Компакт-диски

Среди множества способов хранения аудио- и видеоинформации наибольшую плотность записи и емкость носителя обеспечивает оптический компакт-диск (*compact disc — CD*). Появление компакт-дисков связано с началом разработки в 1977 г. фирмами *Sony* и *Philips* принципиально новых цифровых грампластинок. Цифровые грампластинки явились результатом многих лет постепенного развития техники записи и цифровых методов передачи информации. В итоге и были созданы цифровые грампластинки типа CD. Качество записи у таких дисков существенно выше и эффективнее, чем у долгоиграющей пластинки; компакт-диск имеет небольшую массу и габаритные размеры, их очень удобно хранить, они фактически не подвержены износу, на воспроизведении не сказывается пыль и небольшие царапины; заранее можно запрограммировать воспроизведение желательного отрывка записи на диске.

Цифровой компакт-диск (рис. 1.54) представляет прозрачную поликарбонатную круглую пластину толщиной 1,2 мм и диаметром 120 мм, покрытую тончайшим слоем металла (алюминий, золото, серебро и др.). Вся полезная информация записывается на светоотражающий слой в виде дорожек мельчайших выемок (*питов*, от англ. *pit* — ямка, углубление) и отражающих площадок (флэтов, от англ. *flat* — плоскость) (рис. 1.54, *а*). Глубина питы — 0,1 мкм, ширина — 0,6 мкм (в 30 раз тоньше человеческого волоса). При этом действует цифровой способ записи информации питы и флэты соответствуют единицам и нулям цифрового сигнала.

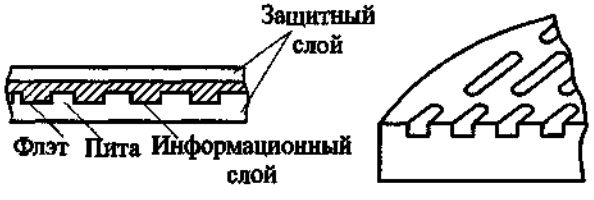



**Рис. 1.54.** Цифровой компакт-диск:
*а* — структура; *б* — питы на металле

Для записи информации используются девять длин пит (рис. 1.54, *б*). Минимальная длина питы около 0,9 мкм, длины следующих пит превышают эту длину в 4/3, 5/3, 2, 7/3, 8/3, 3, 10/3, 11/3 раз. Расположенные в ряд питы образуют дорожку записи. Высокий уровень развития механики позволяет стабильно воспроизводить сигнал при использовании соответствующих оптических устройств (лазерного облучателя компакт-диска и светочувствительного элемента — фотодиода) и техники цифровой обработки сигнала.

В электронной системе воспроизведения информации используется оптический полупроводниковый лазер с площадью излучения 0,5×0,5 мм$^2$, мощностью излучения 2...3 мВт, длиной волны 0,78 мкм. Луч лазера фокусируется на дорожке записи (диаметр пятна 2 мкм) и отражается флэтами металлической пленки на фотодиод. Если же луч падает на питу, он рассеивается и не отражается к фотодиоду. Таким образом, с помощью флэтов и пит осуществляется модуляция отраженного сигнала и считывание информации. Царапины со стороны непрозрачной защитной пленки не страшны, если не повреждены питы, так как луч лазера туда не доходит. Царапины менее 1 мм со стороны прозрачного диска также не оказывают особого влияния, поскольку на этой поверхности диска луч лазера еще не сфокусирован и имеет диаметр около 1 мм; небольшие царапины не создают препятствий для прохождения луча, и не возникают помехи воспроизведению. Информационный сигнал при цифровой записи с помощью аналого-цифрового преобразователя превращается в цифровую форму аналогично записи, используемой в компьютере. Эта последовательность импульсов используется при записи сигнала на компакт-диск в процессе формирования последовательности флэтов и пит различной длины. При воспроизведении промодулированный импульсами отраженный оптический сигнал попадает на фотодиод оптической воспроизводящей системы, а после усиления и обработки — на ЦАП, где превращается в аналоговую форму.

До последнего времени запись на компакт-диск производилась только с одной стороны; длительность звучания превышает 1 ч (максимально 74 мин). В последнее время были разработаны другие, аналогичные цифровые носители аудио- и видеоинформации — *цифровой видеодиск — digital video disc* (DVD). При внешней схожести диск DVD отличается от компакт-диска внутренней структурой. При той же толщине он может иметь до четырех информационных слоев (по два с каждой стороны), причем один слой (внешний) из каждой пары полупрозрачный, а другой (внутренний) — полностью отражающий. Сейчас появились видеозвуковые пластинки диаметром 20 см, на которых записаны звуковые программы с видеосюжетами.

### Системы видеоконференцсвязи

Системы видеоконференцсвязи предназначены для проводимых на расстоянии по каналам связи конференций, совещаний, деловых переговоров и т. д. (кстати, видеоконференцсвязь можно осуществлять и по глобальной сети Internet). Другими словами, эти системы используются для передачи малоподвижных изображений с голосовым сопровождением. Как правило, требования к ка-

честву воспроизводимого изображения менее жесткие, чем в вещательном телевидении. Первая сеть видеоконференцсвязи начала действовать в 1984 г. в ФРГ. В современных цифровых системах видеоконференцсвязи наиболее часто применяются цифровые потоки со скоростью передачи информации 2,048 Мбит/с в Европе и 1,544 Мбит/с в США.

Используемые в настоящее время системы видеоконференцсвязи обеспечивают разрешающую способность по горизонтали в 320 пикселей на строку и по вертикали в 286 строк. Объем информации на кадр равен 700 Кбит и скорость передачи (при 25 кадрах в секунду) должна составлять около 18 Мбит/с. Применение специального вида импульсно-кодовой модуляции сокращает этот поток в два раза — до 9 Мбит/с, а эффективное кодирование дополнительно сокращает цифровой поток до 7 Мбит/с. При разработке систем видеоконференцсвязи учитывается, что передаваемые по каналам изображения малоподвижны и изменяются во времени от кадра к кадру менее 25 % пикселей. В процессе преобразования цифрового потока обеспечивается передача лишь изменений пикселей, что обеспечивает сокращение цифрового потока еще не менее чем в 4 раза. Таким образом, формируемый цифровой поток составляет менее 2 Мбит/с.

В создаваемой в последние годы интегральной цифровой сети предполагается в системе видеоконференцсвязи использовать более эффективные методы сжатия визуальной информации. При этом скорость цифрового потока может быть уменьшена до 384 Кбит/с. В системах видеоконференцсвязи часто используются полиэкранные системы отображения информации.

### Видеотелефонные системы

В странах со стандартом телевидения на 625 строк используются видеотелефонные системы (видеотелефоны) с 313 строками в кадре при чересстрочной развертке и частоте кадров 25 Гц. Верхняя граничная частота видеосигнала равна 1,25 МГц. При этом возможен цифровой поток информации в 20 Мбит/с.

Предполагается использование каналов с возможностью передачи информации со скоростью 64 кбит/с или со скоростью, кратной этому значению. Сокращение избыточности изображений в этом случае в основном связано с их низкой подвижностью.

### Факсимильные системы (факсы)

Факсимильная связь является видом документальной связи, предназначенной для передачи не только содержания, но и внешнего вида самого документа. Сущность факсимильного метода передачи состоит в том, что передаваемое изображение (оригинал) разбивается на отдельные элементарные площадки, которые сканируются световым пятном требуемых размеров со скоростью развертки 60, 90, 120, 180 или 240 строк/мин. Пятно формируется светооптической системой, содержащей источник света и оптическое устройство.

Перемещение светового пятна по поверхности оригинала осуществляется развертывающим устройством (РУ). Сигнал яркости, пропорциональный коэффициенту отражения элементарных площадок, преобразуется в цифровой вид и

передается по каналу связи с использованием того либо иного способа модуляции. На приемной стороне эти сигналы преобразуются в элементы изображения и воспроизводятся (записываются) на приемном бланке. Таким образом, аппарат факсимильной связи очень напоминает ксерокс, в котором оригинал и копию разделяют многие километры.

В большинстве практических случаев передаваемые изображения являются статическими и двухуровневыми (черно-белые). Так как изображение обычно состоит только из текста и рисунков, то требуется более высокое пространственное разрешение по сравнению с стандартным телевидением. Обычно такие изображения можно представить в виде сочетания мозаичных или геометрических элементов, которые хранятся в памяти приемного устройства. Факсимильные системы обеспечивают передачу сигналов по телефонным аналоговым или цифровым каналам (общего пользования или специально выделенным). Цифровая передача обычно ведется со скоростями 1 200, 2 400, 4 800 или 9 600 бит/с. В новейших системах при помехоустойчивом кодировании скорости передач достигают 56 Кбит/с.

Современные факс-модемы имеют в своем составе все составные части факсимильных аппаратов за исключением сканирующего и воспроизводящего устройств. Они «умеют» связываться с обыкновенными факсами, при этом принимаемая информация о передаваемом изображении передается на компьютер, где специальной программой передачи факсимильных сообщений преобразуется в один из распространенных графических форматов. В дальнейшем полученный таким образом документ можно отредактировать, вывести на цветной принтер или передать другому корреспонденту, имеющему факс или компьютер с факс-модемом.

### Магнитная запись и воспроизведение звука

Магнитная запись звука обладает рядом преимуществ по сравнению с механической: ее можно сразу воспроизводить, звукозапись используется многократно, записывающая аппаратура — магнитофоны — просты в управлении и сравнительно дешевы.

Магнитный способ записи звука основан на свойстве ферромагнитных тел сильно намагничиваться в сравнительно слабых магнитных полях и сохранять состояние намагниченности после их исчезновения. В бытовых и профессиональных магнитофонах широкое применение получили магнитные ленты. Их основанием (подложкой) являются немагнитные эластичные ленты толщиной 18...55 мкм из синтетических материалов. Подложку покрывают ферромагнитным порошковым слоем толщиной 8...20 мкм.

Для улучшения качества звуковоспроизведения в магнитной записи используется цифровое представление сигналов. Уже много лет разработаны и эксплуатируются стереофонические цифровые магнитофоны.

## Контрольные вопросы

1. Какие основные задачи решают радиоэлектроника и техника связи?
2. Что определяют понятия «информация» и «сообщение»?
3. Какова взаимосвязь между информацией и сообщением?
4. Что такое «сигнал» и какими параметрами он характеризуется?
5. Зачем в радиотехнике передаваемое сообщение преобразуют в электрический сигнал?
6. Что такое радиоволны и на какие диапазоны они разделяются?
7. Что понимают под плотностью потока мощности электромагнитного излучения?
8. Какое строение имеет ионосфера и почему ее состав меняется в течение суток?
9. Какие физические процессы оказывают влияние на распространение радиоволн?
10. Какие возможны пути распространения радиоволн?
11. Что такое критическая частота и максимально применимая частота?
12. При каких условиях волны полностью отражаются от ионосферы?
13. Каковы особенности работы радиолиний на километровых и гектометровых волнах?
14. Почему образуются «замирания» и «зоны молчания» на декаметровых волнах?
15. Что такое диаграмма направленности? Какой вид имеют диаграммы направленности диполя и полуволнового вибратора?
16. Почему возможен переход от полуволнового вибратора к четвертьволновому?
17. Что собой представляет ФАР и на чем основан принцип ее действия?
18. Представьте структурную схему аналоговой системы радиосвязи.
19. Какие основные радиотехнические процессы характерны для систем связи?
20. В чем состоит суть метода уплотнения?
21. Чем отличаются аналоговые и цифровые системы связи?
22. В чем состоит сущность процесса модуляции?
23. Какие виды модуляции используются в системах связи?
24. Что представляет собой процесс детектирования?
25. Какими параметрами характеризуется радиоканал?
26. Опишите передачу информации по каналам связи и влияние помех.
27. Представьте структурную схему телевизионной системы.
28. Для чего в телевизионных системах используют устройство синхронизации?
29. Представьте структурную схему цифровой телевизионной системы.
30. По каким направлениям в настоящее время развивается телевизионное вещание?
31. На каком принципе основано действие радиолокационных систем?
32. Для каких целей применяют радионавигационные системы?
33. Из каких радиотехнических систем состоит система радиотелеуправления?
34. Укажите области применения радиовещания, телевидения, радиосвязи, радиолокации, радиоастрономии, радионавигации и радиотелеметрии.
35. На каком принципе построены современные системы подвижной связи?
36. Как работают сотовые системы мобильной связи?
37. Что собой представляют пейджинговые системы связи?
38. Для каких целей используются транкинговые системы мобильной радиосвязи?
39. Каков принцип построения современных систем беспроводных телефонов?
40. На чем основано действие низкоорбитальных систем спутниковой связи?
41. Какую структуру представляют системы мобильной связи будущего?
42. Что собой представляет компакт-диск?
43. Для чего используют системы видеоконференцсвязи и факсимильные системы?
44. Как осуществляется магнитная запись и воспроизведение звука?

# Глава 2. ОСНОВЫ ТЕОРИИ ДЕТЕРМИНИРОВАННЫХ СИГНАЛОВ

Изменение во времени напряжения, тока, заряда или мощности в электрических цепях называют *электрическим колебанием*. Используемое для передачи информации электрическое колебание является сигналом. В радиотехнике термины «электрическое колебание» и «сигнал» близки по смыслу и часто заменяют друг друга. Сложность процессов в электрических цепях зависит от сложности исходных сигналов. Поэтому целесообразно пользоваться спектром сигналов. Из математики известны ряды и преобразования Фурье, с помощью которых удается представить сигналы совокупностью гармонических составляющих. На практике полезен анализ характеристики, дающий представление о скорости изменения и длительности сигнала. Это удается достичь с помощью корреляционного анализа.

Сигналы, поступающие из источника сообщений (микрофон, передающая телевизионная камера и т. д.), как правило, не могут быть непосредственно переданы по радиоканалу. Это обусловлено их относительной *низкочастотностью*. Чтобы осуществить эффективную передачу сигналов по линиям связи, необходимо перенести их спектр из низкочастотной области в область высоких частот. Данная процедура получила название *модуляции*.

В различных системах передачи информации применяют узкополосные сигналы, спектр которых сосредоточен в окрестности частоты $\omega_0$. При анализе этих сигналов пользуются понятием «аналитический сигнал».

## 2.1. Общие сведения о радиотехнических сигналах

С точки зрения решения многих проблем передачи информации математические модели сигналов и помех являются фундаментом радиотехники. Особенно это относится к радиотехническим сигналам. Математический аппарат, используемый при описании сигналов, позволяет проводить анализ без учета их природы. Однако для понимания сути, характеристик и параметров электрических сигналов удобнее обращаться к их физическому наполнению.

При передаче информации на расстояние с помощью разных радиосистем (особенно систем связи) используют многообразные радиотехнические сигналы. Традиционно *радиотехническими* принято считать электрические (а теперь и оптические) сигналы, относящиеся к радиодиапазону. С математической точки зрения всякий радиотехнический сигнал можно представить некоторой функцией времени $u(t)$, которая характеризует изменение его мгновенных значений напряжения (такое представление применяют чаще всего), тока, заряда или мощ-

ности. Подобное описание радиотехнического сигнала наиболее распространено и удобно, но оно не исключает и другие виды его представления.

Каждый класс сигналов имеет свои особенности и требует специфических методов описания и анализа. Одним из ключевых компонентов представления и обработки сигналов является анализ. Основной целью анализа служит сравнение сигналов друг с другом для выявления их сходства и различия. Различают три основные составляющие анализа электрических сигналов:

- измерение числовых параметров сигналов; к таким параметрам, прежде всего, относят энергию, среднюю мощность и среднее квадратическое значение;
- разложение сигнала на элементарные составляющие либо для их рассмотрения по отдельности, либо для сравнения свойств различных сигналов; такое разложение проводят с использованием рядов и интегральных преобразований, важнейшими из которых являются ряды и преобразование Фурье;
- количественное измерение степени «похожести» различных сигналов, их параметров и характеристик; такое измерение производят с применением аппарата корреляционного анализа.

**Математические модели сигналов**. Изучать сигналы можно экспериментально с помощью различных приборов и устройств — осциллографов, анализаторов спектров, коррелометров, приемников, вольтметров и пр. Такой эмпирический подход имеет существенный недостаток. Явления, фиксируемые наблюдателем, всегда выступают как частные, единичные проявления, лишенные той степени обобщенности, которая позволила бы судить об их фундаментальных свойствах, предсказывать результаты в изменившихся условиях.

Для того чтобы сделать сигналы объектами изучения и расчетов, следует указать способ их математического описания, т. е. создать *математическую модель* исследуемого сигнала. Математической моделью сигнала называют его описание с помощью математического аппарата (функций, векторов, распределений и т. д.), позволяющее делать выводы об особенностях сигнала, применяя при этом формальные процедуры (например, математические преобразования). Математической моделью сигнала может быть, например, некая функциональная зависимость, аргументом которой является время. Создание модели физического сигнала — первый существенный шаг на пути систематического изучения его свойства. Прежде всего, математическая модель позволяет абстрагироваться от конкретной природы физического носителя сигнала.

В радиотехнике каждому классу сигналов соответствует свое математическое представление, своя математическая модель, причем одна и та же математическая модель может практически всегда адекватно описывать напряжение, ток, заряд, мощность, напряженность электромагнитного поля и т. д. Реальные радиотехнические сигналы весьма сложны по структуре и конкретны, а описывающие их математические модели — абстрактны. Поэтому между исследуемым сигналом-оригиналом и его математическими моделями не всегда удается получить полного совпадения во всех отношениях, когда любую модель можно рассматривать как оригинал или как обобщенную модель для остальных.

По существу любое аналитическое описание реального сигнала является его упрощением, которое получают путем сосредоточения внимания на наиболее важных обстоятельствах (зависимостях) и исключения остальных для данного исследования несущественных. Например, в спектральном представлении главное внимание сосредоточено на частотном составе сигнала, а временны́е зависимости его параметров из рассмотрения практически исключены.

Для математического моделирования сигналов можно использовать формулы, уравнения, функциональные зависимости, матрицы, графики, таблицы, графы, законы распределения вероятностей, векторы, многомерные пространства и другие аналитические образы. При моделировании сигналов нельзя ограничиваться статической фиксацией сведений о сигнале-оригинале. Для получения дальнейших выводов и обобщений необходимо уметь преобразовать имеющуюся информацию по определенным правилам. Эти правила, будучи заложенными в аналитическую модель, превращают ее из статической в *динамическую*. Очевидно, что выбор модели реального сигнала — процесс в значительной степени творческий. Наиболее распространенными способами представлений (описаний) сигналов являются временно́й, спектральный, аналитический, статистический, векторный, графический и геометрический. Функции, описывающие сигналы, могут принимать как вещественные, так и комплексные значения. Поэтому в дальнейшем в книге часто будем говорить о *вещественных* и *комплексных сигналах*. Использование того или другого принципа представления — дело математического «удобства». Зная математические модели сигналов, можно сравнивать эти сигналы между собой, устанавливать их тождество и различие, проводить классификацию сигналов по ряду специфических признаков.

### *Классификация радиотехнических сигналов*

Прежде чем приступить к изучению каких-либо новых явлений, процессов или объектов, в науке всегда стремятся провести их классификацию по возможно бо́льшим признакам. Для рассмотрения и анализа сигналов выделим их основные классы. Это необходимо по двум причинам. Во-первых, проверка принадлежности сигнала к конкретному классу — процедура анализа. Во-вторых, для представления и анализа сигналов разных классов зачастую приходится использовать разные средства и подходы. Основные понятия, термины и определения в области радиотехнических сигналов устанавливает национальный (ранее, государственный) стандарт «Сигналы радиотехнические. Термины и определения». Радиотехнические сигналы чрезвычайно разнообразны. Часть краткой классификации сигналов по ряду признаков приведена на рис. 2.1. Более подробно сведения о ряде понятий изложены далее.

Радиотехнические сигналы удобно рассматривать в виде математических функций, заданных во времени и физических координатах. С этой точки зрения сигналы обычно описывается одной (одномерный сигнал; $n = 1$), двумя (двумерный сигнал; $n = 2$) или более (многомерный сигнал $n > 2$) независимыми переменными. Одномерные сигналы являются функциями только времени, а многомерные, кроме того, отражают положение в $n$-мерном пространстве.




**Рис.2.1**. Классификация радиотехнических сигналов

Будем для определенности и упрощения в основном рассматривать одномерные сигналы, зависящие от времени, однако материал книги допускает обобщение и на многомерный случай, когда сигнал представляется в виде конечной или бесконечной совокупности точек, например в пространстве, положение которых зависит от времени. В телевизионных системах сигнал черно-белого изображения можно рассматривать как функцию $f(x, y, t)$ двух пространственных координат и времени, представляющую интенсивность излучения в точке $(x, y)$ в момент времени $t$ на катоде. При передаче цветного телевизионного сигнала имеем три функции $f(x, y, t)$, $g(x, y, t)$, $h(x, y, t)$, определенные на трехмерном множестве (можно рассматривать эти три функции также как компоненты трехмерного векторного поля). Кроме того, различные виды телевизионных сигналов могут возникать при передаче телевизионного изображения совместно со звуком.

Многомерный сигнал — упорядоченная совокупность одномерных сигналов. Многомерный сигнал создает, например, система напряжений на зажимах многополюсника (рис. 2.2). Многомерные сигналы описывают сложными функциями, и их обработка чаще возможна в цифровой форме. Поэтому многомерные модели сигналов особенно полезны в случаях, когда функционирование сложных систем анализируется с помощью компьютеров. Итак, *многомерные*, или *векторные*, сигналы состоят из множества одномерных сигналов

$$\mathbf{U}(t) = \{u_1(t), u_2(t), ..., u_n(t)\},$$

где $n$ — целое число, размерность сигнала.

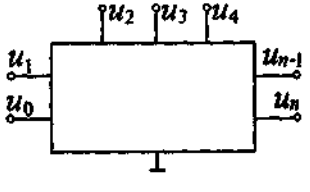



**Рис. 2.2.** Система напряжений многополюсника

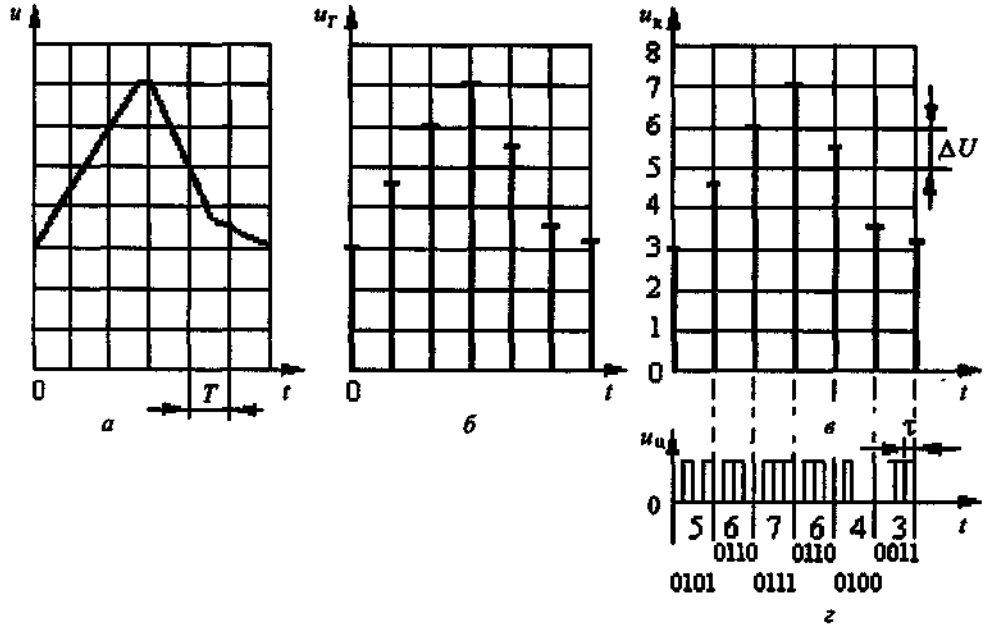



**Рис. 2.3.** Радиотехнические сигналы:
*а* — аналоговый; *б* — дискретный; *в* — квантованный; *г* — цифровой

По особенностям структуры временно́го представления (рис. 2.3) все радиотехнические сигналы делятся на аналоговые (*analog*), дискретные (*discrete-time*; от лат. *discretus* — разделенный, прерывистый) и цифровые (*digital*).

Если физический процесс, порождающий одномерный сигнал, можно представить непрерывной функцией времени $u(t)$ (рис. 2.3, *а*), то такой сигнал называют *аналоговым* (*непрерывным*), или, более обобщенно, *континуальным* (*continuos* — многоступенчатым), если последний имеет скачки, разрывы по оси амплитуд. Заметим, что традиционно термин «аналоговый» используют для описания сигналов, которые непрерывны во времени. *Непрерывный сигнал* можно трактовать как действительное или комплексное колебание во времени $u(t)$, являющейся функцией непрерывной действительной временно́й переменной $t$. Понятие «аналоговый» сигнал связано с тем, что его любое мгновенное значение *аналогично* закону изменения соответствующей физической величины во времени. Примером аналогового сигнала является некоторое напряжение, которое подано на вход осциллографа, в результате чего на экране возникает непрерывная кривая как функция времени. Поскольку современная обработка непрерывных сигналов с использованием резисторов, конденсаторов, операционных усилителей и т. п. имеет мало общего с аналоговыми компьютерами, термин «аналоговый» сегодня представляется не совсем неудачным. Более корректным было бы называть *непрерывной обработкой сигналов* то, что сегодня обычно называют аналоговой обработкой сигналов. Соответственно в учебнике сведено к минимуму использование термина «аналоговые сигналы» и он заменен термином «непрерывные сигналы» везде, где это оправдано.

В радиоэлектронике и технике связи широко применяются импульсные системы, устройства и цепи, действие которых основано на использовании дискретных сигналов. Например, электрический сигнал, отражающий речь, является непрерывным как по уровню, так и по времени, а датчик температуры, выдающий ее значения через каждые 10 мин, служит источником сигналов, непрерывных по значению, но дискретных по времени.

Дискретный сигнал получают из аналогового путем специального преобразования. Процесс преобразования аналогового сигнала в последовательность отсчетов называется *дискретизацией* (*sampling*), а результат такого преобразования — *дискретным сигналом* или *дискретным рядом* (*discrete series*). Простейшая математическая модель *дискретного сигнала* $u_T(t)$ — последовательность точек на временно́й оси, взятых, как правило, через равные промежутки времени $T = \Delta t$, называемые *периодом дискретизации* (или *интервалом, шагом дискретизации*; *sample time*), и в каждой из которых заданы значения соответствующего непрерывного сигнала (рис. 2.3, *б*). Величина, обратная периоду дискретизации, называется *частотой дискретизации* (*sampling frequency*): $f_д = 1/T$ (другое обозначение $f_д = 1/\Delta t$). Соответствующая ей угловая (круговая) частота определяется следующим образом: $\omega_д = 2\pi/\Delta t$.

Дискретные сигналы могут быть созданы непосредственно источником информации (в частности, дискретные отсчеты сигналов датчиков в системах управления). Простейшим примером дискретных сигналов могут служить сведения о температуре, передаваемые в программах новостей радио и телевидения, в паузах же между таким передачами сведений о погоде обычно нет. Не следует думать, что дискретные сообщения обязательно преобразуют в дискретные сигналы, а непрерывные сообщения — в непрерывные сигналы. Чаще всего именно непрерывные сигналы используют для передачи дискретных сообщений (в качестве их переносчиков, т. е. несущей). Дискретные же сигналы можно использовать для передачи непрерывных сообщений.

Очевидно, что в общем случае представление непрерывного сигнала набором дискретных отсчетов приводит к определенной потере полезной информации, так как мы ничего не знаем о поведении сигнала в промежутках между отсчетами. Однако, как будет показано в главе 8 п.8.2 «Цифровое представление сигналов. Теорема Котельникова», существует класс аналоговых сигналов, для которых такой потери информации практически не происходит, и поэтому они могут быть *с высокой степенью точности восстановлены* по значениям своих дискретных отсчетов.

Разновидностью дискретных сигналов является *цифровой сигнал* (*digital signal*). В процессе преобразования дискретных отсчетов сигнала в цифровую форму (обычно в двоичные числа) производится его *квантование по уровню* (*quantization*) напряжения $\Delta$. При этом значения уровней сигнала можно пронумеровать двоичными числами с конечным, требуемым числом разрядов. Сигнал, дискретный во времени и квантованный по уровню, называют *цифровым сигналом*. Кстати, сигналы, квантованные по уровню, но непрерывные во времени,

на практике встречаются редко. В цифровом сигнале дискретные значения сигнала $u_Д(t)$ вначале квантуют по уровню (рис. 2.3, *в*) и затем квантованные отсчеты дискретного сигнала заменяют числами $u_Ц(t)$, чаще всего реализованными в двоичном коде, который представляют *высоким* (*единица*) и *низким* (*нуль*) уровнями потенциалов напряжения — короткими импульсами длительностью τ (рис. 2.3, *г*). Такой код называют *униполярным*.

Поскольку отсчеты могут принимать конечное множество значений уровней напряжения (см. например второй отсчет на рис. 2.3, *г*, который в цифровом виде практически равновероятно может быть записан как числом 5 — 0101, так и числом 4 — 0100), то при представлении сигнала неизбежно происходит его округление. Возникающие при этом ошибки округления называются *ошибками* (или *шумами*) *квантования* (*quantization error*, *quantization noise*). Последовательность чисел, представляющая сигнал при цифровой обработке, является *дискретным рядом* (*discrete series*). Числа, составляющие последовательность, являются значениями сигнала в отдельные (дискретные) моменты времени и называются *цифровыми отсчетами сигнала* (*samples*). Далее квантованное значение сигнала представляется в виде набора импульсов, характеризующих нули («0») и единицы («1») при представлении этого значения в двоичной системе счисления (рис. 2.3, *г*). Набор импульсов используют для амплитудной модуляции несущего колебания и получения кодово-импульсного радиосигнала.

В результате цифровой обработки не получается ничего «физического», только цифры. А цифры — это абстракция, способ описания информации, содержащейся в сообщении. Следовательно, нам необходимо иметь что-то физическое, что будет представлять цифры или «являться носителем» цифр. Итак, сущность цифровой обработки состоит в том, что *физический* сигнал (напряжение, ток и т. д.) преобразуется в последовательность *чисел*, которая затем подвергается *математическим* преобразованиям в вычислительном устройстве. Трансформированный цифровой сигнал (последовательность чисел) при необходимости может быть преобразован обратно, в напряжение или ток.

Цифровая обработка сигналов предоставляет широкие возможности по передаче, приему и преобразованию информации, в том числе и те, которые не могут быть реализованы с помощью аналоговой техники. Заметим, что в главе 8 будут в основном рассмотрены принципы математического описания, анализа и методов обработки дискретных сигналов. Здесь прежде всего следует пояснить некоторые терминологические тонкости, которые, возможно, уже были замечены внимательным читателем: речь идет как о *дискретной*, так и о *цифровой* обработке сигналов, а фактически рассматриваются *дискретные* сигналы и их обработка. Это связано с тем, что на практике при анализе и обработке сигналов чаще всего цифровые сигналы заменяют дискретными, а их отличие от цифровых интерпретируют как шум квантования. В связи с этим эффекты, связанные с квантованием по уровню и оцифровкой сигналов, в большинстве случаев не будут приниматься во внимание. Можно сказать, что и в дискретных и цифровых цепях (в частности, в цифровых фильтрах) обрабатывают дискретные сигналы, только внутри структуры цифровых цепей эти сигналы представлены числами.

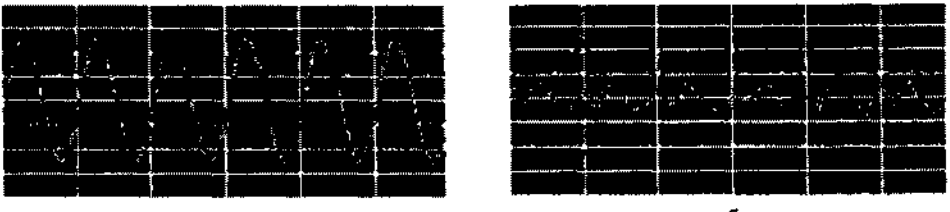
Поэтому здесь и далее термины «дискретные сигналы» и «цифровые сигналы», а также «дискретные фильтры» и «цифровые фильтры» будут использоваться как синонимы. Вместе с тем ряду эффектов, а также некоторым проблемам, связанным с конечной точностью представления дискретных сигналов числами в вычислительных устройствах, внимание будет уделено в главе 8. Вычислительные устройства, предназначенные для обработки сигналов, могут оперировать с цифровыми сигналами. Существуют также устройства, построенные в основном на базе аналоговой схемотехники, которые работают с дискретными сигналами, представленными в виде импульсов различной амплитуды, длительности или частоты повторения.

Одним из основных признаков, по которым различаются сигналы, является предсказуемость сигнала (его значений) во времени.

*По математическому представлению* (по степени наличия *априорной*, от лат. *a priori* — из предшествующего, т. е. *доопытной информации*) все радиотехнические сигналы принято делить на две основные группы: детерминированные (регулярные; *determined*) и случайные (*casual*) сигналы (рис. 2.4).

*Детерминированными* называют радиотехнические сигналы, мгновенные значения которых в любой момент времени достоверно известны, т. е. предсказуемы с вероятностью, равной единице. Детерминированные сигналы описываются заранее заданными функциями времени. Кстати, мгновенное значение сигнала — это мера того, на какое значение и в каком направлении переменная отклоняется от нуля; таким образом, мгновенные значения сигнала могут быть как положительными, так и отрицательными (рис. 2.4, *а*). Простейшими примерами детерминированного сигнала являются гармоническое колебание с известной начальной фазой, высокочастотные колебания, модулированные по известному закону, последовательность или пачка импульсов, форма, амплитуда и временно́е положение которых заранее известны.

Если бы передаваемое по каналам связи сообщение было детерминированным, т. е. заранее известным с полной достоверностью, то его передача была бы бессмысленной. Такое детерминированное сообщение по сути дела не содержит никакой новой информации. Поэтому сообщения следует рассматривать как случайные события (или случайные функции, случайные величины). Иначе говоря, должно существовать некоторое множество вариантов сообщения (например, множество различных значений давления, выдаваемых датчиком), из



*а* *б*

**Рис. 2.4.** Сигналы:
*а* — детерминированный; *б* — случайный

которых реализуют с определенной вероятностью одно. В связи с этим и сигнал является случайной функцией. Детерминированный сигнал не может быть носителем информации. Его можно использовать лишь для испытаний радиотехнической системы передачи информации или тестирования отдельных ее устройств. Случайный характер сообщений, а также помех обусловил важнейшее значение теории вероятностей в построении теории передачи информации.

Детерминированные сигналы разделяют на *периодические* и *непериодические* (*импульсные*). Сигнал конечной энергии, существенно отличный от нуля в течение ограниченного интервала времени, соизмеримого со временем завершения переходного процесса в системе, для воздействия на которую он предназначен, называют *импульсным* сигналом.

*Случайными* называют сигналы, мгновенные значения которых в любой момент времени не известны и не могут быть предсказаны с вероятностью, равной единице. Фактически для случайных сигналов можно знать только вероятность того, что он примет какое-либо значение.

Может показаться, что понятие «случайный сигнал» не совсем корректно. Но это не так. Например, напряжение на выходе приемника тепловизора, направленного на источник ИК-излучения, представляет хаотические колебания, несущие разнообразную информацию об анализируемом объекте. Строго говоря, все сигналы, встречающиеся на практике, являются случайными и большинство из них представляют хаотические функции времени (рис. 2.4, *б*). Как ни парадоксально на первый взгляд, но сигналом, несущим полезную информацию, может быть только случайный сигнал. Информация в таком сигнале заложена во множестве амплитудных, частотных (фазовых) или кодовых изменений передаваемого сигнала. Сигналы связи во времени меняют мгновенные значения, причем эти изменения могут быть предсказаны лишь с некоторой вероятностью, меньшей единицы. Таким образом, сигналы связи являются в некотором роде случайными процессами, поэтому и их описание осуществляется посредством методов, аналогичных методам описания случайных процессов.

В процессе передачи полезной информации радиотехнические сигналы могут быть подвергнуты тому или иному преобразованию. Это обычно отражают в их названии: сигналы *модулированные*, *демодулированные* (детектированные), *кодированные* (декодированные), *усиленные*, *задержанные*, *дискретизированные*, *квантованные* и др.

*По назначению*, которое сигналы имеют в процессе модуляции, их можно разделить на *модулирующие* (первичный сигнал, который модулирует несущее колебание) или *модулируемые* (несущее колебание).

*По принадлежности* к тому или иному виду радиотехнических систем, и в частности систем передачи информации, различают «*связные*», *телефонные*, *телеграфные*, *радиовещательные*, *телевизионные*, *радиолокационные*, *радионавигационные*, *измерительные*, *управляющие*, служебные (в том числе *пилот-сигналы*) и другие сигналы.

Приведенная краткая классификация радиотехнических сигналов не полностью охватывает все их разнообразие.

### Помехи и шумы в радиотехнических системах

Радиотехнические сигналы редко присутствуют в электрических цепях в чистом виде. Практически всегда на них накладываются шумы и помехи. При этом полезный сигнал искажается при передаче и сообщение воспроизводится с некоторой ошибкой. Причиной ошибок являются как искажения, вносимые самим каналом, так и различного вида помехи, воздействующие на сигнал в процессе передачи. Наличие шума, наложенного на сигнал, «затеняет», или маскирует, его; это ограничивает способность приемника принимать точные решения о значении символов, а следовательно, ограничивает скорость передачи информации. В собственно радиоэлектронных устройствах канала передачи информации имеются два основных источника шумов: дискретная структура тока в усилительных элементах (транзисторы, микросхемы и т. д.) и тепловое движение свободных электронов в проводниках электрической цепи. При этом временные и частотные характеристики канала определяют так называемые *линейные искажения*. Кроме того, радиоканал может вносить и нелинейные искажения, обусловленные нелинейностью тех или иных его звеньев, цепей или устройств. Следует отличать искажения от помех, имеющих случайный характер. Если линейные и нелинейные искажения обусловлены известными характеристиками канала, то в принципе они могут быть исключены их надлежащей коррекцией. Помехи же заранее, как правило, не известны, и поэтому практически не могут быть полностью устранены. Борьба с помехами (шумами) является одной из главных проблем радиотехники.

В общем случае под *радиотехнической помехой* понимают случайный сигнал, однородный с полезным и действующий одновременно с ним. Для систем передачи информации помеха — это любое случайное воздействие на полезный сигнал, ухудшающее верность приема и воспроизведения передаваемых по линии связи сообщений.

Радиотехнические помехи классифицируют по ряду признаков, поскольку они разнообразны и по происхождению, и по физическим свойствам.

*По месту возникновения* помехи делят на *внешние* и *внутренние*. Причиной возникновения *внешних* помех являются природные процессы и работа различных технических устройств. В диапазонах дециметровых и менее волн имеют значение и *космические помехи*, связанные с электромагнитными процессами, происходящими на Солнце, звездах и других внеземных объектах. В диапазоне оптических частот существенное значение имеет *квантовый шум*, вызванный дискретной природой сигнала.

В радиоканалах встречаются *атмосферные помехи*, обусловленные электрическими процессами в атмосфере, прежде всего грозовыми разрядами. Их энергия сосредоточена в области километровых и гектометровых волн.

Сильные помехи создают промышленные установки. Это так называемые *индустриальные помехи*, возникающие из-за резких изменений тока в мощных электрических цепях всевозможных электротехнических устройств. Сюда относят помехи от электротранспорта, электрических двигателей, медицинских ус-

тановок, систем зажигания двигателей внутреннего сгорания и т. п. Распространенным видом внешних помех являются помехи от посторонних радио- и телестанций, систем военного назначения. Они обусловлены нарушением регламента распределения частот, недостаточной стабильностью частот генераторов и плохой фильтрацией гармоник сигнала, а также нелинейными процессами в каналах, ведущими к так называемым *перекрестным искажениям* (проявляются в переносе модуляции с мешающего внеполосного сигнала на полезный).

Основным видом внешних помех в проводных каналах связи являются *импульсные шумы* и *прерывание связи*. Появление импульсных помех часто связано с автоматической коммутацией каналов и перекрестными наводками. Прерывание связи — явление, при котором сигнал в линии резко затухает или исчезает.

Следует отметить *умышленные активные* и *пассивные помехи, ложные цели и ловушки*. Эти помехи, как правило, устанавливают в системах военного и специального назначения для целей маскировки, радиопротиводействия, «глушения» различных средств радиовещания и повышения секретности.

*Внутренние* помехи обусловлены процессами, происходящими при работе самого радиотехнического устройства. Практически в любом диапазоне частот имеют место внутренние шумы радиотехнических устройств, связанные с хаотическим движением носителей заряда в усилительных приборах, резисторах и других элементах. Эти помехи особенно сказываются при передаче информации в диапазонах дециметровых и меньших длин волн, где другие помехи невелики.

Аналитически влияние помехи $r(t)$ на полезный сигнал $u(t)$ в общем виде можно выразить оператором $Y$

$$z(t) = Y\{s[u(t)], r(t)\}, \tag{2.1}$$

где функция $s[u(t)]$ отражает искаженный полезный сигнал.

Возможны два сочетания полезного сигнала и шума. Если оператор $Y$ в (2.1) вырождается в линейную сумму сигнальной составляющей и помехи

$$z(t) = s(t) + r(t), \tag{2.2}$$

то помеху называют *аддитивной* (от англ. *addition* — сложение; термин «аддитивный» означает, что шум просто суммируется с сигналом).

Если же оператор $Y$ может быть представлен в виде произведения некоторого коэффициента $k(t)$ (здесь $k(t)$ — случайный процесс) и сигнала

$$z(t) = k(t)u(t), \tag{2.3}$$

то помеху называют *мультипликативной* (от англ. *multiplication* — умножение).

Мультипликативные помехи обусловлены случайными изменениями параметров радиоканала. Они проявляются в изменении уровня сигнала. Простейшие случаи — телефонная линия связи с плохими электрическими контактами, налагающиеся на сигнал грозовые разряды, изменяющийся коэффициент усиления приемника и т. д. В реальных каналах передачи информации обычно имеют место и аддитивные, и мультипликативные помехи, и поэтому

$$z(t) = k(t)u(t) + r(t). \tag{2.4}$$

*По основным свойствам* аддитивные помехи делят на три класса: сосредоточенные по спектру (узкополосные помехи), импульсные помехи (сосредоточенные во времени) и флуктуационные (*fluctuation*; распределенные по частоте и времени) помехи, не ограниченные ни во времени, ни по спектру.

*Сосредоточенными по спектру* называют помехи, основная часть мощности которых приходится на отдельные участки диапазона частот, меньше полосы пропускания радиотехнической системы. Помехи, наводимые в радиотехнических цепях от промышленной силовой сети частотой 50 Гц, являются сосредоточенными. Эффективность их подавления в значительной мере определяется достоверностью априорных данных о частотном спектре передаваемого сигнала.

*Импульсной* (сосредоточенной во времени) *помехой* называют регулярную или хаотическую последовательность импульсных сигналов, однородных с полезным сигналом. Импульсные и сосредоточенные помехи часто называют *наводками*. Источниками таких помех являются цифровые и коммутирующие элементы радиотехнических цепей или работающего рядом с ними устройства. К импульсным помехам относят многие виды атмосферных и индустриальных помех. В зависимости от частоты следования импульсов одна и та же помеха может воздействовать как импульсная на приемник с широкой полосой пропускания и как флуктуационная на приемник с относительной узкой полосой пропускания. На практике импульсные помехи рассматривают как случайный, относительно широкополосный (тем шире, чем короче импульсы помехи) процесс, состоящий из отдельных редких, случайно распределенных во времени и по амплитуде импульсов. Для устранения влияния помех целесообразно исключить причины их возникновения. Способы борьбы с помехами в значительной мере зависят от их спектрального состава, вида передаваемого сигнала и помехи.

*Флуктуационная помеха* (*флуктуационный шум*) представляет случайный процесс с нормальным распределением — гауссовский процесс (*normal distribution, Gaussian distribution*; закон Гаусса; закон Муавра–Лапласа; Карл Гаусс — Carl Gauss; 1777–1855; немецкий математик и астроном). Эти помехи имеют место практически во всех реальных каналах связи, и их называют *шумами*.

Флуктуационный шум представляют наибольший интерес как в теоретическом, так и в практическом отношении. С физической точки зрения аддитивные флуктуационные помехи порождаются в системах связи различного рода флуктуациями, т. е. случайными отклонениями тех или иных физических величин (параметров) от их средних значений. Среди таких шумов можно, прежде всего, назвать внутренние шумы электронных усилителей.



К. Гаусс

Различают следующие виды флуктуационных шумов:

- *тепловой* (*thermal noise*; шум Джонсона);
- *фликкер-шум* (*flicker noise*; иногда, розовый шум);
- *дробовый* (*schrot noise*; квантовый).

**Тепловые шумы резисторов.** Одной из главных причин возникновения шума являются флуктуации объемной плотности электрического заряда в резисторах (и резистив-

ных элементах) из-за хаотического теплового движения носителей. В любом резисторе всегда имеются свободные электроны, находящиеся в хаотическом тепловом движении. При этом может оказаться, что в определенный момент времени в одном направлении проходит больше электронов, чем в другом. Значит, даже в отсутствие внешней ЭДС мгновенное значение тока, текущего через резистор, отлично от нуля. Эти мгновенные изменения тока вызывают на выводах резистора шумовую разность потенциалов. Среднее значение такого напряжения равно нулю, а переменная составляющая проявляется как шум.

Важное значение для радиотехники и систем связи имеет спектром мощности шумового напряжения на концах резистора. Его определяют по формуле Найквиста, известной из физики:

$$W_0 = 2kTR,$$

где $R$ — сопротивление резистора, Ом; $k = 1{,}38 \cdot 10^{-23}$ Дж/К — постоянная Больцмана; $T$ — абсолютная температура резистора в градусах Кельвина.

Часто удобнее пользоваться односторонним энергетическим спектром, который задают в области положительных частот [$В^2$/Гц]

$$N_0 = 2W_0 = 4kTR. \qquad (2.5)$$

Спектральную плотность мощности теплового шума можно оценить из следующего примера: при $T = 300$ К и $R = 20$ кОм значение

$$N_0 = 4 \cdot 1{,}38 \cdot 10^{-23} \cdot 300 \cdot 20000 = 3{,}31 \cdot 10^{-16} \text{ В}^2/\text{Гц},$$

откуда его среднее квадратическое значение напряжение $U_ш = 3{,}31 \cdot 10^{-16}$ В/Гц $^{-2}$.

Несмотря на малый уровень, тепловой шум может явиться решающим фактором, ограничивающим реальную чувствительность приемных устройств.

Спектральная плотность мощности теплового шума *одинакова* для всех частот, представляющих интерес для большинства систем связи; другими словами, источник теплового шума на всех частотах излучает с равной мощностью на единицу ширины полосы — от постоянной составляющей до частоты порядка $10^{12}$ Гц. Следовательно, простая модель теплового шума предполагает, что его спектральная плотность мощности равномерна и достаточно точно соответствует модели белого шума. Тепловой шум практически устанавливает нижнюю границу напряжения шумов радиотехнического устройства. Одной из лучших визуализаций теплового шума служит «рябь» на экране не настроенного телевизора. В заключение укажем, что приведенные соотношения можно использовать также при анализе теплового шума в избирательных цепях.

Специфическим для электронных приборов является фликер-шум, который возникает в результате разного рода поверхностных явлений.

**Фликкер-шум** — шум, спектральная плотность которого изменяется с частотой по закону $1/f$ (с примерно постоянной спектральной мощностью на декаду — изменение в десять раз). В электровакуумных и газоразрядных приборах шумы такого вида вызываются *испарением атомов*. Часто фликер-шумом называют любой шум, спектральная плотность которого уменьшается с увеличением час-

тоты. Обычно на частотах выше 10 кГц фликкер-шумами пренебрегают. Наиболее яркий пример фликкер-шума — шум пролетающего вертолета. В телевидении и видеотехнике фликкер-шум (мерцание) — раздражающая помеха, главным образом связанная с кадровой синхронизацией и отображением полей.

**Дробовой (квантовый) шум** обусловлен неравномерным движением дискретных носителей электрического тока в электронных приборах — диодах, транзисторах, микросхемах и лампах; он имеет равномерный спектр, т. е. является белым; в отличие от резисторов флуктуации возникают не за счет хаотического теплового движения электронов, а вследствие статистической независимости их упорядоченного перемещения.

Поскольку тепловой шум присутствует во всех системах передачи информации и связи и для их большинства является заметным источником помех, характеристики теплового шума (аддитивный, белый и гауссов) часто применяются для моделирования шума в системах связи. Гауссов шум с нулевым средним полностью характеризуется *дисперсией*, поэтому эту модель особенно просто использовать и при детектировании сигналов и при проектировании оптимальных приемников. Будем полагать (если не оговорено другое), что полезный сигнал подвергается искажению со стороны *аддитивного белого гауссового шума* (*additive white Gaussian noise* — AWGN) *с нулевым средним*, несмотря на то, что иногда такое упрощение чересчур сильно.

*По виду частотного спектра* помехи делят на белый и нестационарный шумы. *Белый шум* (см. гл. 3) содержит гармонические составляющие с одинаковой амплитудой и случайной начальной фазой, которые равномерно распределены практически по всему частотному радиодиапазону — от постоянной составляющей до частоты порядка $10^{12}$ Гц. В теории оптимальной фильтрации часто вводят понятие «квазибелого шума» (от лат. *quasi* — якобы; также — почти, похожий), параметры и характеристики которого близки к показателям белого шума. *Нестационарный шум* имеет неравномерный спектр.

В зависимости от спектра помехи могут быть сплошными или дискретными (селективными). Сигнал *сплошной* помехи характеризуется распределением его мощности по широкому спектру частот — это уже упоминавшийся белый шум. *Селективная* помеха характеризуется тем, что ее мощность сосредоточена либо на одной частоте, либо в очень узкой полосе частот.

Фактически между сигналом и помехой нет принципиального различия. Более того, они существуют в единстве, хотя и противоположны по действию. Так излучение антенны передатчика является полезным сигналом для антенны приемника, которому предназначено это излучение, и помехой для всех антенн других приемников. Электромагнитное излучение звезд — одна из причин космического шума в СВЧ-диапазоне, и поэтому это — помеха для радиосистем. С другой стороны, это излучение является весьма полезным сигналом, по которому определяют некоторые физико-химические свойства звезд.

Хорошее техническое проектирование может устранить большинство шумов путем экранирования, фильтрации, выбора модуляции и оптимального местоположения приемника. Например, радиоастрономические исследования про-

водят в пустынных местах, вдали от естественных источников шума. Тепловые шумы заметно уменьшаются при охлаждении их источника. Однако в целом борьба с помехами чрезвычайно сложна и является и искусством, и наукой.

С математической точки зрения информационные случайные сигналы (сигналы случайного характера, несущие передаваемую информацию) и шумы подчиняются одним вероятностным законам, поэтому они получили обобщенное название *случайные колебания* или *случайные процессы*.

Для анализа случайных сигналов применяют методы статистической радиотехники, базирующейся на математическом аппарате теории вероятности и теории случайных процессов. Однако использование теории вероятности для анализа случайных процессов затруднено ввиду сложности исследования и практических расчетов. Поэтому с целью упрощения и наглядности анализа работу радиотехнических цепей часто рассматривают при воздействии детерминированных сигналов. Для учета же случайного характера реального сигнала в качестве его математической модели используют не отдельную детерминированную функцию $u(t)$, а совокупность подобных функций $\{u_k(t)\} = u_1(t), u_2(t), ...$, образующих случайный процесс, в котором будет заключена полезная информация.

Случайный характер сигналов и помех обусловил важнейшее значение теории вероятностей в построении радиосистем. Вероятностные свойства сигналов и сообщений, а также среды, в которой передается сигнал, позволяют определить количество передаваемой информации и ее потери в процессе передачи.

### *Способы аналитического описания сигналов*

Функции, описывающие сигналы, могут быть представлены как вещественными, так и комплексными значениями. Поэтому в радиотехнике говорят о *вещественных* и, условно, о *комплексных сигналах*. Применение той или иной формы описания сигнала — дело математического удобства.

Для анализа прохождения сигнала через радиотехнические цепи необходимо его представить в удобной математической форме. В радиотехнике широкое применение нашли два способа математического представления сигналов: спектральный и временно́й. Такое представление возможно вследствие принципа *дуальности* (от англ. *duality* — двойственности, симметрии; например, сопротивления $R$ и проводимости $Y$; $R = 1/Y$) частоты и времени, поскольку $f = 1/t$.

При временно́м способе анализа сигнал отражают непрерывной функцией времени или совокупностью элементарных импульсов, следующих друг за другом через определенные интервалы времени. Временна́я форма представления сигнала позволяет легко определить такие важные его характеристики, как энергия, мощность и длительность. Спектральный способ позволяет представлять сигналы либо в виде взвешенной суммы гармонических составляющих, либо в виде суммы комплексных экспонент с частотами, обычно кратными друг другу (образующими арифметическую прогрессию).

Часто анализ радиотехнических сигналов строится на их геометрической или векторной интерпретации. Причем известно, что спектральное и векторное представления сигналов взаимосвязаны.

### Интегральное преобразование сигналов

При ряде условий для функции $u(t)$, описывающей сигнал во временно́й области (области определения) $0 - T$, существует интегральное преобразование

$$S(j\omega) = \dot{S}(\omega) = S(\omega) = \int_0^T \psi(\omega, t)u(t)dt, \tag{2.6}$$

где $\psi(\omega, t)$ — заданная функция (ядро интегрального преобразования).

Интегральное преобразование позволяет осуществить переход от временно́й области определения функции к области частоты $\omega$. Операциям над сигналом во временно́й области соответствуют операции в области частоты, которые зачастую проще. Значит расширяются возможности проводимого анализа.

Формула, позволяющая восстановить сигнал $u(t)$ по известной комплексной функции $S(\omega)$, называют формулой обращения интегрального преобразования

$$u(t) = \int_0^T \phi(t, \omega)S(\omega)d\omega, \tag{2.7}$$

где $\phi(t, \omega)$ — базисная (*basis*) функция.

Выражения (2.6) и (2.7) устанавливают взаимно однозначное соответствие между сигналом $u(t)$ и его интегральным преобразованием — спектром $S(\omega)$. Функция $S(\omega)$ является интегральным представлением сигнала и может быть исходной при его описании. Из интегральных преобразований, наиболее используемым при анализе сигналов, является преобразование Фурье.

### Комплексная форма представления сигналов

Часто при описании и анализе некоторых видов сигналов (в первую очередь узкополосных) бывает удобной комплексная форма их представления

$$\dot{u}(t) = |\dot{u}(t)|e^{j\varphi(t)}, \tag{2.8}$$

где $|\dot{u}(t)|$, $\varphi(t)$ — соответственно модуль и фаза комплексной величины $\dot{u}(t)$.

Комплексная функция $\dot{u}(t)$ может быть также представлена в виде

$$\dot{u}(t) = \mathrm{Re}[\dot{u}(t)] + j\,\mathrm{Im}[\dot{u}(t)], \tag{2.9}$$

где Re, Im — действительная и мнимая части комплексной функции.

Из формул (2.8) и (2.9) получим:

$$|\dot{u}(t)| = \sqrt{\mathrm{Re}^2[\dot{u}(t)] + \mathrm{Im}^2[\dot{u}(t)]}; \tag{2.10}$$

$$\varphi(t) = \mathrm{arctg}\,\frac{\mathrm{Im}[\dot{u}(t)]}{\mathrm{Re}[\dot{u}(t)]}. \tag{2.11}$$

По формальному аналитическому представлению сигнал может быть действительным или комплексным, т. е. состоящим из вещественной и мнимой частей. Комплексные числа появляются не как дань математической общности (хотя возможность вместо двух выражений $e^a \cos b$ и $e^a \sin b$ записать одно $e^{(a+jb)}$ сильно облегчает выкладки), и тем более не стоит видеть здесь математическую мистику («корень квадратный из минус единицы»). С их помощью удобно записывать синфазную (совпадающую по фазе с некоторым сигналом) и квадратурную (отличающуюся по фазе от этого сигнала на 90°) составляющие сигнала.

### Векторное представление сигналов

Комплексную форму сигналов удобно отражать точками на плоскости — одна координата отражает действительную, вторая — мнимую часть. Тогда сложение сигналов станет сложением соответствующих сигналам векторов, а умножение — поворотом векторов на плоскости (с умножением их длин, равным модулям этих чисел; углы же, равные аргументам чисел, складывают). Последовательное возведение комплексного числа в степень становится вращением выражающего это число вектора вокруг начала координат. Проекция данного вектора на одну из осей координат будет представлять нарастающие, затухающие или же с постоянной амплитудой колебания — в зависимости от того, больше ли единицы модуль данного комплексного числа, меньше или равен ей.

Итак, при векторном представлении комплексный сигнал — это вектор на комплексной плоскости с действительной осью — осью абсцисс и мнимой осью — осью ординат (рис. 2.5). Вектор на плоскости вращается в положительном направлении (против часовой стрелки) со скоростью $\omega_0$. Длина вектора равна модулю комплексного сигнала, угол между вектором и осью абсцисс — аргументу $\varphi(t)$. Проекции вектора на оси координат равны соответственно действительной и мнимой частям комплексной величины. Заметим, что на так называемой *амплитудно-фазовой плоскости — диаграмме* (на одной оси амплитуда, а на другой — фаза) сигнал может быть представлен в виде точки, соответствующей концу вектора. Такое представление используют для описания видов модуляции в модемах. Комплексная форма представления сигналов подключает пространственное воображение, расширяет рамки анализа, упрощает математические операции, проводимые с сигналами.

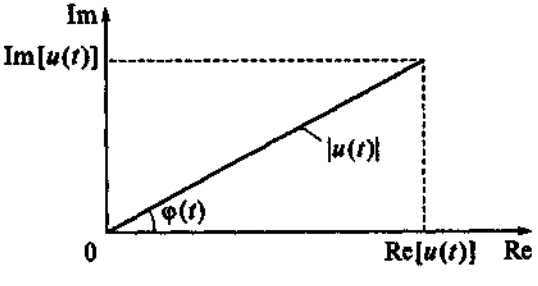



**Рис. 2.5.** Графическое представление комплексной формы сигнала

Векторное представление сигналов базируется на функциональном анализе — разделе математики, объединяющем методы и подходы топологии, классического анализа и линейной алгебры, и позволяющим создать стройную аналитическую теорию сигналов. В геометрической форме элемент $U$ в $n$-мерном пространстве представляют в виде точки или вектора с координатами $u_1$, $u_2$, ..., $u_n$. При такой интерпретации множеству сигналов ставят в соответствие *линейное векторное пространство* $L$. Сигналы в этом пространстве изображают векторами и операции с сигналами заменяют операциями с векторами. Если число членов множества $n$ стремится к бесконечности, то принято говорить о *бесконечномерном* пространстве $L$.

Пространство $L$ называется *нормированным*, если введена *норма*, т. е. расстояние между началом координат и какой-либо точкой пространства. Для вещественного и комплексного сигналов, определяемых на временно́м интервале $t_1 - t_2$ (часто удобнее интервал обозначать как $0 - T$), норма соответственно запишется как:

$$\|u\| = \sqrt{\int_{t_1}^{t_2} u^2(t)dt}\,; \qquad \|u\| = \sqrt{\int_{t_1}^{t_2} u(t)u^*(t)dt}\,, \tag{2.12}$$

где $u^*(t)$ — сигнал, комплексно-сопряженный сигналу $u(t)$.

Норма представляет геометрическую трактовку линейного пространства сигналов и по своему смыслу соответствует длине вектора сигнала. Нормированные пространства функций имеют важное значение для спектрального представления сигналов. Если сигнал является дискретным, то операцию интегрирования в (2.12) следует заменить операцией суммирования по всем отсчетам сигнала.

Еще одним фундаментальным понятием линейного пространства сигналов является *метрика*. Пространство сигналов называется *метрическим*, если введен способ определения *метрики* — расстояния $d(u, v)$ между двумя его элементами (здесь сигналами), например $u(t)$ и $v(t)$. Метрика — неотрицательное число, которое независимо от способа задания должно удовлетворять ряду известных в математике аксиом (для упрощения не приводятся).

Обычно используют такую аналитическую запись *метрики* пространства:

$$d(u, v) = \sqrt{\sum_{i=1}^{n} (u_i - v_i)^2}\,. \tag{2.13}$$

Пространство функций с метрикой (2.13) называют *n-мерным евклидовым пространством*. Если математические модели сигналов являются комплексными функциями, то приходим уже к *комплексному линейному пространству*.

Кроме нормы и метрики вводят *скалярное произведение сигналов*.

$$(u, v) = \int_{t_2} u(t)v(t)dt. \tag{2.14}$$



Д. Гильберт

Скалярное произведение сигналов (функций) обладает рядом свойств:

- $(u_k, u_n) \geq 0$;
- $(u + v) = (v + u)$;
- $(\alpha u, v) = \alpha(u, v)$, где $\alpha$ — вещественное число;
- $(u + v, s) = (u, s) + (v, s)$.

Полное линейное пространство с такими свойствами называют *вещественным гильбертовым пространством H* (Давид Гильберт — *David Hilbert*; немецкий математик; 1862–1943).

При анализе комплексных сигналов можно определить *комплексное гильбертово пространство*, введя в нем скалярное произведение

$$(u, v) = \int_{t_1}^{t_2} u(t)v^*(t)dt, \qquad (2.15)$$

такое, что $(u, v) = (u + v)^*$. В (2.15) $v^*(t)$ — сигнал, комплексно-сопряженный сигналу $v(t)$.

Для скалярного произведения сигналов справедливо фундаментальное *неравенство Коши–Буняковского–Шварца* (см. гл. 9)

$$|(u, v)| \leq \|u\| \cdot \|v\|. \qquad (2.16)$$

Сигнал, описываемый выражением

$$u(t) = \sum_{i=1}^{N} c_i v_i(t), \qquad (2.17)$$

есть $n$-мерный вектор линейного пространства (рис. 2.6).

Ортонормированная (ортогональная и нормированная к 1) система базисных функций $\{v_i(t)\}$ образует координатную систему в $n$-мерном евклидовом пространстве (частном случае гильбертового). Функции $v_i(t)$ представляют единичные векторы (*орты*, откуда ортонормированный, нормированные по длине к 1), коэффициенты $c_n$ — проекции вектора сигнала $u(t)$ на оси координат.

Координаты вектора — скалярное произведение функций $u(t)$ и $v_i(t)$

$$c_k = \int_{t_1}^{t_2} u(t)v_i(t)dt. \qquad (2.18)$$

Необходимо отметить, что путем поворота системы координат относительно ее начала можно получить бесчисленное множество различных координатных систем.

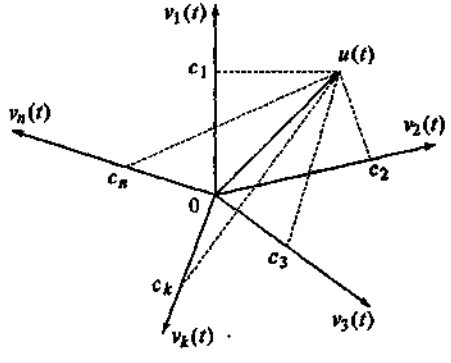



**Рис. 2.6.** Векторное представление сигнала

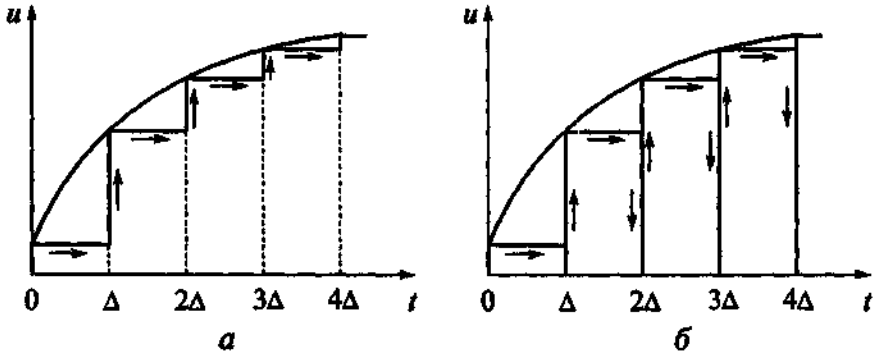



**Рис. 2.7.** Способы динамического представления сигналов:
*а* — ступенчатыми функциями; *б* — прямоугольными импульсами
(стрелками показаны направления изменения во времени элементарных слагаемых)

Замена координатной системы означает замену системы базисных функций, используемых при разложении реальных сигналов. Свойства же векторов, отражающие основные свойства радиотехнических сигналов, остаются неизменными, в том числе длины векторов, расстояния и углы между векторами.

### Представление сигналов динамическими моделями

Применяют два способа динамического представления сигналов (рис. 2.7). Согласно первому — в качестве элементарных сигналов используют ступенчатые функции, возникающие через равные интервалы времени Δ (рис. 2.7, *а*). Высота каждой ступеньки равна приращению сигнала на интервале Δ. При втором способе представления элементарными сигналами служат прямоугольные импульсы длительностью Δ. Импульсы примыкают друг к другу и образуют последовательность, вписанную в кривую или описанную вокруг нее (рис. 2.7, *б*).

### Энергетические характеристики сигналов

На практике одной из важных составляющих анализа сигналов является измерение их количественных параметров. Наиболее часто специалистами используются такие параметры радиотехнического сигнала, как энергия и средняя мощность (и связанное с ней среднее квадратическое значение).

Для сигнала $u(t)$ *среднее квадратическое значение* (*root mean square — RMS*) определяется как корень квадратный из среднего квадрата его мгновенного значения (корень квадратный из средней мощности) за интервал времени $t_1 - t_2$

$$U = \sqrt{\frac{1}{T}\int_{t_1}^{t_2} u^2(t)dt}\,. \qquad (2.19)$$

Во все формулы, определяющие энергетические параметры сигнала, должно входить сопротивление нагрузки $R$.

Однако, если энергия и мощность интересуют нас не как физические величины, а как средство *сравнения* различных сигналов, этот параметр можно из формул исключить (т. е. принять сопротивление $R$ = 1 Ом). Тогда получим определения энергии и средней мощности, принятые в теории сигналов.

*Энергия* сигнала $u(t)$ [$В^2 \cdot с$] за интервал времени $t_1 - t_2$ (это квадрат нормы сигнала)

$$Э = \| u(t) \|^2 = \int_{t_1}^{t_2} u^2(t)dt. \qquad (2.20)$$

Именно такая энергия выделяется на резисторе с сопротивлением 1 Ом, если на его зажимы подано напряжение $u(t)$.

*Средняя мощность* (*average power*) сигнала [$В^2$] на том же интервал времени:

$$P = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} u^2(t)dt. \qquad (2.21)$$

Для периодического сигнала с периодом $T = t_2 - t_1$ средняя мощность:

$$P = \frac{1}{T} \int_{0}^{T} u^2(t)dt. \qquad (2.22)$$

## 2.2. Спектральное представление детерминированных сигналов

Часто математическое описание даже несложных детерминированных сигналов является весьма трудной задачей. Поэтому в радиотехнике используют оригинальный прием, при котором реальные, сложные по структуре и форме сигналы заменяют (представляют) набором (взвешенной суммой) математических моделей, описываемых элементарными функциями. Это дает важный инструмент для анализа прохождения сигналов через цепи. Кроме того, представление сигнала в виде ряда может использоваться и как исходное при его описании и анализе. При этом можно существенно упростить обратную задачу — *синтез* сложных сигналов из совокупности элементарных функций.

### *Спектральное представление периодических сигналов. Ряды Фурье*

**Обобщенный ряд Фурье.** Фундаментальная идея спектрального представления сигналов (функций) восходит к временам более чем 200-летней давности и принадлежит знаменитому французскому физику и математику Жану Б. Фурье (*Jean B. Fourier*; 1768–1830). Рассмотрим системы элементарных ортогональных функций (сигналов), каждая из которых получается из одной исходной — функции-прототипа.



Ж. Фурье



Л. Эйлер

Эта функция-прототип выполняет роль «строительного блока», а искомая аппроксимация получается соответствующим комбинированием одинаковых по структуре блоков. При этом если «хорошая» аппроксимация получается при использовании небольшого числа блоков, то достигается значительное сжатие информации (о сжатии информации см. далее).

Ж. Фурье удалось доказать оригинальную теорему, которая в буквальном смысле ошеломила его скептически настроенных оппонентов, включая видных математиков и физиков того времени: П. Лапласа (*P. Laplace*), Ж. Био (*J. Biot*), С. Пуассона, Леонарда Эйлера (*Leonard Euler*; 1707–1783; знаменитый швейцарский математик) и др. Фурье свел единую функцию, трудно поддающуюся математическому описанию, к более удобным в обращении рядам гармонических тригонометрических функций, которые в сумме дают исходную функцию. Фурье показал, что любую сложную функцию можно представить (аппроксимировать) в виде конечной или бесконечной суммы ряда кратных гармонических колебаний с определенными амплитудами, частотами и начальными фазами. Этой функцией может быть, в частности, ток или напряжение в цепи.

Солнечный луч, разложенный призмой на спектр цветов, представляет физический аналог математических преобразований Фурье (рис. 2.8). Свет, выходящий из призмы, разделен в пространстве на отдельные «чистые» цвета, или частоты. В спектре имеется средняя амплитуда на каждой частоте. Таким образом, функция интенсивности от времени трансформировалась в функцию амплитуды в зависимости от частоты.

Столь простое и наглядное представление сложного изменения во времени какой-либо физической величины в виде суммы ряда простейших гармониче-

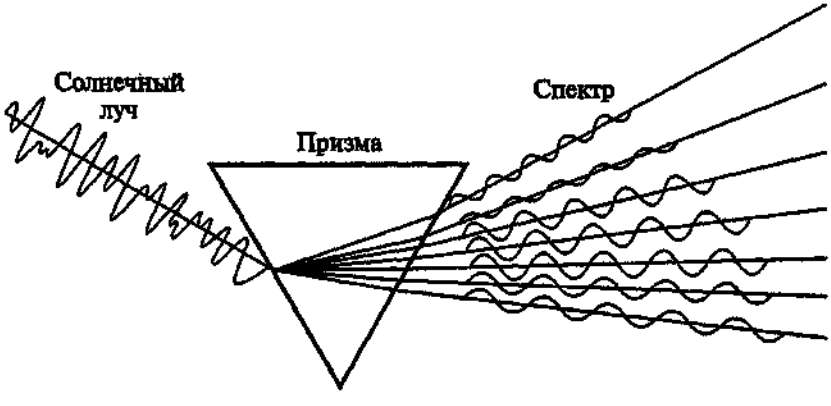



**Рис. 2.8.** Разложение света на «цвета»

ских колебаний могло показаться, на первый взгляд, лишь математическим трюком. Но это не трюк. Простой пример иллюстраций рассуждений Фурье показан на рис. 2.9. Периодическая, достаточно сложная по форме кривая (рис. 2.9, *а*), — это сумма двух гармоник разных, но кратных частот: первой (рис. 2.9, *б*) и удвоенной (рис. 2.9, *в*).

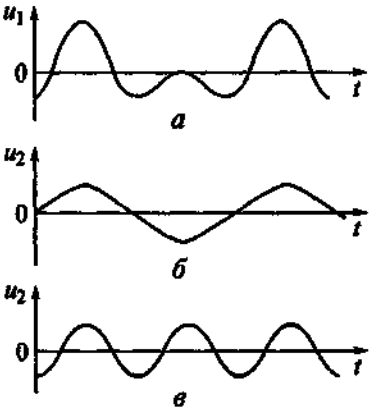



**Рис. 2.9.** К анализу Фурье:
*а* — сложное колебание; *б*, *в* — первый и второй аппроксимирующие сигналы

При помощи анализа Фурье сложная функция представляется суммой гармонических составляющих, каждая из которых имеет свою частоту, амплитуду и начальную фазу. Преобразование Фурье определяет функции, представляющие амплитуду и фазу гармонических составляющих, соответствующие конкретной частоте (напомним, что амплитуда сигнала представляет модуль наибольшего его отклонения от нуля, следовательно, амплитуда всегда положительна, а фаза — это начальная точка синусоиды; кстати, иногда в литературе по цифровой обработке сигналов вместо термина «модуль» употребляют термин «абсолютное значение»). Преобразование можно получить двумя разными математическими методами, один из которых применяют, когда исходная функция непрерывна, а другой — когда она состоит из множества отдельных дискретных измерений. Если исследуемая функция получена из значений с определенными дискретными интервалами, то ее можно разбить на последовательный ряд синусоидальных функций с дискретными частотами — от самой низкой, основной или главной частоты, и далее с частотами, вдвое, втрое и т. д. выше (т. е. кратно превышающими) основной. Впрочем, данный ряд может содержать и постоянную составляющую. Такая сумма составляющих и называется рядом Фурье.

### Ортогональные сигналы

Наиболее удобным способом спектрального описания исследуемого сигнала по Фурье является его аналитическое представление с помощью системы некоторых взаимосвязанных элементарных функций времени, являющихся ортогональными. Обратимся к определениям, характеризующим основные свойства ортогональных и ортонормированных систем функций.

Пусть имеется гильбертово пространство сигналов $u_0(t)$, $u_1(t)$, ..., $u_n(t)$ с конечным значением энергии, определенных на конечном или бесконечном отрезке времени $(t_1, t_2)$. На этом отрезке зададим бесконечную систему (подмножество) взаимосвязанных элементарных функций времени и назовем ее *базисной*

$$v_0(t), v_1(t), v_2(t), ..., v_i(t), ..., \tag{2.23}$$

где $i = 1, 2, 3...$

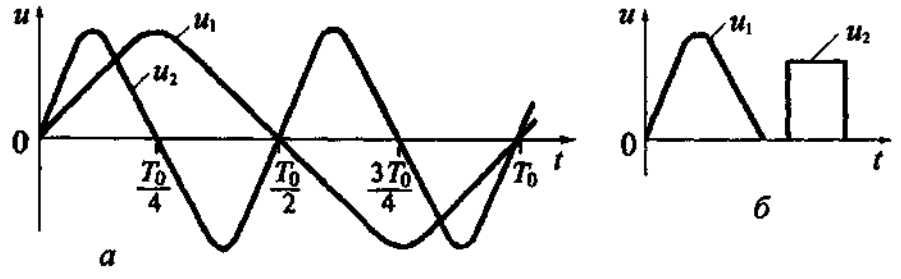



**Рис. 2.10.** Ортогональность сигналов:
*а* — на интервале; *б* — из-за разновременности появления

Функции $u(t)$ и $v(t)$ ортогональны на интервале $t_1 - t_2$, если их скалярное произведение

$$\int_{t_1}^{t_2} u(t)v(t)dt = 0, \tag{2.24}$$

при условии, что ни одна из этих функций не равна тождественно нулю.

В математике так задают в гильбертовом пространстве сигналов *ортогональный координатный базис*, т. е. систему ортогональных базисных функций.

Свойство ортогональности функций (сигналов) связано с интервалом их определения (рис. 2.10). Например, два гармонических сигнала $u_1(t) = \sin(2\pi t/T_0)$ и $u_2(t) = \sin(4\pi t/T_0)$ (т. е. с частотами $f_0 = 1/T_0$ и $2f_0$ соответственно) ортогональны на любом интервале времени, длительность которого равна целому числу полупериодов $T_0$ (рис. 2.10, *а*). Следовательно, в первом периоде сигналы $u_1(t)$ и $u_2(t)$ ортогональны на интервале $0 - T_0/2$; но на интервале $0 - 3T_0/4$ — неортогональны. На рис. 2.10, *б* сигналы ортогональны из-за разновременности их появления.

Представление сигнала $u(t)$ элементарными моделями существенно упрощается, если выбрана система базисных функций $v_i(t)$, обладающих свойством *ортонормированности*. Из математики известно, если для любой пары функций из ортогональной системы (2.23) выполняется условие

$$\int_{t_1}^{t_2} v_i(t)v_k(t)dt = \begin{cases} 1, & i = k; \\ 0, & i \neq k, \end{cases} \tag{2.25}$$

то система функций, описываемая (2.23), *ортонормированна*.

В математике такую систему базисных функций вида (2.23) называют *ортонормированным базисом*.

Пусть на заданном интервале времени $t_1$, $t_2$ действует произвольный электрический сигнал $u(t)$ и для его представления используется ортонормированная система идеализированных функций (2.23). Проектирование произвольного электрического сигнала $u(t)$ на оси координатного базиса называется *разложением в обобщенный ряд Фурье*.

Это разложение имеет вид

$$u(t) = \sum_{i=0}^{\infty} c_i v_i(t), \tag{2.26}$$

где $c_i$ — некоторые постоянные коэффициенты.

Для определения коэффициентов $c_k$ обобщенного ряда Фурье выберем одну из базисных функций (2.23) $v_k(t)$ с произвольным номером $k$.

Умножим обе части выражения (2.26) на эту функцию и проинтегрируем результат по времени

$$\int_{t_1}^{t_2} u(t)v_k(t)dt = \sum_{i=0}^{\infty} c_i \int_{t_1}^{t_2} v_i(t)v_k(t)dt.$$

Вследствие ортонормированности базиса выбранных функций в правой части этого равенства все члены суммы при $i \neq k$ обратятся в нуль. Ненулевым останется только единственный член суммы с номером $i = k$, поэтому

$$c_k = \int_{t_1}^{t_2} u(t)v_k(t)dt. \tag{2.27}$$

Произведение вида $c_k v_k(t)$, входящее в обобщенный ряд Фурье (2.26), представляет собой *спектральную составляющую* сигнала $u(t)$, а совокупность коэффициентов (проекций векторов сигнала на оси координат) $\{c_0, c_1, \ldots, c_k, \ldots, c_n\}$ полностью определяет анализируемый сигнал $u(t)$ и называется его *спектром* (от лат. *spectrum* — представление, образ).

Суть *спектрального представления* (часто *анализа*) сигнала состоит в определении коэффициентов $c_n$ (экспериментально или аналитически) в соответствии с (2.27). Заметим, что у относительно гладких функций спектр быстро убывает (с ростом номера коэффициенты быстро стремятся к нулю). Для относительно «изрезанных» функций спектр убывает медленно, так как для представления разрывов и «изломов» функции нужны синусоиды с большими частотами.

Достаточно распространенное использование в теории сигналов обобщенного ряда Фурье связано с его очень важным свойством: при выбранной ортонормированной системе функций $v_k(t)$ и фиксированном числе слагаемых ряда (2.26) он обеспечивает наилучшее представление заданного сигнала $u(t)$. Это свойство рядов Фурье широко известно.

При спектральном представлении сигналов наибольшее применение получили ортонормированные базисы тригонометрических (синусоидальных и косинусоидальных) функций. Это обусловлено следующим. Во-первых, гармонические колебания наиболее просто генерировать; гармонические сигналы инвариантны относительно преобразований, осуществляемых стационарными линейными электрическими цепями.

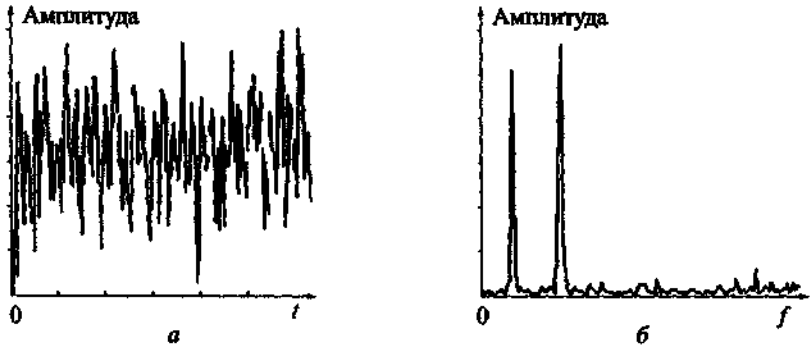



**Рис. 2.11.** Спектральное представление сигнала:
*а* — временна́я диаграмма; *б* — амплитудный спектр

Гармоническое колебание теоретически полностью сохраняет форму при прохождении через любую линейную цепь с постоянными параметрами, а изменяются при этом лишь его амплитуда и начальная фаза. Во-вторых, широко используется хорошо разработанный в теории электрических цепей символический метод. Операцию представления детерминированных сигналов в виде совокупности постоянной составляющей (*constant component*; *direct current*) и суммы гармонических колебаний с кратными частотами принято называть *спектральным разложением* (*представлением*).

Наглядно оценим временно́е и спектральное представление аналогового сигнала (рис. 2.11). На рис. 2.11, *а* показана временна́я диаграмма достаточно сложного по форме непрерывного сигнала, а на рис. 2.11, *б* — его спектральное разложение. Рассмотрим спектральное представление простейшего вида детерминированных сигналов — периодических. При этом они представляются в виде суммы либо гармонических функций, либо комплексных экспонент с частотами, образующими арифметическую прогрессию.

*Периодическим* называют любой электрический сигнал (функцию) $u_п(t)$, повторяющийся через регулярные интервалы времени (рис. 2.12) и удовлетворяющий условию (для удобства будем записывать $u_п(t) = u(t)$):

$$u(t) = u_п(t) = u(t \pm nT), \tag{2.28}$$

где $T$ — период повторения или следования импульсов; $n = 0, 1, 2, \ldots$

Если величина $T$ является периодом сигнала $u(t)$, то периодами для него будут и кратные ей значения: $2T$, $3T$ и т. д.

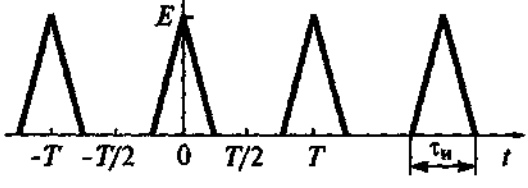



**Рис. 2.12.** Периодический сигнал

Периодическая последовательность импульсов (в радиотехнике такие, обычно однополярные, импульсы прямоугольной, трапецеидальной и другой формы называют *видеоимпульсами*; исторически происхождение термина связано с тем, что впервые импульсы были применены в телевидении) имеет вид:

$$u(t) = \sum_{n=-\infty}^{\infty} u_0(t - nT). \qquad (2.29)$$

Здесь $u_0(t)$ — форма одиночного импульса, характеризующаяся амплитудой (высотой) $h = E$; длительностью $\tau_и$; периодом следования $T = 1/F$ ($F$ — частота); положением импульсов во времени относительно тактовых точек, например $t = 0$.

Если форма импульса не прямоугольная, появляется ряд дополнительных параметров: крутизна фронта и среза (заднего фронта), длительность вершины и т. д.

При спектральном анализе периодических сигналов весьма удобна ортогональная система (2.23) в виде гармонических функций с кратными частотами:

$$1,\ \sin\omega_1 t,\ \cos\omega_1 t,\ \sin 2\omega_1 t,\ \cos 2\omega_1 t,\ \ldots,\ \sin n\omega_1 t,\ \cos n\omega_1 t, \qquad (2.30)$$

где $\omega_1 = 2\pi/T$ — частота следования импульсов.

Вычисляя интегралы по (2.24), легко убедиться в ортогональности этих функций на интервале $-T/2$, $T/2$. Любая функция удовлетворяет условию периодичности (2.29), поскольку частоты их кратны. Если систему (2.30) записать как

$$\sqrt{\frac{1}{T}}, \sqrt{\frac{2}{T}}\cos\omega_1 t, \sqrt{\frac{2}{T}}\sin\omega_1 t, \sqrt{\frac{2}{T}}\sin 2\omega_1 t, \sqrt{\frac{2}{T}}\cos 2\omega_1 t, \ldots, \sqrt{\frac{2}{T}}\cos n\omega_1 t, \sqrt{\frac{2}{T}}\sin n\omega_1 t, \qquad (2.31)$$

то получим ортонормированный базис гармонических функций.

Представим периодический сигнал наиболее распространенной в теории сигналов тригонометрической (синусно-косинусной) формой ряда Фурье:

$$u(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos n\omega_1 t + b_n \sin n\omega_1 t). \qquad (2.32)$$

Из курса математики известно, что разложение (2.29) существует, т. е. ряд сходится, если функция (в данном случае сигнал) $u(t)$ на интервале $[-T/2, T/2]$ удовлетворяет *условиям Дирихле* (в отличие от теоремы Дирихле их часто трактуют упрощенно; П. Дирихле (*P. Dirichlet*; 1805–1859; немецкий математик)):

- не должно быть разрывов 2-го рода (с уходящими в бесконечность ветвями);
- функция ограничена и имеет конечное число разрывов 1-го рода (скачков);
- имеет конечное число экстремумов (т. е. максимумов и минимумов).



П. Дирихле

В (2.32) имеются следующие компоненты анализируемого сигнала:

– постоянная составляющая

$$a_0 = \frac{2}{T}\int_{-T/2}^{T/2} u(t)dt; \tag{2.33}$$

– амплитуды косинусоидальных составляющих:

$$a_n = \frac{2}{T}\int_{-T/2}^{T/2} u(t)\cos n\omega_1 t dt; \tag{2.34}$$

– амплитуды синусоидальных составляющих:

$$b_n = \frac{2}{T}\int_{-T/2}^{T/2} u(t)\sin n\omega_1 t dt. \tag{2.35}$$

Спектральную составляющую с частотой $\omega_1$ в радиотехнике называют *первой* (*основной*) *гармоникой*, а составляющие с частотами $n\omega_1$ ($n > 1$) — *высшими гармониками* периодического сигнала.

Из курса математики известно, что если сигнал представляет собой четную функцию времени $u(t) = u(-t)$, то в тригонометрической записи ряда Фурье (2.32) отсутствуют синусоидальные коэффициенты $b_n$, так как в соответствии с формулой (2.35) они обращаются в нуль. Для сигнала $u(t)$, описываемого нечетной функцией времени, наоборот, согласно формуле (2.34), нулю равны косинусоидальные коэффициенты $a_n$, и ряд содержит составляющие $b_n$ (кстати, постоянная составляющая $a_0$ также отсутствует). Заметим, что пределы интегрирования (от $-T/2$ до $T/2$) не обязательно должны быть такими, как в приведенных формулах (2.33 – 2.35). Интегрирование может производиться по *любому* интервалу времени шириной $T$ — результат от этого не изменится. Конкретные пределы выбираются из соображений удобства вычислений; например, может оказаться проще выполнять интегрирование от 0 до $T$ или от $-T$ до 0, и т. д.

Раздел математики, устанавливающий соотношение между функцией времени $u(t)$ и спектральными коэффициентами $a_n$, $b_n$, называют *гармоническим анализом* вследствие связи функции $u(t)$ с синусоидальными и косинусоидальными членами этой суммы. Далее спектральный анализ в основном ограничен рамками гармонического анализа, находящего исключительное применение.

Часто применение синусно-косинусной формы ряда Фурье не совсем удобно, поскольку для каждого значения индекса суммирования $n$ (т. е. для каждой гармоники с частотой $n\omega_1$) в формуле (2.32) фигурируют два слагаемых — косинус и синус. С математической точки зрения удобнее эту формулу представить эквивалентным рядом Фурье в *вещественной форме*:

$$u(t) = A_0 + \sum_{n=1}^{\infty} A_n \cos(n\omega_1 t - \varphi_n), \tag{2.36}$$

где $A_0 = a_0/2$; $A_n = \sqrt{a_n^2 + b_n^2}$ — амплитуда; а $\varphi_n = \text{arctg}(b_n/a_n)$ — начальная фаза $n$-й гармоники сигнала.

Иногда в соотношении (2.36) перед $\varphi_n$ ставят знак «+», тогда начальную фазу гармоник записывают как $\varphi_n = -\text{arctg}(b_n/a_n)$.

В радиоэлектронике и теории сигналов также широко используют комплексную форму ряда Фурье. Она получается из вещественной формы ряда представлением косинуса в виде полусуммы комплексных экспонент. Представление вытекает из формулы Эйлера $e^{jx} = \cos x + j\sin x$:

$$\cos x = 0{,}5(e^{jx} + e^{-jx}). \tag{2.37}$$

Применив данное преобразование к вещественной форме ряда Фурье (2.36), получим суммы комплексных экспонент с положительными и отрицательными показателями:

$$u(t) = A_0 + \sum_{n=1}^{\infty} \frac{A_n}{2}\left(e^{jn\omega_1 t - \varphi_n} + e^{-jn\omega_1 t + \varphi_n}\right). \tag{2.38}$$

А теперь будем трактовать в (2.38) экспоненты при частоте $\omega_1$ со знаком минус в показателе как члены ряда с отрицательными номерами. В рамках этого же подхода коэффициент $A_0$ станет членом ряда с нулевым номером. После несложных преобразований приходим к комплексной форме ряда Фурье

$$u(t) = \sum_{n=-\infty}^{\infty} \mathbf{C}_n e^{jn\omega_1 t}, \tag{2.39}$$

где

$$\mathbf{C}_n = \dot{\mathbf{C}}_n = \mathbf{C}_n(j\omega) = \frac{1}{T}\int_{-T/2}^{T/2} u(t)e^{-jn\omega_1 t}\,dt \tag{2.40}$$

— комплексная амплитуда $n$-й гармоники.

Определим связь между коэффициентами тригонометрической и комплексной форм ряда Фурье. Очевидно, что

$$C_0 = |\mathbf{C}_0| = A_0 = a_0/2;\ C_n = |\mathbf{C}_n| = |\mathbf{C}_{-n}| = A_n/2. \tag{2.41}$$

Можно также показать, что коэффициенты:

$$a_n = 2\mathbf{C}_n\cos\varphi_n; \quad b_n = 2\mathbf{C}_n\sin\varphi_n. \tag{2.42}$$

Если $u(t)$ является *четной* функцией, коэффициенты ряда $\mathbf{C}_n$ будут *вещественными*, а если $u(t)$ — функция *нечетная*, коэффициенты ряда станут *мнимыми*.

Из формулы (2.39) нетрудно выяснить, что спектральное представление периодического сигнала комплексной формой ряда Фурье содержит как положительные, так и отрицательные частоты. Однако *отрицательные частоты* в природе не существуют, и это не физическое понятие, а математическая абстракция

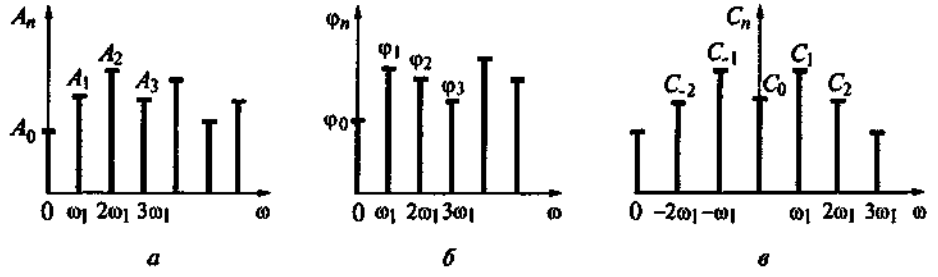



**Рис. 2.13.** Спектры периодического сигнала:
*а* — амплитудный; *б* — фазовый; *в* — амплитудный спектр комплексного ряда Фурье

(физический смысл отрицательной частоты — вращение в направлении, противоположном тому, которое принято за положительное). Они появляются как следствие формального представления гармонических колебаний комплексной формой. Легко показать, что при переходе от комплексной формы записи (2.39) к тригонометрической (2.36) «отрицательная частота» пропадает.

Наиболее наглядно о спектре сигнала судят по его графическому изображению — спектральной диаграмме. Различают *амплитудно-частотные* и *фазочастотные спектры* (не следует путать с амплитудно- и фазочастотными *характеристиками* электрических цепей). Совокупность амплитуд гармоник $A_n$ называют *амплитудным спектром*, их фаз $\varphi_n$ — *фазовым спектром*. Совокупность $C_n = |\mathbf{C}_n|$ является *комплексным амплитудным спектром* (рис. 2.13). На спектральных диаграммах по оси абсцисс откладывают текущую частоту, а по оси ординат — либо вещественную, либо комплексную амплитуду, или фазу соответствующих гармонических составляющих анализируемого сигнала.

Спектр периодического сигнала называют *линейчатым* или *дискретным*, так как он состоит из отдельных линий, высота которых равна амплитуде $A_n$ гармоник. Из всех видов спектров наиболее информативен амплитудный, поскольку с его помощью можно оценить количественное содержание тех или иных гармоник в частотном составе сигнала. В теории сигналов доказано, что амплитудный спектр есть *четная функция частоты*, а фазовый — *нечетная*.

Отметим *эквидистантность* (равноудаленность от начала координат) комплексного спектра периодических сигналов: симметричные (положительные и отрицательные) частоты, на которых расположены спектральные коэффициенты тригонометрического ряда Фурье, образуют эквидистантную последовательность (..., $-n\omega_1$, ..., $-2\omega_1$, $-\omega_1$, 0, $\omega_1$, $2\omega_1$,..., $n\omega_1$...), обязательно содержащую частоту $\omega = 0$ и имеющую шаг $\omega_1 = 2\pi/T$. Сами коэффициенты могут принимать любые, в том числе и нулевые, значения.

**Пример 2.1.** Рассчитать амплитудный и фазовый спектры периодической последовательности прямоугольных импульсов, имеющих амплитуду $E$, длительность $\tau_и$, период повторения $T$. Сигнал — функция четная (рис. 2.14).

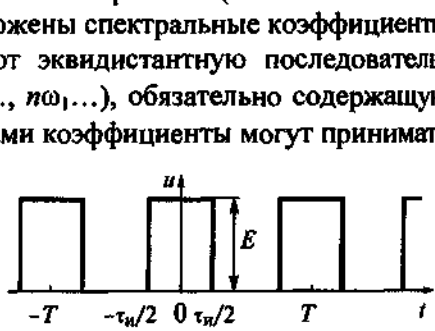



**Рис. 2.14.** Последовательность прямоугольных импульсов

Р е ш е н и е. Из курса математики известно, что одиночный идеальный прямоугольный видеоимпульс описывается следующим уравнением:

$$u(t) = E[\sigma(t + \tau_и/2) - \sigma(t - \tau_и/2)], \tag{2.43}$$

т. е. он формируется как разность двух единичных функций $\sigma(t)$ (функций включения), сдвинутых во времени на $\tau_и$.

Последовательность прямоугольных импульсов представляет собой известную сумму одиночных импульсов:

$$u(t) = \sum_{k=0}^{\infty} E\left[\sigma(t + kT + \tau_и/2) - \sigma(t - kT - \tau_и/2)\right] \tag{2.44}$$

Отметим, что идеальные прямоугольные импульсы в природе практически не встречаются. В реальных импульсах время изменения сигнала от нулевых до амплитудных значений (и обратно) всегда имеет конечную длительность.

Поскольку заданный сигнал является четной функцией времени и в течение одного периода действует только на интервале $-\tau_и/2$, $\tau_и/2$, то, согласно (2.33):

$$\frac{a_0}{2} = A_0 = \frac{1}{T}\int_{-\tau_и/2}^{\tau_и/2} E dt = \frac{E}{T}\tau_и = \frac{E}{q}, \tag{2.45}$$

где $q = T/\tau_и$.

Анализируя полученную формулу, можно заметить, что период следования и длительность импульсов входят в нее не обособленно, а в виде отношения. Этот параметр $q$ — отношение периода к длительности импульсов — называют *скважностью* периодической последовательности импульсов (в зарубежной литературе вместо скважности используют обратную величину — *коэффициент заполнения*, от анг. *duty cycle*, равный $\tau_и/T$); при $q = 2$ последовательность прямоугольных импульсов, когда длительности импульсов и промежутков между ними становятся равными, называют *меандром* (от греч. *узор* геометрический орнамент). Кстати, меандр может быть как однополярным, так и разнополярным.

В силу четности функции, описывающей анализируемый сигнал, в ряде Фурье, наряду с постоянной составляющей, будут присутствовать только косинусоидальные составляющие (2.34)

$$a_n = \frac{2}{T} = \int_{-\tau_и/2}^{\tau_и/2} E\cos n\omega_1 t dt = \frac{2E}{q}\frac{\sin(n\omega_1\tau_и/2)}{n\omega_1\tau_и/2}. \tag{2.46}$$

В правой части данной формулы второй сомножитель имеет вид элементарной функции $(\sin x)/x$. В математике эту функцию обозначают как $\mathrm{sinc}(x) = \mathrm{sinc} = (\sin x)/x$, причем только при значении $x = 0$ она равна единице $\left[\lim_{x\to 0}(\sin x/x) = 1\right]$, проходит через нуль в точках $x = \pm\pi, \pm 2\pi, \ldots$ и затухает с ростом аргумента $x$ (рис. 2.15). Окончательно тригонометрический ряд Фурье (2.32), который

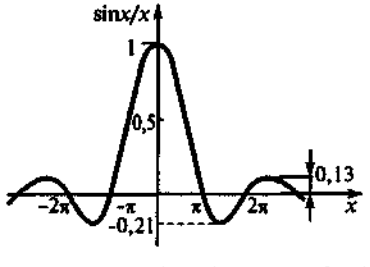



**Рис. 2.15.** График функции $\sin x/x$

аппроксимирует заданный сигнал, записывают в следующей форме

$$u(t) = \frac{E}{q}\left[1 + 2\sum_{n=1}^{\infty} \frac{\sin(n\omega_1 \tau_и/2)}{n\omega_1 \tau_и/2} \cos n\omega_1 t\right]. \qquad (2.47)$$

Функция sin*c* имеет лепестковый характер. Говоря о ширине лепестков, следует подчеркнуть, что для графиков дискретных спектров периодических сигналов возможны два варианта градуировки горизонтальной оси — в номерах гармоник и частотах. Например, на рис. 2.13 градуировка оси ординат соответствует частотам. Ширина лепестков, измеренная в числе гармоник, равна скважности импульсной последовательности. Отсюда следует важное свойство спектра последовательности прямоугольных импульсов — в нем отсутствуют (имеют нулевые амплитуды) гармоники с номерами, кратными скважности. При скважности импульсов, равной трем, исчезает каждая третья гармоника. Если бы скважность была бы равна двум, то в спектре остались бы лишь нечетные гармоники основной частоты.

Из формулы (2.46) и рис. 2.15 следует, что коэффициенты ряда высших гармоник сигнала имеют отрицательный знак. Это связано с тем, что начальная фаза этих гармоник равна π. Поэтому (2.46) принято представлять в измененном виде:

$$A_n = |a_n| = \frac{2E}{\pi n}\left|\sin\frac{n\omega_1 \tau_и}{2}\right|. \qquad (2.48)$$

При такой записи ряда Фурье значения амплитуд всех высших гармонических составляющих на графике спектральной диаграммы положительны (рис. 2.16, *а*).

Амплитудный спектр анализируемого сигнала в значительной степени зависит от отношения периода повторения $T$ и длительности импульса $\tau_и$, т. е. от скважности $q$. Расстояние по частоте между соседними гармониками спектра равно частоте следования импульсов $\omega_1 = 2\pi/T$.

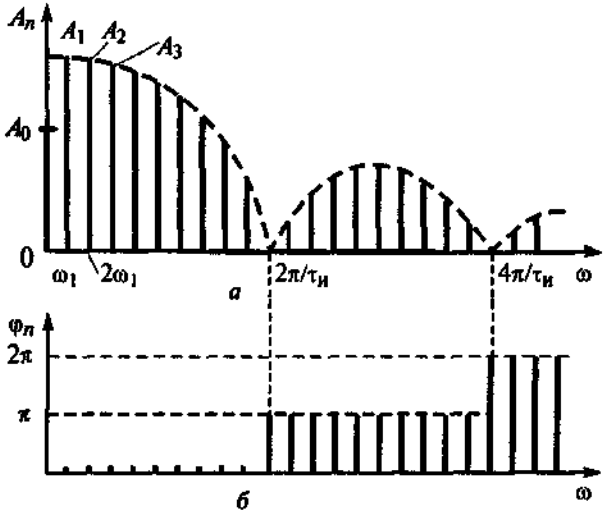



**Рис. 2.16.** Спектры последовательности прямоугольных импульсов:
*а* — амплитудный; *б* — фазовый

Ширина лепестков спектра последовательности, измеренная в единицах частоты, равна $2\pi/\tau_и$, т. е. обратно пропорциональна длительности импульсов. Отметим, что при одной и той же длительности импульса $\tau_и$ с увеличением периода их повторения $T$ основная частота $\omega_1$ уменьшается, и спектр становится плотнее. Ту же картину наблюдают, если укорачивают длительность импульса $\tau_и$ при неизменном периоде $T$. Амплитуды всех гармоник при этом уменьшаются. Это проявление общего закона (принципа неопределенности В. Гейзенберга), чем короче длительность сигнала, тем шире его спектр.

Фазы составляющих определим из формулы $\varphi_n = \operatorname{arctg}(b_n/a_n)$. Так как здесь коэффициенты $b_n = 0$, то

$$\varphi_n = \operatorname{arctg}\frac{0}{a_n} = \pm m\pi, \qquad (2.49)$$

где $m = 0, 1, 2, \ldots$

Соотношение (2.49) показывает, что при вычислениях фаз спектральных составляющих имеем дело с математической неопределенностью. Для ее раскрытия обратимся к формуле (2.46), согласно которой амплитуды гармоник периодически меняют знак в соответствии с изменением знака функции $\sin(n\omega_1\tau_и/2)$. Изменение знака в (2.46) эквивалентно сдвигу фазы этой функции на $\pi$. Следовательно, когда данная функция положительна, фаза гармоники $\varphi_n = 2m\pi$, а когда отрицательна — $\varphi_n = (2m + 1)\pi$ ( рис. 2.16, *б*).

Заметим, что хотя амплитуды составляющих в спектре прямоугольных импульсов и уменьшаются с ростом частоты (см. рис. 2.16, *а*), этот спад довольно медленный (амплитуды убывают обратно пропорционально частоте). Для передачи таких импульсов без искажений необходима бесконечная полоса частот канала связи. Для сравнительно малозаметных искажений граничное значение полосы частот должно быть во много раз больше значения, обратного длительности импульса. Однако все реальные каналы имеют конечную полосу пропускания, что приводит к искажениям формы переданных импульсов.

При расчетах спектров периодических сигналов вычисление бесконечной суммы ряда Фурье вызывает определенные трудности и не всегда требуется, поэтому ограничиваются суммированием конечного количества слагаемых. Точность аппроксимации исходного сигнала зависит от числа суммируемых гармонических составляющих. Рассмотрим это на примере аппроксимации суммой из восьми первых гармоник последовательности прямоугольных импульсов (рис. 2.17). Сигнал имеет вид однополярного меандра с периодом повторения $T$, амплитудой $E = 1$ и длительностью импульсов $\tau_и = T/2$ (заданный сигнал — функция четная — рис. 2.17, *а*; скважность $q = 2$). Аппроксимация показана на рис. 2.17, *б*, причем на графиках показано число суммируемых гармоник.

Необходимо отметить, что в проводимой аппроксимации заданного периодического импульсного сигнала (см. рис. 2.14) тригонометрическим рядом Фурье (2.32) суммирование первой и высших гармоник будет осуществляться только по нечетным коэффициентам $n$, так как при четных их значениях и длительности импульса $\tau_и = T/2 = \pi/\omega_1$ величина $\sin(n\omega_1\tau_и/2) = \sin(n\pi/2)$ обращается в нуль (см. представление рядами Фурье четных функций).

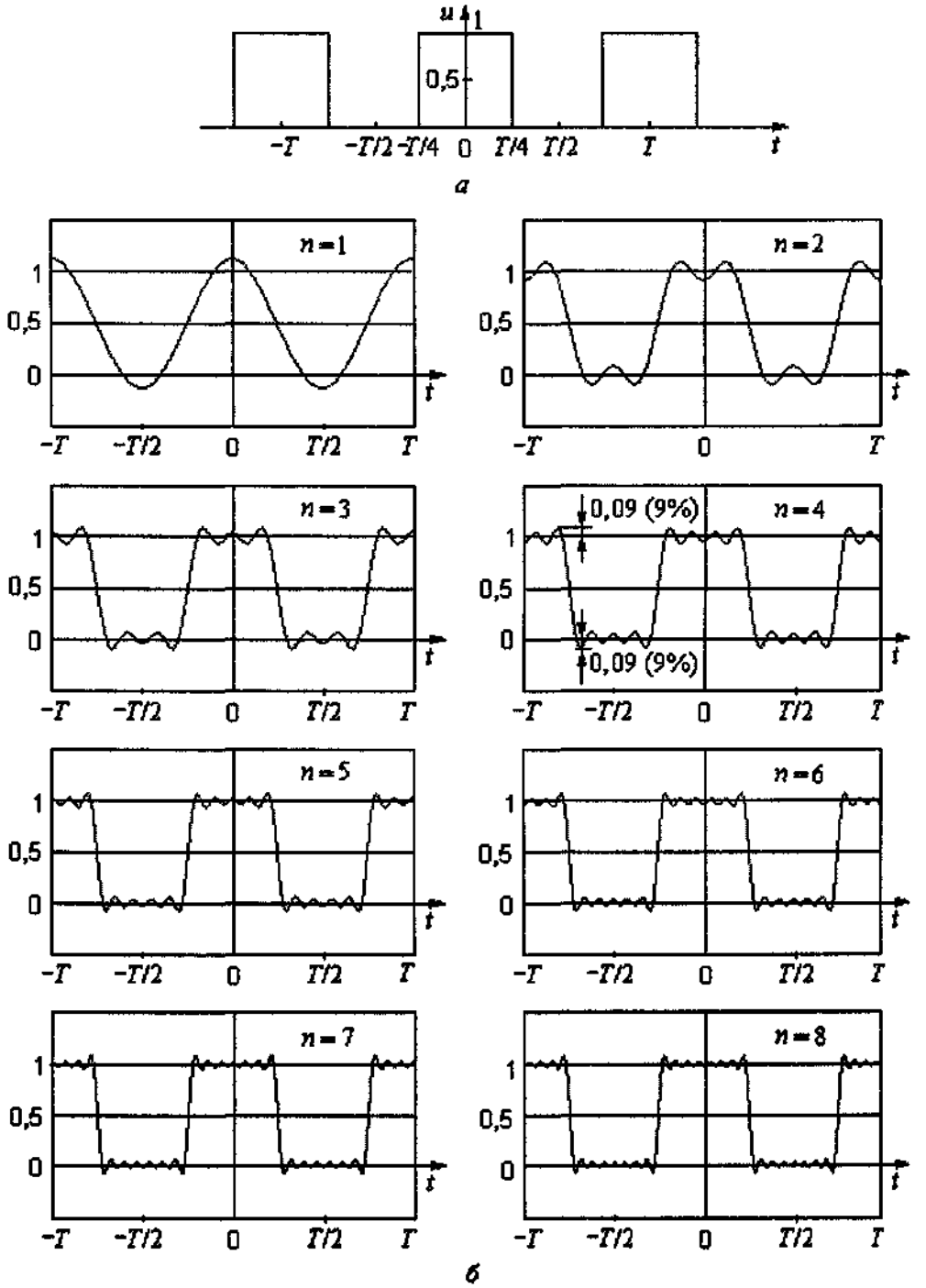



**Рис. 2.17.** Формирование меандра суммой гармоник:
*а* — заданный сигнал; *б* — промежуточные стадии суммирования ряда Фурье

Тригонометрическая форма ряда Фурье (2.47) для заданного сигнала

$$u(t)=\frac{E}{2}+\frac{2E}{\pi}\sum_{n=1}^{\infty}\frac{\sin(n\pi/2)}{n}\cos n\omega_1 t. \qquad (2.50)$$

Для удобства представления ряд Фурье (2.50) можно записать упрощенно:

$$u(t)=\frac{E}{2}+\frac{2E}{\pi}\left[\cos\left(\frac{2\pi}{T}t\right)-\frac{1}{3}\cos\left(3\frac{2\pi}{T}t\right)+\frac{1}{5}\cos\left(5\frac{2\pi}{T}t\right)-\cdots\right]. \qquad (2.51)$$

Из (2.51) видно, что гармоники, аппроксимирующие меандр, нечетные, имеют чередующиеся знаки, а их амплитуды обратно пропорциональны номерам. Отметим, что последовательность прямоугольных импульсов плохо подходит для представления рядом Фурье — она содержит пульсации и скачки, а сумма любого числа гармонических составляющих с любыми амплитудами всегда будет непрерывной функцией. Поэтому поведение ряда Фурье в окрестностях разрывов представляет особый интерес. Из графиков рис. 2.17, *б* нетрудно заметить, как с увеличением числа суммируемых гармоник напряжения результирующая аппроксимирующая функция все точнее приближается к форме исходного прямоугольного сигнала $u(t)$ везде, кроме точек ее разрыва. В окрестности точки разрыва суммирование ряда Фурье дает наклонный участок, причем крутизна наклона результирующей аппроксимирующей функции возрастает с ростом числа суммируемых гармоник. В самой точке разрыва (обозначим ее как $t=t_0$) ряд Фурье $u(t_0)$ сходится к полусумме правого и левого пределов

$$u(t_0)=\frac{1}{2}\left(\lim_{t\to t_0-0}u(t)+\lim_{t\to t_0+0}u(t)\right).$$

На примыкающих к разрыву участках аппроксимируемой кривой сумма ряда Фурье дает заметные пульсации, причем на рис. 2.16 видно, что амплитуда основного выброса этих пульсаций не уменьшается с ростом числа суммируемых гармоник — выброс лишь сжимается по горизонтали, приближаясь к точке разрыва. При $n\to\infty$ в точках разрыва амплитуда выброса остается постоянной, а его ширина будет бесконечно узкой. Это связано со сходимостью ряда Фурье.

Вопрос о сходимости всегда возникал при суммировании бесконечного ряда чисел и был предметом дискуссии между Фурье и Ж. Лагранжем (*J. Lagrange*; 1736–1813; французский математик). Фурье полагал, что любая функция может быть представлена таким разложением. Лагранж же возражал, приводя в пример функции с резкими изломами и указывая, что все слагаемые являются гладкими и представить излом ими нельзя. Окончательный ответ состоял в том, что приближение может быть сколь угодно точно в среднеквадратическом смысле, однако, на местах изломов остается некий скачок, причем по мере увеличения количества слагаемых скачок становится у́же, но отнюдь не ниже (хотя заключенная в нем энергия и стремится к нулю). Обратимся к классическому примеру:



Д. Гиббс

достигнете ли вы когда-нибудь стены, если с каждым шагом будете проходить половину оставшегося расстояния? Первый шаг приведет вас к отметке на половине пути, второй — к отметке на трех его четвертях, а после пятого шага преодолеете уже почти 97 % пути. Вы почти дошли до цели, однако, сколько бы еще шагов вперед ни сделали, никогда не достигнете ее в строгом математическом смысле. Можно лишь доказать математически, что в конце концов вы сможете приблизиться на любое заданное, сколько угодно малое расстояние. Данное доказательство будет эквивалентно демонстрации того, что сумма чисел 1/2, 1/4, 1/8, 1/16 и т. д. стремится к единице.

Это явление, присущее всем рядам Фурье для сигналов с разрывами первого рода (например, скачками, как на фронтах прямоугольных импульсов), называют *эффектом Гиббса* (*Gibbs phenomenon*; жаргонное, *уши Гиббса*; Д. Гиббс — *J. Gibbs*; 1839–1903; американский физик и математик, который дал ему пояснение). В математике показывают, что значение первого (самого большого) выброса амплитуды в аппроксимируемой кривой составляет около 9 % от уровня скачка (см. рис. 2.17, $n = 4$). Эффект Гиббса приводит к неустранимой погрешности аппроксимации периодических импульсных сигналов с разрывами первого рода. Для ряда практических приложений наличие эффекта Гиббса вызывает определенные проблемы. Например, в звуковоспроизводящих системах (в частности, в микрофонах и динамиках) это явление называют «звоном» или «дребезгом». При этом каждый резкий согласный или любой другой внезапный звук может сопровождаться коротким свистящим, неприятным для слуха, звуком.

Ряд Фурье может быть применен для представления не только периодических сигналов, но и сигналов конечной длительности. При этом оговаривается временно́й интервал, для которого строится ряд Фурье, а в остальные моменты времени сигнал считается равным нулю. Для расчета коэффициентов ряда такой подход фактически означает *периодическое продолжение* сигнала за границами рассматриваемого интервала.

Отметим, что и природа (например, слух человека) использует принцип гармонического анализа сигналов. Виртуальное преобразование Фурье человек производит всякий раз, когда слышит звук: ухо автоматически выполняет это, представляя звук — колебательное движение распространяющихся звуковых волн в газообразной, жидкой или твердых средах — в виде спектра последовательных значений громкости для тонов различной высоты. Мозг человека автоматически превращает эту информацию в воспринимаемый звук.

**Гармонический синтез.** В теории сигналов наряду со спектральным представлением (гармоническим анализом) сигналов широко используется *гармонический синтез* — получение заданных электрических колебаний сложной формы путем суммирования ряда гармонических составляющих их спектра. По существу, выше и был проведен синтез периодической последовательности прямоугольных импульсов суммой из нескольких гармоник. На практике такие операции осуществляют на компьютере, как это и показано на рис. 2.17, *б*.

## *Спектральное представление непериодических сигналов. Преобразование Фурье*

Метод рядов Фурье допускает определенное обобщение, позволяющее получать спектральные характеристики и непериодических сигналов. Среди них для радиотехники интерес представляют импульсные (одиночные) сигналы.

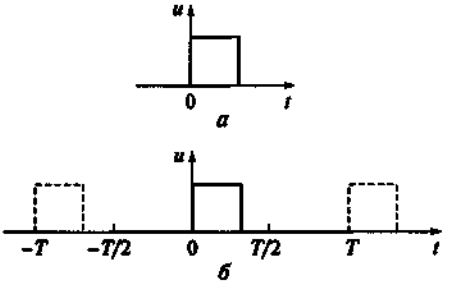



**Рис. 2.18.** Непериодические сигналы:
*а* — одиночный импульс;
*б* — условное периодическое представление

Преобразование Фурье (*Fourier transform*) является инструментом спектрального анализа *непериодических* (импульсных) сигналов (их еще называют сигналами *конечной длительности*, или *финитными*, т. е. *пространственно ограниченными*). Такие сигналы отличны от нуля только на ограниченном промежутке времени; иногда говорят, что сигнал *существует* на конечном временно́м интервале. Очевидно, что сигнал конечной длительности будет иметь и конечную энергию — если только он не содержит разрывов второго рода (с уходящими в бесконечность ветвями функции).

Для иллюстрации перехода от ряда к интегральному преобразованию Фурье применяют не вполне строгий математически, но зато понятный аналитический подход. В теории спектрального представления непериодических импульсных сигналов используют искусственный прием, мысленно (формально) заменяя одиночные сигналы периодическими (как бы превращая одиночный импульсный сигнал в виде отдельного вагончика в бесконечный поезд, составленный из одинаковых частей-вагончиков) с бесконечно большим периодом следования $T \to \infty$ (рис. 2.18). Положим, что некоторая функция $u(t)$ аналитически описывает одиночный импульсный сигнал конечной длительности (рис. 2.18, *а*). Мысленно дополнив его такими же импульсными сигналами, следующими с некоторым интервалом $T$ (штриховые импульсы на рис. 2.18, *б*), получим периодическую последовательность аналогичных импульсов $u_n(t) = u(t \pm nT)$.

Для того чтобы вне искусственно введенного интервала времени $0 \ldots T$ исходный сигнал был равен нулю, необходимо увеличить период повторения этих импульсов. В пределе, при увеличении длительности периода и $T \to \infty$, все импульсы уйдут вправо и влево в бесконечность и периодическая последовательность импульсов $u_n(t)$ вновь станет одиночным импульсом $u(t)$. В этом случае выражения (2.39) и (2.40) сохраняют смысл. Подставив соотношение (2.40) в формулу (2.39), запишем периодическую функцию

$$u_n(t) = \sum_{n=-\infty}^{\infty} \mathrm{C}_n \mathrm{e}^{jn\omega_1 t} = \sum_{n=-\infty}^{\infty} \left[ \frac{1}{T} \int_{-T/2}^{T/2} u(t) \mathrm{e}^{-jn\omega_1 t} dt \right] \mathrm{e}^{jn\omega_1 t}.$$

Так как период следования импульсов $T = 2\pi/\omega_1$, то

$$u_n(t) = \frac{1}{2\pi}\sum_{n=-\infty}^{\infty}\left[\int_{-T/2}^{T/2} u(t)e^{-jn\omega_1 t}dt\right]e^{jn\omega_1 t}\omega_1. \qquad (2.52)$$

Нетрудно заметить, что при увеличении периода следования импульсов $T$ гармоники располагаются ближе друг к другу по частоте (линейный спектр становится все более плотным), а общий уровень спектральных составляющих становится все меньше. При этом вид вычисляемого интеграла (2.40) не меняется. В предельном случае, когда $T \to \infty$, равные расстояния между спектральными линиями уменьшатся настолько, что спектр станет *сплошным*, а амплитуды отдельных спектральных составляющих окажутся бесконечно малыми. При этом частота следования импульсов $\omega_1 = 2\pi/T \to 0$ и превращается в $d\omega$, дискретная переменная $n\omega_1$ — в мгновенную (текущую) частоту $\omega$, а сумма трансформируется в интеграл. Периодическая последовательность импульсов $u_n(t)$ станет одиночным импульсом $u(t)$, и выражение (2.52) запишется в виде

$$u(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} u(t)e^{-j\omega t}dt\right]e^{j\omega t}d\omega. \qquad (2.53)$$

Интеграл в скобках есть комплексная функция частоты. Обозначив его

$$\mathbf{S}(\omega) = \dot{S}(\omega) = S(j\omega) = \int_{-\infty}^{\infty} u(t)e^{-j\omega t}dt, \qquad (2.54)$$

получим

$$u(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathbf{S}(\omega)e^{j\omega t}d\omega. \qquad (2.55)$$

Соотношения (2.54) и (2.55) носят фундаментальный характер в теории сигналов и определяют соответственно *прямое* и *обратное преобразования Фурье* (*direct, inverse Fourier transform*). Они связывают между собой вещественную функцию времени $u(t)$ и комплексную функцию частоты $\mathbf{S}(\omega)$.

Если использовать не угловую частоту $\omega$, а циклическую $f = \omega/(2\pi)$, то формулы прямого (2.54) и обратного (2.55) преобразования Фурье становятся еще более симметричными, отличаясь лишь знаком в показателе экспоненты:

$$\mathbf{S}(f) = \int_{-\infty}^{\infty} u(t)e^{-j2\pi ft}dt; \qquad (2.56)$$

$$u(t) = \int_{-\infty}^{\infty}\mathbf{S}(f)e^{j2\pi ft}df. \qquad (2.57)$$

Известно, что преобразования (2.54) и (2.55) существуют, если анализируемая функция $u(t)$ удовлетворяет условиям Дирихле (по аналогии с периодическим сигналом), к которым добавляется требование *абсолютной интегрируемости* сигнала

$$\int_{-\infty}^{\infty} |u(t)|\,dt < \infty. \qquad (2.58)$$

Однако с привлечением математического аппарата так называемых *обобщенных функций* возможно выполнение интегрального анализа Фурье и для ряда сигналов, не удовлетворяющих этим требованиям (речь идет о неинтегрируемых сигналах). В математике под обобщенной функцией понимают непрерывный функционал. К таким функциям относятся, в частности, δ-функция, единичная функция и т. д.

Итак, прямое преобразование Фурье (2.54) ставит в соответствие сигналу, заданному во времени, его спектральную функцию. При этом осуществляется переход из *временно́й области* (*time domain*) *в частотную область* (*frequency domain*). Преобразование Фурье является взаимно-однозначным, поэтому представление сигнала в частотной области (спектральная функция) содержит ровно столько же информации, сколько и исходный сигнал, заданный во временно́й области. Принципиально важно, что спектральная плотность — комплексная функция частоты, одновременно несущая информацию, как об амплитуде, так и о фазе элементарных гармоник.

Поскольку интеграл Фурье (2.54) содержит непрерывную (сплошную) последовательность спектральных составляющих анализируемого сигнала с бесконечно малыми амплитудами, то функцию $S(\omega)$ называют *спектральной функцией* (часто говорят *спектральная плотность* (*spectral density*) или просто *спектр*; последнее название не совсем корректно). Она характеризует интенсивность сплошного распределения амплитуд гармоник непериодического сигнала вдоль оси частот ω. В этом основное отличие спектральной плотности непериодического сигнала от дискретного спектра периодического сигнала, в котором каждая гармоническая составляющая имеет вполне определенное значение частоты и отстоит от соседней на величину $\omega_1 = 2\pi/T$.

Дискретный спектр периодического сигнала и спектральная плотность непериодического сигнала имеют разные размерности. Размерность амплитудного спектра периодического сигнала совпадает с размерностью самого сигнала — [В] или [А], а размерность спектральной плотности амплитуд определяется отношением размерности сигнала к размерности частоты — [В/Гц] или [А/Гц].

Поскольку анализируемый непериодический сигнал $u(t)$ и его спектральная плотность $S(\omega)$ взаимно-однозначно связаны прямым и обратным преобразованиями Фурье, то последние позволяют аналитически отыскать спектральную плотность по заданной форме сигнала, и наоборот, его форму по полученной спектральной плотности.

В общем случае $S(\omega)$ является комплексной величиной. Как комплексная величина она записывается в виде

$$S(\omega) = |S(\omega)|e^{j\varphi(\omega)} = |S(\omega)|e^{j\varphi(\omega)}, \tag{2.59}$$

где $|S(\omega)| = S(\omega)$, $\varphi(\omega)$ — соответственно модуль и аргумент комплексной величины, т. е. амплитудный и фазовый спектры сигнала.

В математике доказывается, что прямое преобразование Фурье четного сигнала $u(t)$ всегда дает вещественную функцию частоты $\omega$, а нечетного сигнала $u(t)$ — всегда мнимую функцию частоты.

Нетрудно показать, что интеграл

$$\int_{-\infty}^{\infty} u(t)e^{j\omega t}\,dt = S(-\omega) = S^*(\omega) \tag{2.60}$$

представляет собой комплексно-сопряженную спектральную плотность непериодического сигнала.

**Симметрия преобразований Фурье (взаимозаменяемость частоты и времени).** Преобразования Фурье вещественных сигналов в комплексной области не симметричны по $j$. При переходе от прямого преобразования Фурье к обратному преобразованию Фурье необходимо изменение знака при $j$. Вместе с тем в преобразованиях имеется и существенная симметрия. Пусть *четному* сигналу $u(t)$ соответствует *вещественный* спектр $S(\omega)$, который, в свою очередь, будет являться *четной* функцией частоты $\omega$, тогда четному сигналу $S(t)$ должен соответствовать спектр $U(\omega) = 2\pi u(\omega)$. Именно «взаимозаменяемость» аргументов $t$ и $\omega$, входящих в ядро ортонормированного базиса $\exp(\pm j\omega t)$, и подразумевают, говоря о симметрии (дуальности) пары интегральных преобразований Фурье (2.54) и (2.55). Симметрия очевидна, если рассматривать *комплексные сигналы*. Данное свойство удобно использовать в тех случаях, когда по известной паре преобразований Фурье можно найти временну́ю функцию, спектр которой соответствует временно́й функции известного сигнала, и наоборот.

**Пример 2.2.** Задан прямоугольный импульс напряжения, имеющий амплитуду $E$ и длительность $\tau_и$ (рис. 2.19, *а*). Определить его спектральную плотность.

Р е ш е н и е. Поскольку анализируемый прямоугольный импульс расположен на временно́м интервале $-\tau_и/2$, $\tau_и/2$, то, в соответствии с (2.54), получим

$$S(\omega) = \int_{-\tau_и/2}^{\tau_и/2} Ee^{-j\omega t}\,dt = E\int_{-\tau_и/2}^{\tau_и/2}(\cos\omega t - j\sin\omega t)\,dt = E\tau_и\frac{\sin(\omega\tau_и/2)}{\omega\tau_и/2}. \tag{2.61}$$

На рис. 2.19, *б* показан модуль спектральной плотности $|S(\omega)| = S(\omega)$ исследуемого прямоугольного импульса.

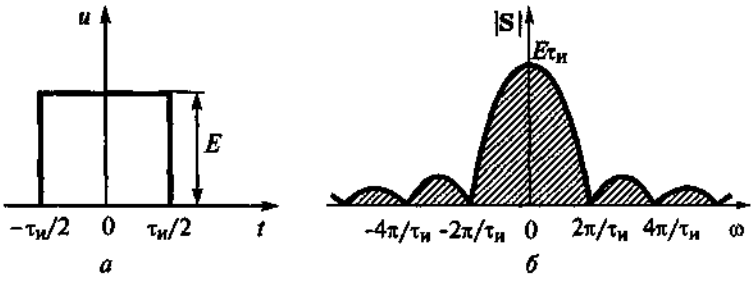



**Рис. 2.19.** Прямоугольный импульс:
*а* — временна́я диаграмма; *б* — модуль спектральной плотности

Сравнив выражения для спектральной плотности одиночного прямоугольного импульса (2.61) и спектра периодической последовательности таких же импульсов (2.46), нетрудно заметить, что модуль спектральной плотности и огибающая (*огибающая* — воображаемая линия, которая действительно «огибает» сигнал по его амплитудным значениям и имеет смысл максимальных значений его мгновенной амплитуды) гармоник дискретного спектра совпадают по форме и отличаются лишь масштабом по оси амплитуд.

В табл. 2.1 приведены графики ряда импульсов и их амплитудных спектров в области положительных частот, из которых видно, что почти все реально применяемые в радиотехнике импульсы ограниченной длительности теоретически имеют бесконечный амплитудный спектр.

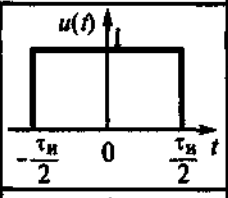
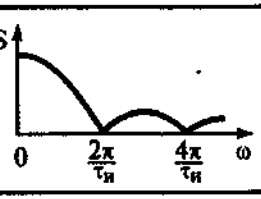
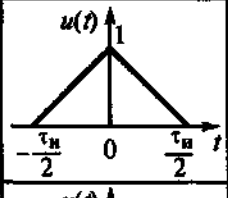
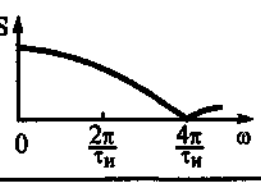
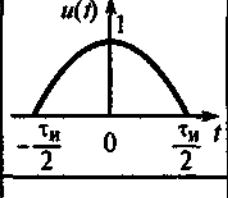
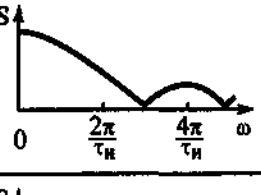
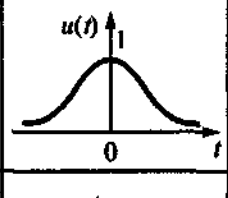
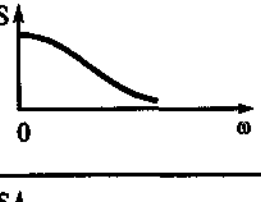
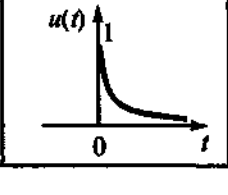
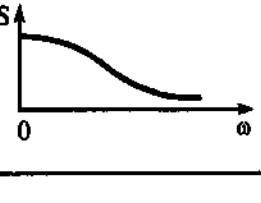
Известный в математике и физике принцип неопределенности (*Uncertainty principle*) В. Гейзенберга (*V. Heisenberg*; 1901–1976; немецкий физик, лауреат Нобелевской премии) применительно к одномерным сигналам можно трактовать так: чем сильнее сигнал $u(t)$ локализован во времени (т. е. чем компактнее «колокол», накрывающий его), тем шире его спектр $S(\omega)$ «размазан» в пространстве частот $\omega$. И наоборот — чем меньше ширина спектра $S(\omega)$, тем больший интервал времени занимает анализируемый сигнал $u(t)$. В общем случае, в рамках теории радиотехнических сигналов принцип неопределенности Гейзенберга означает, что невозможно одновременно с заданной степенью точности зафиксировать частоту сигнала и время его возникновения. И тот и другой параметры фиксируются с некоторой ошибкой, т. е. истинное значение параметров сигнала находится внутри некоторого «окна». Если считать это окно прямоугольным, то его площадь будет равна произведению частоты и времени, и она величина постоянная. Поэтому улучшение разрешения по частоте сопровождается ухудшением разрешения по времени и наоборот.



В. Гейзенберг

У относительно гладких функций спектр быстро убывает (с ростом номера коэффициенты резко стремятся к нулю). Для относительно «изрезанных» функций спектр убывает медленно, поскольку для представления разрывов и «изломов» функции нужны гармоники с большими частотами.

Таблица 2.1. **Спектры некоторых распространенных импульсов**

| Сигнал $u(t)$ | | Спектральная плотность $S=\|S\|$ | Амплитудный спектр |
|---|---|---|---|
| $u(t)=\begin{cases}1, & -\frac{\tau_и}{2}\le t\le\frac{\tau_и}{2};\\ 0, & t<-\frac{\tau_и}{2};\ t>\frac{\tau_и}{2}\end{cases}$ |  | $\tau_и\dfrac{\sin\frac{\omega\tau_и}{2}}{\frac{\omega\tau_и}{2}}$ |  |
| $u(t)=\begin{cases}1-\frac{\lvert t\rvert}{\tau_и/2}, & \lvert t\rvert\le\frac{\tau_и}{2};\\ 0, & \lvert t\rvert>\frac{\tau_и}{2}\end{cases}$ |  | $\frac{\tau_и}{2}\left(\dfrac{\sin\frac{\omega\tau_и}{4}}{\frac{\omega\tau_и}{4}}\right)^2$ |  |
| $u(t)=\begin{cases}\cos\omega_0 t, & -\frac{\tau_и}{2}\le t\le\frac{\tau_и}{2};\\ 0, & t<-\frac{\tau_и}{2};\ t>\frac{\tau_и}{2}\end{cases}$ |  | $\frac{4\tau_и}{\pi}\dfrac{\cos\frac{\omega\tau_и}{2}}{1-\left(\frac{\omega\tau_и}{2}\right)^2}$ |  |
| $u(t)=e^{-(\alpha t)^2}$ |  | $\frac{\sqrt{\pi}}{\alpha}e^{-\left(\frac{\omega}{2\alpha}\right)^2}$ |  |
| $u(t)=\begin{cases}e^{-\alpha t}, & t\ge 0;\\ 0, & t<0\end{cases}$ $\alpha>0$ |  | $\frac{1}{\alpha+j\omega}$ |  |

В радиотехнике шириной спектра считают эффективную область частот $F_э$, в пределах которой сконцентрировано 90...95 % энергии сигнала. Для колокольного и экспоненциального импульсов, имеющих теоретически бесконечную длительность, для удобства расчетов также вводят понятие «эффективной длительности сигнала» $T_э$, подразумевая под этим интервал времени, в пределах которого сосредоточена основная доля его энергии.

### Основные свойства преобразований Фурье

Для практического использования важна связь между рядом преобразований исходжного сигнала и соответствующими этим преобразованиям изменениями его спектральной плотности.

**1. Сложение, усиление и ослабление сигналов (теорема линейности).** К линейным операциям относят сложение, усиление и ослабление сигналов, поэтому к ним применимо свойство линейности. Если имеется совокупность детерминированных сигналов

$$u_1(t),\ u_2(t),\ldots,\ u_i(t),\ldots,\ u_N(t),$$

обладающих спектральными плотностями

$$\mathbf{S}_1(\omega),\ \mathbf{S}_2(\omega),\ \ldots,\ \mathbf{S}_i(\omega),\ \ldots,\ \mathbf{S}_N(\omega),$$

то суммарному (разностному) значению сигналов

$$u_c(t) = u_1(t) + u_2(t) + \ldots + u_i(t) + \ldots + u_N(t)$$

соответствует сумма (разность) их спектральных плотностей

$$\mathbf{S}_c(\omega) = \mathbf{S}_1(\omega) + \mathbf{S}_2(\omega) + \ldots + \mathbf{S}_i(\omega) + \ldots + \mathbf{S}_N(\omega).$$

Данная теорема имеет элементарное доказательство: достаточно в прямое преобразование Фурье (2.54) подставить сумму исходных сигналов.

В общем виде теорему линейности записывают следующим образом:

$$a_i \sum_{i=0}^{N} u_i(t) \leftrightarrow a_i \sum_{i=0}^{N} \mathbf{S}_i(\omega), \tag{2.62}$$

где $a_i$ — произвольные числовые коэффициенты; $i = 0, 1, \ldots, N$.

**2. Сдвиг сигнала во времени (теорема запаздывания).** Пусть сигнал $u_1(t)$ со спектральной плотностью $\mathbf{S}_1(\omega)$ задержан на некоторое время $t_c$. В этом случае $u_2(t) = u_1(t - t_c)$, и спектральная плотность задержанного сигнала в соответствии с прямым преобразованием Фурье (2.54) имеет вид:

$$\mathbf{S}_2(\omega) = \int_{-\infty}^{\infty} u_2(t) e^{-j\omega t} dt = \int_{-\infty}^{\infty} u_1(t - t_c) e^{-j\omega t} dt.$$

Введя новую переменную интегрирования $\tau = t - t_c$, получим

$$\mathbf{S}_2(\omega) = \int_{-\infty}^{\infty} u_1(\tau) e^{-j\omega\tau} e^{-j\omega t_c} d\tau = \mathbf{S}_1(\omega) e^{-j\omega t_c}. \tag{2.63}$$

Итак, сдвиг исходного сигнала во времени на некоторый интервал $t_c$ приводит к тому, что спектр задержанного сигнала оказывается равен $\mathbf{S}_1(\omega)$, умноженному на комплексную экспоненту $e^{-j\omega t_c}$. Амплитудный же спектр сигнала не меняется (ведь модуль такой комплексной экспоненты равен 1; к тому же здравый смысл подсказывает, что соотношение между амплитудами спектральных составляющих из-за сдвига сигнала во времени измениться не должно). При этом фазовый спектр приобретает дополнительное слагаемое $-\omega t_c$, линейно зависящее от частоты. На практике сдвиг исходного сигнала во времени осущест-

вляют при аудио- и видеозаписи на различные типы носителей. Теорема запаздывания показывает, что сколь долго (теоретически) ни хранилась бы такая запись, спектр (и форма) сигнала теоретически не претерпит изменений.

Заметим, что если в результате какого-либо преобразования сигнала его *спектр умножается* на некоторую функцию, не зависящую от преобразуемого сигнала, это означает, что данное преобразование может быть выполнено линейной системой с постоянными параметрами (см. гл. 4).

**3. Смещение спектра сигнала (теорема смещения).** Смысл данной теоремы заключается в следующем: если $S_1(\omega)$ — спектральная плотность сигнала $u_1(t)$, то спектральная плотность $S_2(\omega + \Omega)$, полученная путем сдвига исходного спектра по оси частот на величину $\Omega$, соответствует сигналу $u_2(t) = u_1(t)e^{-j\Omega t}$. Действительно, согласно (2.54):

$$S_2(\omega) = \int_{-\infty}^{\infty} u_2(t)e^{-j\omega t}dt = \int_{-\infty}^{\infty} u_1(t)e^{-j(\omega+\Omega)t}dt = S_1(\omega + \Omega). \qquad (2.64)$$

Это преобразование спектра импульсного сигнала применяют в системах передачи информации и связи, либо при переносе спектра сигнала из одной полосы частот в другую, либо при модуляции. Формула (2.61) показывает, что в результате таких преобразований спектр сигнала смещается на величину $\Omega$, равную частоте сдвига.

**4. Изменение масштаба времени.** Предположим, что в исходном сигнале $u_1(t)$ изменен масштаб времени таким образом, что аргумент $t$ умножен на некоторый постоянный коэффициент $b$ и $u_2(t) = u_1(bt)$. Если $b > 1$, то происходит «сжатие» исходного сигнала; если же $0 < b < 1$ — исходный сигнал «растягивается» во времени. Докажем это положение.

Спектральная плотность измененного во времени сигнала

$$S_2(\omega) = \int_{-\infty}^{\infty} u_2(t)e^{-j\omega t}dt = \int_{-\infty}^{\infty} u_1(bt)e^{-j\omega t}dt.$$

Введя новую переменную $\tau = bt$, получим

$$S_2(\omega) = \frac{1}{b}\int_{-\infty}^{\infty} u_1(t)e^{-j\omega\tau/b}d\tau, \qquad (2.65)$$

откуда

$$S_2(\omega) = \frac{1}{b}S_1\left(\frac{\omega}{b}\right). \qquad (2.66)$$

Увеличение длительности импульсного сигнала любой формы в $b$ раз сопровождается сжатием ширины его спектра во столько же раз, и наоборот, уменьшение длительности о сигнала приводит к расширению его спектра.

**5. Спектр произведения сигналов (теорема о свертке спектров).** Прежде чем определить данный спектр, введем важное для теории сигналов понятие «свертки» двух функций. Известно, что скалярное произведение двух функций $f(t)$ и $h(t)$ определяется формулой

$$u(t)=\int_{-\infty}^{\infty} f(\tau)h(t-\tau)d\tau = f(t)*h(t). \qquad (2.67)$$

Это соотношение имеет фундаментальное значение в теоретической радиотехнике. Интеграл (2.67) в математике и теории цепей называют *сверткой* (от англ. — *convolution*) двух функций или сигналов (знак * в формуле — символ свертки).

Пусть сигналы $f(t)$ и $h(t)$ имеют спектральные плотности $\mathbf{F}(\omega)$ и $\mathbf{H}(\omega)$ соответственно. Тогда их произведение $u(t) = f(t)h(t)$ будет характеризовать спектральная плотность

$$\begin{aligned}\mathbf{S}(\omega)&=\int_{-\infty}^{\infty}u(t)e^{-j\omega t}dt=\int_{-\infty}^{\infty}f(t)h(t)e^{-j\omega t}dt=\int_{-\infty}^{\infty}h(t)e^{-j\omega t}\frac{1}{2\pi}\int_{-\infty}^{\infty}\mathbf{F}(\tau)e^{j\tau t}d\tau dt=\\&=\frac{1}{2\pi}\int_{-\infty}^{\infty}\mathbf{F}(\tau)\left\{\int_{-\infty}^{\infty}h(t)e^{-j(\omega-\tau)t}dt\right\}d\tau=\frac{1}{2\pi}\int_{-\infty}^{\infty}\mathbf{F}(\tau)\mathbf{H}(\omega-\tau)d\tau=\frac{1}{2\pi}\mathbf{F}(\omega)*\mathbf{H}(\omega).\end{aligned} \qquad (2.68)$$

Здесь при выводе формулы (2.68) сигнал $f(t)$ выражен через свою спектральную плотность $\mathbf{F}(\omega)$ с заменой переменной $\omega$ на $\tau$.

Итак, согласно (2.68) спектральная плотность произведения двух сигналов есть свертка их спектральных плотностей (умноженная на коэффициент $1/(2\pi)$), т. е. свертка, осуществленная уже́ в частотной области. Данное соотношение имеет чрезвычайно важное значение в теоретической радиотехнике и теории связи. Оно связывает спектральный и временно́й подходы к анализу импульсных сигналов и служит для целей исследования прохождения подобных сигналов через линейные и линейные параметрические цепи.

Нетрудно убедиться, что операция свертки коммутативна, т. е. допускает изменение в порядка следования преобразуемых функций

$$\mathbf{S}(\omega)=\frac{1}{2\pi}\mathbf{F}(\omega)*\mathbf{H}(\omega)=\frac{1}{2\pi}\mathbf{H}(\omega)*\mathbf{F}(\omega). \qquad (2.69)$$

Теорема Рэлея и равенство Парсеваля. Приняв в (2.68) значение частоты $\omega = 0$, приходим к выводу известной в математике теоремы (*обобщенной формулы*) Дж. Рэлей (*J. Rayleigh*; 1842–1919; английский физик и математик; лауреат Нобелевской премии 1904 г.) для сигналов

$$\int_{-\infty}^{\infty}f(t)h(t)dt=\frac{1}{2\pi}\int_{-\infty}^{\infty}\mathbf{F}(\tau)\mathbf{H}(-\tau)d\tau=\frac{1}{2\pi}\int_{-\infty}^{\infty}\mathbf{F}(\omega)\mathbf{H}(-\omega)d\omega=\frac{1}{2\pi}\int_{-\infty}^{\infty}\mathbf{F}(\omega)\mathbf{H}^{*}(\omega)d\omega. \qquad (2.70)$$


М. Парсеваль


Д. Рэлей

Здесь учтено соотношение (2.60), согласно которому $\mathbf{H}(-\omega) = \mathbf{H}^*(\omega)$. Легко запоминающаяся трактовка формулы (2.70) такова: скалярное произведение двух непрерывных сигналов с точностью до коэффициента $1/(2\pi)$ пропорционально скалярному произведению их спектральных плотностей. Формула Рэлея относится к классу обобщенных функций и обладает важным положением, касающимся спектральных свойств ряда неинтегрируемых сигналов.

При $f(t) = h(t) = u(t)$ из теоремы Рэлея вытекает известное равенство Парсеваля (Марк Парсева́ль; *Marc Parseval*; 1755–1836; французский математик)

$$\int_{-\infty}^{\infty} u^2(t)dt = \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathbf{S}(\omega)\mathbf{S}^*(\omega)d\omega = \frac{1}{2\pi}\int_{-\infty}^{\infty}\left|\mathbf{S}(\omega)\right|^2 d\omega = \frac{1}{\pi}\int_{0}^{\infty}\left|\mathbf{S}(\omega)\right|^2 d\omega. \qquad (2.71)$$

**6. Умножение сигнала на гармоническую функцию.** Умножим исходный непрерывный сигнал $u(t)$, спектральная плотность $\mathbf{S}(\omega)$ которого известна, на гармоническую функцию единичной амплитуды (для упрощения примем начальную фазу гармонического сигнала равной нулю)

$$f(t) = u(t)\cos\omega_0 t. \qquad (2.72)$$

Посмотрим, что произошло со спектром при таком преобразовании:

$$\mathbf{F}(\omega) = \int_{-\infty}^{\infty} u(t)\cos\omega_0 t\, e^{-j\omega t}dt = \int_{-\infty}^{\infty} u(t)\frac{e^{j\omega_0 t} + e^{-j\omega_0 t}}{2}e^{-j\omega t}dt =$$

$$= \frac{1}{2}\int_{-\infty}^{\infty} u(t)e^{-j(\omega-\omega_0)t}dt + \frac{1}{2}\int_{-\infty}^{\infty} u(t)e^{-j(\omega+\omega_0)t}dt = \frac{1}{2}\mathbf{S}(\omega-\omega_0) + \frac{1}{2}\mathbf{S}(\omega+\omega_0). \qquad (2.73)$$

Итак, спектр исходного сигнала при его умножении на гармоническую функцию «раздвоился» — распался на два слагаемых вдвое меньшего уровня, чем исходный (1/2 перед каждым из слагаемых), смещенных на частоту сигнала $\pm\omega_0$ соответственно влево $(\omega - \omega_0)$ и вправо $(\omega + \omega_0)$ по оси частот. Несложно показать, что если в гармоническом сигнале имеется начальная фаза $\varphi_0$, то при первом слагаемом в (2.73) будет множитель $e^{j\varphi_0}$, а при втором — $e^{-j\varphi_0}$.

### *Спектры некоторых неинтегрируемых сигналов*

При введении понятия преобразования Фурье были указаны возможности его применимости: выполнение условий Дирихле и абсолютная интегрируемость (сходимость) сигнала (2.58). Ряд широко применяемых в радиотехнике сигналов не удовлетворяют этим условиям, поэтому их прямое преобразование Фурье осуществить невозможно.


П. Дирак

Вместе с тем, используя свойства обобщенных функций, можно применить преобразование Фурье и к таким сигналам, получив при этом вполне осмысленный и практически полезный результат.

В математике известны и широко используются три специфических функции: дельта-функция, гармонический сигнал и функция единичного скачка. Некоторые свойства этих функций позволяют устранить отмеченное препятствие. Правда, при этом оказывается, что соответствующие спектральные плотности будут уже не обычными, классическими, а *обобщенными* функциями. Дельта-функция (рис. 2.20) относится к обобщенным функциям и является распространенным *тестовым сигналом*, который применяется для анализа характеристик цепей.

**Дельта-функция и ее спектр.** Рассмотрим теоретическую модель бесконечно короткого импульса с бесконечно большой амплитудой (ее изображают жирной стрелкой, рис. 2.20, *а*), аналитически определяемого как

$$\delta(t)=\begin{cases}\infty, & t=0;\\ 0, & t\neq 0.\end{cases} \tag{2.74}$$

Площадь такого импульса всегда равна единице:

$$\int_{-\infty}^{\infty}\delta(t)dt=1. \tag{2.75}$$

Функцию $\delta(t)$ называют *дельта-функцией, единичным импульсом, функцией Дирака* (*delta-function, Dirac function*; предложена Полем Дираком; Paul Dirac; 1902–1984; английский физик).

Поскольку интеграл от дельта-функции (2.75) дает безразмерную единицу, размерность дельта-функции обратна размерности ее аргумента. Например, дельта-функция времени имеет размерность 1/с, т. е. размерность циклической частоты [Гц].

При сдвиге дельта-функции по оси времени на некоторый интервал $t_0$ (см.

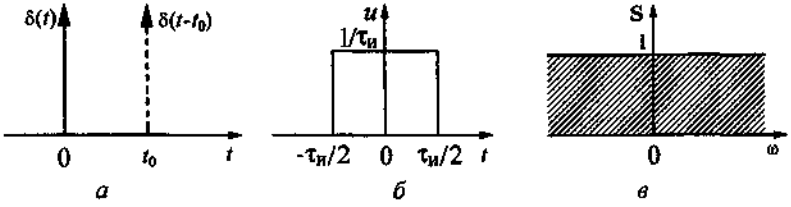



**Рис. 2.20.** Дельта-функция:
*а* — графическое представление;
*б* — дельта-функция в виде прямоугольного импульса; *в* — спектральная плотность

рис. 2.20, *а*) определения (2.74) и (2.75) можно записать в более общей форме:

$$\delta(t-t_0)=\begin{cases}\infty, & t=t_0;\\ 0, & t\neq t_0;\end{cases} \tag{2.76}$$

$$\int_{-\infty}^{\infty}\delta(t-t_0)dt=1. \tag{2.77}$$

Разумеется, сигнал в виде дельта-функции невозможно реализовать физически. Однако теоретически дельта-функцию можно рассматривать как предел, к которому стремится прямоугольный импульс длительностью $\tau_и$ и амплитудой $1/\tau_и$ при $\tau_и \to 0$ (рис. 2.20, *б*).

Дельта-функция обладает важнейшим свойством, благодаря которому она получила широкое применение в математике, радиотехнике, теории связи и т. д. Пусть имеется непрерывная функция (аналоговый сигнал) времени $f(t)$. Тогда, согласно формулам (2.76) и (2.77), справедливо следующее соотношение:

$$\int_{-\infty}^{\infty}f(t)\delta(t-t_0)dt=f(t_0)\int_{-\infty}^{\infty}\delta(t-t_0)dt=f(t_0). \tag{2.78}$$

Соотношение (2.78) становится понятным, если учесть, что по определению функция $\delta(t-t_0)$ будет равна нулю на всей оси времени, кроме точки $t=t_0$. Это позволяет сделать интервал интегрирования бесконечно малым, включающим в себя точку $t_0$. В этом интервале функция $f(t)$ принимает единственное постоянное значение $f(t_0)$ в точке $t=t_0$, которое можно вынести за знак интеграла. Соотношение (2.78) характеризует *фильтрующее* (*выделяющее*, или *стробирующее*, от слова «строб» — короткий прямоугольный импульс, применяемый в радиолокации для выделения сегментов колебаний) свойство дельта-функции.

Спектральную плотность дельта-функции $\delta(t-t_0)$ найдем с помощью формулы (2.54):

$$S(\omega)=\int_{-\infty}^{\infty}\delta(t-t_0)e^{-j\omega t}dt.$$

Используя фильтрующее свойство дельта-функции (2.78), получим

$$S(\omega)=e^{-j\omega t_0}\int_{-\infty}^{\infty}\delta(t-t_0)dt=e^{-j\omega t_0}. \tag{2.79}$$

Из (2.79) следует, что при $t_0=0$ спектральная плотность дельта-функции

$$S(\omega)=1.$$

Итак, теоретически дельта-функция имеет равномерный (сплошной и бесконечный) спектр с единичной амплитудой на всех частотах (рис. 2.20, *в*).

Физически интерпретировать свойства дельта-функции просто. В момент возникновения дельта-функции (при $t = 0$) все элементарные гармонические составляющие ее спектра в бесконечной полосе частот складываются *когерентно* (синфазно), поскольку в соответствии с формулой (2.79) спектральная плотность дельта-функции вещественна. Поэтому в момент времени $t = 0$ наблюдают бесконечно большую амплитуду импульса дельта-функции.

Следует обязательно иметь в виду, что правая часть равенства (2.79) является размерной единицей — это единичная площадь импульса. Если дельта-функция представляет собой импульс напряжения, то размерность спектральной плотности $S(\omega)$ — вольт·секунда (В·с).

Понятие дельта-функции широко используют в радиотехнике и теории связи при исследовании воздействия очень коротких импульсов напряжения длительностью $\tau_и$ на линейные электрические цепи. При этом не обязательно, чтобы теоретически длительность реального импульса была бесконечно мала, а амплитуда бесконечно велика. Оказывается на практике достаточным является условие, чтобы длительность реального импульса была много меньше периода собственных свободных колебаний исследуемой электрической цепи.

Дельта-функцию можно представить в виде обратного преобразования Фурье (2.55) от ее спектральной плотности $S_\delta(\omega) = \delta(\omega) = 1$:

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

Учитывая условие дуальности частоты $\omega$ и времени $t$, последнее выражение можно записать следующим образом:

$$\delta(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{j\omega t} dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-j\omega t} dt. \qquad (2.80)$$

Перемена знака в показателе степени экспоненты в этом случае не влияет на значение интеграла (вследствие взаимозаменяемости частоты и времени).

Гармонический сигнал и его спектр. Найдем спектральную плотность гармонического (косинусоидального) сигнала единичной амплитуды

$$u(t) = \cos\omega_0 t.$$

Подставив в прямое преобразование Фурье (2.54) заданный сигнал и воспользовавшись уже упоминавшейся формулой Эйлера (2.37), находим

$$S(\omega) = \int_{-\infty}^{\infty} \cos\omega_0 t e^{-j\omega t} dt = 0{,}5\int_{-\infty}^{\infty} e^{-j(\omega-\omega_0)t} dt + 0{,}5\int_{-\infty}^{\infty} e^{-j(\omega+\omega_0)t} dt.$$

На основании формулы (2.80) последнее соотношение можно записать как

$$S(\omega) = \pi[\delta(\omega - \omega_0) + \delta(\omega + \omega_0)]. \qquad (2.81)$$

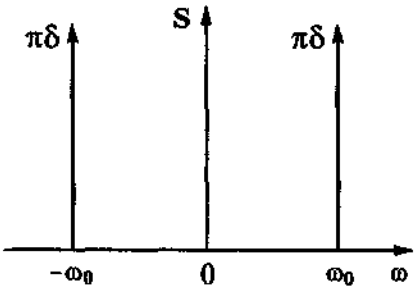



**Рис. 2.21.** Спектр гармонического сигнала

Как и следовало ожидать, гармоническому (в данном случае косинусоидальному) сигналу с единичной амплитудой и частотой $\omega_0$ соответствует дискретный спектр, состоящий лишь из двух линий бесконечно большой амплитуды в виде дельта-функций (с множителями $\pi$), расположенных симметрично относительно нуля на частотах $-\omega_0$ и $\omega_0$ (рис. 2.21).

По аналогии со спектральной плотностью косинусоидального сигнала можно показать, что синусоидальному сигналу $u(t) = \sin\omega_0 t$ отвечает очень похожая спектральная плотность

$$S(\omega) = \pi[\delta(\omega - \omega_0) - \delta(\omega + \omega_0)]. \qquad (2.82)$$

Знак минус в формуле (2.82) для спектральной плотности появляется вследствие нечетности функции синусоидального сигнала относительно оси ординат.

**Экспоненциальный импульс и его спектр.** Экспоненциальный импульс принято относить к радиотехническим сигналам с «полубесконечной» длительностью (рис. 2.22).

При единичной амплитуде экспоненциального импульса (рис. 2.22, *а*) имеем такую аналитическую запись:

$$u(t) = \begin{cases} 0, & t < 0; \\ e^{-\alpha t}, & t \geq 0, \end{cases} \qquad (2.83)$$

где $\alpha > 0$ — вещественный параметр.

Прямым вычислением по (2.54) и (2.83) находим

$$S(\omega) = \int_{-\infty}^{\infty} e^{-\alpha t} e^{-j\omega t} dt = \int_{-\infty}^{\infty} e^{-(\alpha + j\omega)t} dt = -\frac{1}{\alpha + j\omega} e^{-(\alpha + j\omega)t} \Bigg|_0^{\infty}.$$

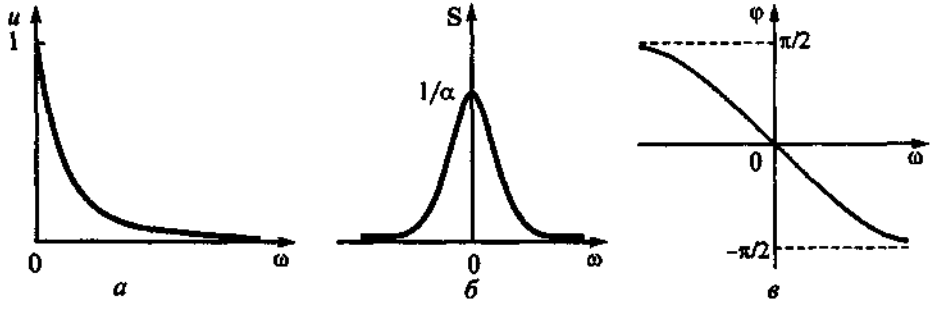



**Рис. 2.22.** Экспоненциальный импульс:
*а* — графическое представление; *б* — спектральная плотность; *в* — фазовый спектр

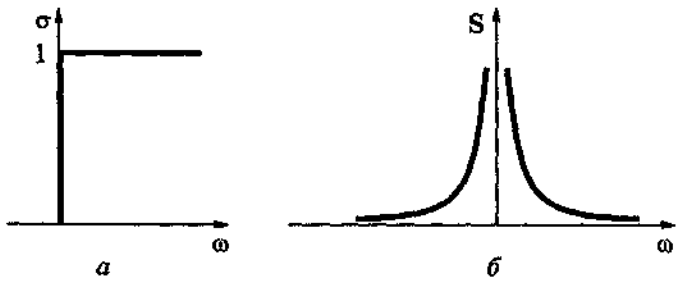



**Рис. 2.23.** Единичная функция:
*а* — графическое представление; *б* — спектральная плотность

Подстановка пределов интегрирования в формулу (2.83) дает следующий результат (рис. 2.22, *б*)

$$S(\omega) = \frac{1}{\alpha + j\omega}. \qquad (2.84)$$

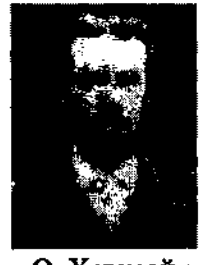
Можно показать, что фазовый спектр рассмотренного экспоненциального импульса (рис. 2.22, *в*) описывается выражением

$$\varphi(\omega) = -\operatorname{arctg}(\omega/\alpha).$$

**Единичная функция и ее спектр.** Рассмотрим еще один элементарный электрический сигнал (рис. 2.23), используемый для анализа радиотехнических цепей и описываемый обобщенной функцией. Несколько упрощенное аналитическое выражение единичной функции (рис. 2.23, *а*) принято в радиотехнике и теории электрических цепей записывать следующим образом:

$$\sigma(t) = \begin{cases} 0, & t < 0; \\ 1, & t \geq 0. \end{cases} \qquad (2.85)$$

Функцию $\sigma(t)$ называют *единичной функцией, функцией включения* (*unit function, unit step*) или *функцией Оливера Хевисайда* (*Oliver Heaviside*; английский физик; 1850–1925). Единичную функцию удобно использовать при создании математических моделей сигналов конечной длительности. Простейшим примером является аналитическое формирование прямоугольного импульса.


О. Хевисайд

Сравнив формулы (2.83) и (2.85), замечаем, что функцию включения легко получить путем предельного перехода из экспоненциального импульса при $\alpha \to 0$

$$\sigma(t) = \begin{cases} 0, & t < 0; \\ \lim\limits_{\alpha \to 0} e^{-\alpha t}, & t \geq 0. \end{cases} \qquad (2.86)$$

Спектральную плотность функции включения определим, выполнив предельный переход в спектральной плотности экспоненциального импульса

$$S(\omega) = \lim_{\alpha \to 0} \frac{1}{\alpha + j\omega} = \lim_{\alpha \to 0} \frac{\alpha}{\alpha^2 + \omega^2} - j \lim_{\alpha \to 0} \frac{\omega}{\alpha^2 + \omega^2}.$$

При $\alpha = 0$ первое слагаемое в правой части этой формулы равно нулю на всех частотах, кроме $\omega = 0$, где оно обращается в бесконечность. Площадь же под кривой функции $\alpha/(\alpha^2 + \omega^2)$ равна постоянной величине

$$\int_{-\infty}^{\infty} \frac{\alpha}{\alpha^2 + \omega^2} d\omega = \int_{-\infty}^{\infty} \frac{1}{\alpha^2 + \omega^2} d\left(\frac{\omega}{\alpha}\right) = \pi$$

независимо от значения $\alpha$.

При $\alpha \to 0$ пределом первого слагаемого есть функция $\pi\delta(\omega)$, второго — $1/(j\omega)$. Значит, спектральная плотность единичной функции (рис. 2.23, *б*):

$$S(\omega) = \pi\delta(\omega) - j/\omega. \qquad (2.87)$$

**Постоянный сигнал и его спектр.** Простейшим неинтегрируемым сигналом является постоянное напряжение (рис. 2.24). Спектральную плотность постоянного напряжения единичной амплитуды (рис. 2.24, *а*) можно легко определить, приравняв к нулю в формуле (2.81) для спектральной плотности косинусоидального сигнала частоту $\omega_0$. В результате получим

$$S(\omega) = 2\pi\delta(\omega) = 2\pi\delta(0). \qquad (2.88)$$

Физический смысл формулы (2.88) прост — постоянный во времени сигнал имеет единственную спектральную составляющую (в виде дельта-функции с коэффициентом $2\pi$), расположенную на нулевой частоте (рис. 2.24, *б*).

Особенности преобразований Фурье. Это преобразование неприменимо к ряду сложных сигналов, содержащих множество конечных скачков на ограниченном интервале времени. Однако в общем случае ряд Фурье всегда сходится, если анализируемый сигнал представляет результат реального физического измерения. Если исходная функция задает значение для каждого действительного числа, ее можно разложить на гармоники всех возможных частот; эти функции объединяют посредством записи и вычисления интеграла Фурье.

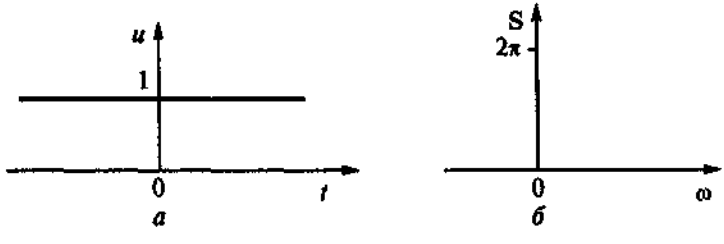



**Рис. 2.24.** Постоянный сигнал:
*а* — графическое представление; *б* — спектральная плотность

Независимо от способа реализации преобразования Фурье для каждой составляющей необходимо указать два параметра. Это могут быть амплитуда и частота, однако ту же роль могут играть и другие пары параметров. Их значения можно выразить в виде одного комплексного числа. Таким представлением пользуются очень широко, так как оно позволяет привлечь математический аппарат алгебры комплексных переменных.

Преобразование Фурье и теория функций комплексных переменных стали необходимыми в численных расчетах при проектировании электрических цепей. При этом представление исходной функции (сигнала) комплексным преобразованием Фурье дает ряд преимуществ при вычислениях. Типичным примером может служить задача расчета тока в заданной электрической цепи при известном значении приложенного к ней напряжения. Если решать эту задачу прямым методом, то приходится иметь дело со сложным дифференциальным уравнением, связывающим функции напряжения и тока. В то же время преобразования Фурье от функций напряжения и тока можно связать алгебраическим уравнением, которое имеет тривиальное решение.

В последние годы использование преобразования Фурье сводится в основном к поиску эффективных способов перехода от сигналов к их спектральной плотности и обратно. Чтобы вычислить интеграл Фурье и произвести преобразование, можно воспользоваться аналитическими методами. И хотя при применении этих методов в повседневной практике могут возникнуть определенные трудности, многие интегралы Фурье уже вычислены и представлены в таблицах и справочниках. Выбор рациональной системы ортогональных функций для спектрального анализа радиотехнических сигналов зависит от поставленной задачи. Из множества задач спектрального анализа наиболее важной является задача приближенного разложения сложных сигналов, при которой требуемая точность обеспечивается при минимуме членов ряда.

Для разложения непрерывных радиотехнических сигналов применяют полиномы и функции Лагерра, Лежандра, Чебышева, Эрмита и др. Для представления ступенчатых и импульсных сигналов используются кусочно-постоянные функции Уолша, Адамара, Радемахера и Хаара. Определенную пользу приносит преобразование Хартли.

В последние годы широко применяют базисные функции типа вейвлетов и анализ на основе фрактальных моделей.

### *Преобразования Хартли в спектральном анализе сигналов*

Одним из недостатков представления сигналов с помощью преобразования Фурье — использование в нем комплексных чисел. Обработку же ряда сигналов в виде последовательности вещественных данных желательно осуществлять в области действительных чисел с помощью взаимно симметричных прямого и обратного преобразований. Поэтому вызывает интерес преобразование сигналов на основе гармонического анализа, свободного от указанных недостатков, но не потерявшего при этом такого свойства, как представление сигналов элементарными функциями. Этим свойством обладает преобразование Хартли (*Hartley*

*Transformation*), которое, несмотря на некоторые преимущества перед преобразованием Фурье, до сих пор не получило практического применения.

Р. Хартли еще в 1942 г. предложил пару интегральных преобразований сигналов — прямое и обратное, основанных на специально введенной им функции

$$\mathrm{cas}t = \sin t + \cos t. \tag{2.89}$$

Для сигнала $u(t)$ прямое и обратное преобразования Хартли имеют вид:

$$H(\omega) = \int_{-\infty}^{\infty} u(t)\mathrm{cas}\omega t dt; \tag{2.90}$$

$$u(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} H(\omega)\mathrm{cas}\omega t d\omega. \tag{2.91}$$

Представляя сигнал как сумму четной и нечетной составляющих, запишем прямое преобразование Хартли в виде

$$H(\omega) = H_{\text{ч}}(\omega) + H_{\text{н}}(\omega). \tag{2.92}$$

Если исследуемый сигнал четная функция $u_{\text{ч}}(t)$, то из (2.90) находим

$$H_{\text{ч}}(\omega) = 2\int_{0}^{\infty} u_{\text{ч}}(t)\cos\omega t dt. \tag{2.93}$$

Для сигнала в виде нечетной функции $u_{\text{н}}(t)$ из соотношения (2.90) имеем следующую форму преобразования Хартли:

$$H_{\text{н}}(\omega) = 2\int_{0}^{\infty} u_{\text{н}}(t)\sin\omega t dt. \tag{2.94}$$

Сравнение преобразований Хартли $H(\omega)$ и Фурье $\mathrm{S}(\omega)$ позволяет установить взаимосвязь между ними. После сопоставлений формул (2.54) и (2.90) получим

$$H(\omega) = [\mathrm{S}(\omega) + \mathrm{S}(-\omega)]/2 - j[\mathrm{S}(\omega) - \mathrm{S}(-\omega)]/2 = \mathrm{Re}[\mathrm{S}(\omega)] - \mathrm{Im}[\mathrm{S}(\omega)]; \tag{2.95}$$

$$\mathrm{S}(\omega) = [H(\omega) + H(-\omega)]/2 + j[H(\omega) - H(-\omega)]/2 = H_{\text{ч}}(\omega) - jH_{\text{н}}(\omega). \tag{2.96}$$

Итак, преобразование Хартли определяется как разность вещественной и мнимой составляющих преобразования Фурье; преобразование Фурье равно разности четной составляющей преобразования Хартли и его нечетной составляющей, умноженной на $j$.

Из (2.95) следует, что для сигнала, описываемого четной функцией,

$$H(\omega) = \mathrm{S}(\omega), \tag{2.97}$$

а для сигнала, описываемого нечетной функцией,

$$H(\omega) = -j\mathrm{S}(\omega). \tag{2.98}$$

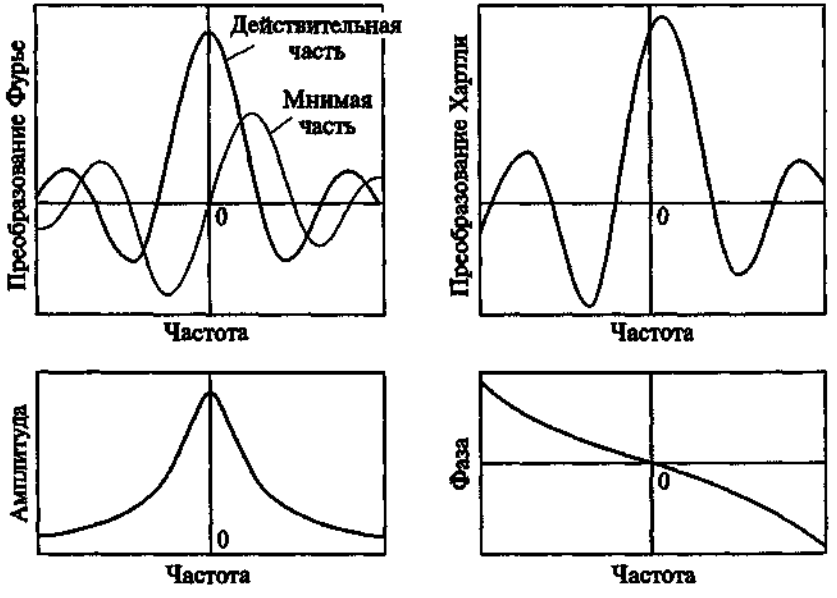



**Рис. 2.25.** Сравнение преобразований экспоненциального импульса по Фурье и Хартли

Преобразование Хартли сходно с преобразованием Фурье и может применяться для спектрального анализа сигналов. В отличие от преобразования Фурье, отображающего вещественные функции в комплексную область и несимметричного по $j$ (происходит изменение знака при переходе от прямого к обратному преобразованию), преобразование Хартли осуществляет прямое и обратное преобразования только в вещественной области с использованием действительной суммы синусоидальных функций и обладает указанной симметрией. Преобразование Хартли, как и преобразование Фурье, трансформирует функции времени в функции частоты, содержащие информацию об амплитуде и фазе.

На рис. 2.25 проведено графическое сравнение спектрального преобразования экспоненциального импульса (см. рис. 2.22, *а*) по Фурье и Хартли. Хотя графики, отражающие эти преобразования, выглядят по-разному, из преобразований Фурье и Хартли можно получить одну и ту же информацию об амплитуде и фазе исследуемого сигнала.

Амплитуда по Фурье определяется квадратным корнем из суммы квадратов действительной и мнимой частей. Амплитуда спектра сигнала по Хартли определяется квадратным корнем из суммы квадратов $H(\omega)$ и $H(-\omega)$. Фаза по преобразованию Фурье определяется арктангенсом мнимой части, деленной на действительную часть, а фаза по Хартли определяется суммой 45° и арктангенса угла от $H(-\omega)$, деленного на $H(\omega)$.

Оценка преобразования Хартли и его особенностей с точки зрения эффективности и удобства позволяет сделать выводы о практике применения при анализе сигналов. Преобразование Хартли используют для вычисления спектра, который аналогичен спектру Фурье. Преобразование Хартли осуществляет гар-

монический анализ по сумме двух составляющих, одна из которых сдвинута по фазе относительно другой на $\pi/2$. Действительные величины, получаемые при применении преобразования Хартли, дают определенные удобства при анализе сигналов и расчете их параметров в практических приложениях. К серьезным преимуществам преобразования Хартли относят и его полную симметричность. Однако преобразование Фурье является более наглядным с физической точки зрения, поскольку в большей степени отвечает установившимся представлениям о спектре сигналов как о совокупности гармонических колебаний.

Основной недостаток преобразования Хартли — отсутствие простой зависимости преобразования от сдвига исследуемого сигнала во времени. Преобразование Хартли приводит к более сложным выражениям, прежде всего для анализа спектра произведения и свертки сигналов. Эти операции встречаются довольно часто при анализе сигналов, и это обстоятельство надо учитывать.

## 2.3. Основы корреляционного анализа детерминированных сигналов

Корреляционный (*correlation*) анализ наряду со спектральным играет большую роль в теории сигналов. Это вид анализа сигналов «пришел» в радиотехнику в конце сороковых и начале пятидесятых годов прошлого века, а фундаментальная теорема Винера–Хинчина связала его и спектральный анализ в единое целое. В радиотехнике корреляционная теория используется, прежде всего, при исследовании случайных процессов, позволяя установить связь между корреляционными и спектральными свойствами случайных сигналов.

Для чего нужна корреляционная теория? Часто возникает задача обнаружения одного сигнала в другом или в помехах. Для надежного обнаружения полезных сигналов и применяется метод *корреляции*, основанный на корреляционной теории. Часто на практике оказывается полезным анализ характеристики, дающей представление о скорости изменения во времени, а также длительности сигнала без разложения его на гармонические составляющие.

Корреляционная функция является одной из основных характеристик радиолокационных сигналов. Она характеризует потенциальные возможности радиолокационной станции по разрешению целей. При рассмотрении в разд. 1.4 импульсной РЛС было отмечено, что информацию об обнаруженном объекте можно получить путем сравнения параметров излучаемого и отраженного (копии) сигналов. Пусть копия сигнала $u(t - \tau)$ смещена относительно своего оригинала $u(t)$ на некоторый заданный интервал времени $\tau$. Для количественной оценки степени отличия (связи) сигнала $u(t)$ и его смещенной во времени копии $u(t-\tau)$ используют *автокорреляционную функцию* (АКФ). Автокорреляционная функция показывает степень сходства между анализируемым сигналом и его сдвинутой копией — чем больше значение корреляционной функции, тем это сходство сильнее. Поэтому термин «*автокорреляция*» (*autocorrelation*) применяют, когда находят связь сигнала с его копией.

Для детерминированного сигнала конечной длительности (финитного сигнала) аналитическая запись АКФ представляет собой интеграл вида

$$B(\tau) = \int_{-\infty}^{\infty} u(t)u(t-\tau)dt. \qquad (2.99)$$

Формула (2.99) показывает, что при отсутствии сдвига копии относительно сигнала ($\tau = 0$) АКФ положительна и достигает максимального значения, равного энергии сигнала:

$$B(0) = \int_{-\infty}^{\infty} u^2(t)dt = Э. \qquad (2.100)$$

Такая энергия [Дж] выделяется на резисторе с сопротивлением в 1 Ом, если к его выводам подключить некоторое напряжение $u(t)$ [В].

Одним из важнейших свойств АКФ является ее четность: $B(\tau) = B(-\tau)$, что нетрудно доказать математически. Действительно, если в выражении (2.99) произвести замену переменной $x = t - \tau$, то

$$\int_{-\infty}^{\infty} u(t)u(t-\tau)dt = \int_{-\infty}^{\infty} u(x+\tau)u(x)dx.$$

Поэтому интеграл (2.99) можно представить в другом виде:

$$B(\tau) = \int_{-\infty}^{\infty} u(t)u(t+\tau)dt. \qquad (2.101)$$

Для периодического сигнала с периодом $T$, энергия которого бесконечно велика (поскольку сигнал существует бесконечное время), вычисление АКФ по (2.99) неприемлемо. В этом случае определяют АКФ за период

$$B_n(\tau) = \frac{1}{T}\int_{-T/2}^{T/2} u(t)u(t-\tau)dt. \qquad (2.102)$$

**Пример 2.3.** Определить АКФ прямоугольного импульса (рис. 2.26). Прямоугольный импульс имеет амплитуду $E$ и длительность $\tau_и$ (рис. 2.26, *а*).

Р е ш е н и е. Для такого сигнала вычисления АКФ удобнее и проще провести графически. Такое построение показано на рис. 2.26, *а – в*, где приведены соответственно: исходный видеоимпульс $u(t) = u$, сдвинутая на $\tau$ его копия $u_\tau(t) = u(t-\tau) = u_\tau$ и их произведение $u(t)u(t-\tau) = uu_\tau$.

Рассмотрим графическое вычисление интеграла (2.99). Очевидно, что произведение $u(t)u(t-\tau)$ не равно нулю на интервале времени, когда имеется наложение друг на друга любых частей сигнала и его копии.

Как следует из рис. 2.26, этот интервал равен $\tau - \tau_и$, если временно́й сдвиг копии меньше длительности импульса. В подобных случаях для прямоугольного импульса АКФ определится как $B(\tau) = E^2(\tau_и - |\tau|)$ при временно́м сдвиге копии на текущее время $|\tau| \le \tau_и$. АКФ прямоугольного импульса имеет вид равнобедренного треугольника с основанием в два раза больше длительности импульса и высотой, определяемой энергией сигнала $B(0) = E^2\tau_и = Э$ (рис. 2.26, *г*).

В радиотехнике и связи часто вводят удобный для анализа и сравнения сигналов числовой параметр — *интервал корреляции* $\tau_к$, аналитически и графически равный ширине основания АКФ. Для данного примера интервал корреляции $\tau_к = 2\tau_и$.

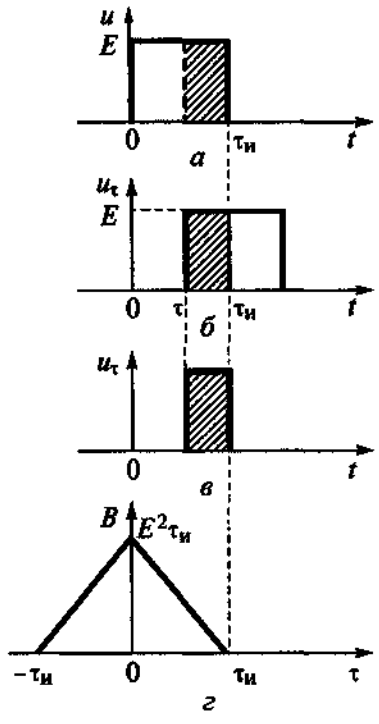



**Рис. 2.26.** Определение АКФ импульса: *а* — импульс; *б* — копия; *в* — произведение сигнала и копии; *г* — АКФ

**Пример 2.4.** Определить АКФ гармонического (косинусоидального) сигнала $u(t) = U_m\cos(\omega t + \varphi_0)$ с ненулевой начальной фазой (рис. 2. 27, *а*).

Р е ш е н и е. Используя формулу (2.102) и обозначив $B_u(\tau) = B(\tau)$, находим

$$B(\tau) = \frac{U_m^2}{T}\int_{-T/2}^{T/2}\cos(\omega t + \varphi_0)\cdot\cos(\omega t + \varphi_0 + \omega\tau)dt = 0{,}5U_m^2\cos\omega\tau.$$

Из этой формулы следует, что АКФ гармонического сигнала тоже является гармонической функцией (рис. 2.27, *б*) и имеет *размерность мощности* ($В^2$; очевидно, что на сопротивлении 1 Ом). Отметим еще один очень важный факт, что вычисленная АКФ не зависит от начальной фазы гармонического сигнала (параметр $\varphi$ в полученном выражении отсутствует).

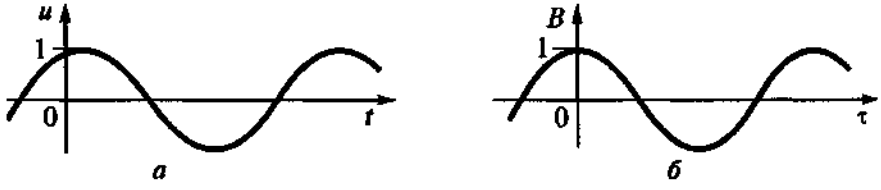



**Рис. 2.27.** Диаграммы к примеру 2.4: *а* — гармонический сигнал; *б* — АКФ гармонического сигнала

Из проведенного анализа следует важный вывод: АКФ практически любого сигнала *не зависит от его фазового спектра.* Следовательно, сигналы, амплитудные спектры которых полностью совпадают, а фазовые различаются, будут иметь одинаковую АКФ. Еще одно замечание заключается в том, что по АКФ нельзя восстановить исходный сигнал (опять же вследствие утраты информации о фазе).

### Связь между АКФ и энергетическим спектром сигнала

Допустим, что некоторый импульсный сигнал $u(t)$ имеет спектральную плотность $S(\omega)$. С помощью (2.99) определим АКФ, записав заданный сигнал $u(t)$ в виде обратного преобразования Фурье (2.55):

$$B(\tau) = \int_{-\infty}^{\infty} u(t)u(t-\tau)dt = \int_{-\infty}^{\infty}\left[\frac{1}{2\pi}\int_{-\infty}^{\infty} S(\omega)e^{j\omega t}d\omega\right]u(t-\tau)dt.$$

Для упрощения выкладок введем новую переменную $x = t - \tau$. Затем, сделав в последней формуле ряд несложных перестановок и выкладок, получим

$$B(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S(\omega)e^{j\omega\tau}\int_{-\infty}^{\infty} u(x)e^{j\omega x}dxd\omega. \qquad (2.103)$$

Здесь интеграл

$$\int_{-\infty}^{\infty} u(x)e^{j\omega x}dx = S^*(\omega) \qquad (2.104)$$

есть функция, *комплексно-сопряженная* со спектральной плотностью сигнала $S(\omega)$.

С учетом (2.104) формула (2.103) примет вид

$$B(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S(\omega)S^*(\omega)e^{j\omega\tau}d\omega. \qquad (2.105)$$

Функцию

$$W_и(\omega) = S(\omega)S^*(\omega) = |S(\omega)|^2 \qquad (2.106)$$

называют *энергетическим спектром* (*спектральной плотностью энергии — energy spectral density*) *сигнала*, который показывает распределение его энергии по оси частот. Размерность энергетического спектра сигнала соответствует величине $W_и(\omega)$ — [(В$^2$ · с)/Гц].

Учитывая соотношение (2.106), окончательно получим выражение для АКФ

$$B(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} W_и(\omega)e^{j\omega\tau}d\omega. \qquad (2.107)$$

Итак, АКФ сигнала представляет собой обратное преобразование Фурье от его энергетического спектра. Имеется и прямое преобразование Фурье от АКФ

$$\mathrm{W}_{\text{и}}(\omega) = \int_{-\infty}^{\infty} B(\tau) e^{-j\omega\tau} d\tau. \qquad (2.108)$$

Итак, с помощью полученных соотношений приходим к очень важному выводу: *прямое преобразование Фурье* (2.108) *автокорреляционной функции определяет энергетический спектр*, а *обратное преобразование Фурье энергетического спектра* (2.107) — *автокорреляционную функцию детерминированного сигнала.*

Эти результаты имеют фундаментальное значение и важны по двум причинам. Во-первых, исходя из распределения энергии по спектру становится возможным оценить корреляционные свойства сигналов — чем шире энергетический спектр сигнала, тем меньше интервал корреляции. Соответственно, чем больше интервал корреляции сигнала, тем короче его энергетический спектр. Во-вторых, соотношения (2.107) и (2.108) позволяют экспериментально определить одну из функций по значению другой. Часто удобнее вначале получить АКФ, а затем с помощью прямого преобразования Фурье вычислить энергетический спектр. Этот прием широко применяют при анализе свойств сигналов в реальном масштабе времени, т. е. без временно́й задержки при его обработке.

### Взаимокорреляционная функция двух сигналов

Если надо оценить степень связи между сигналами $u_1(t)$ и $u_2(t)$, то используют *взаимокорреляционную функцию* (ВКФ; *cross-correlation function*, CCF)

$$B_{12}(\tau) = \int_{-\infty}^{\infty} u_1(t) u_2(t-\tau) dt. \qquad (2.109)$$

При $\tau = 0$ ВКФ равна так называемой *взаимной энергии двух сигналов*

$$Э_{12} = B_{12}(0) = \int_{-\infty}^{\infty} u_1(t) u_2(t) dt. \qquad (2.110)$$

Значение ВКФ не меняется, если вместо задержки второго сигнала $u_2(t)$ рассматривать опережение его первым сигналом $u_1(t)$, поэтому

$$B_{12}(\tau) = \int_{-\infty}^{\infty} u_1(t) u_2(t-\tau) dt = \int_{-\infty}^{\infty} u_2(t) u_1(t+\tau) dt. \qquad (2.111)$$

АКФ является частным случаем ВКФ, если сигналы одинаковы

$$u_1(t) = u_2(t) = u(t).$$

В отличие от АКФ ВКФ двух сигналов $B_{12}(\tau)$ не является четной и необязательно максимальна при $\tau = 0$, т. е. при отсутствии временно́го сдвига сигналов.

## 2.4. Аналоговая модуляция несущих колебаний

Как уже отмечалось, под *модуляцией* понимают процесс (медленный по сравнению с периодом несущего колебания), при котором один или несколько параметров несущего колебания изменяют по закону передаваемого сообщения. Получаемые в процессе модуляции колебания называют *радиосигналами*. В аналоговых системах связи радиосигналы передаются непрерывно во времени, и при модуляции могут изменяться амплитуда, частота или фаза несущего гармонического колебания.

В зависимости от того, какой из названных параметров несущего колебания подвергается изменению, различают два основных вида аналоговой модуляции: *амплитудную* и *угловую*. Последний вид модуляции, в свою очередь, разделяется на *частотную* и *фазовую*.

В современных цифровых системах передачи информации широкое распространение получила *квадратурная* (амплитудно-фазовая, или фазоамплитудная — ФАМ; *amplitude phase modulation*) *модуляция*, при которой одновременно изменяются и амплитуда и фаза сигнала. Этот тип модуляции относят как к аналоговым, так и цифровым видам.

В настоящее время все большая часть информации, передаваемой по разнообразным каналам радио связи и другим типам связи, существует в импульсном и цифровом виде. Это означает, что передаче подлежит не непрерывный (аналоговый) модулирующий сигнал, а импульсные последовательности и последовательность целых чисел, которые могут принимать значения из некоторого фиксированного конечного множества. Поэтому в подобных радиосистемах применяются и будут применяться различные виды *импульсной* и *цифровой модуляции*, при которой радиосигналы представляются в виде так называемых радиоимпульсов.

Используют также несколько экзотические виды модуляции, как спектральная и поляризационная.

*Спектральная модуляция* — модуляция, при которой несущая модулируется по частоте в сочетании с третьим, кодовым сигналом. Спектральную модуляцию применяют в военной технике и пакетных радиосетях передачи данных.

*Поляризационная модуляция* — модуляция в оптическом диапазоне частот, основанная на изменении угла плоскости поляризации света.

### *Радиосигналы с аналоговыми видами модуляции*

Долгие годы, по существу начиная с изобретения радио, при передаче информации использовались и до сих пор широко используются аналоговые виды модуляции. Это связано в основном с простотой схемотехнического решения таких видов передачи сигналов. Очевидно, что аналоговые системы (например, радиовещание, телевидение и пр.) еще какое-то время будут существовать наряду с импульсными и цифровыми системами передачи информации.

При анализе модулированных колебаний и их спектров удобно использовать ряд основных формул тригонометрии:

$$\cos(\alpha-\beta)=\cos\alpha\cos\beta+\sin\alpha\sin\beta;\qquad \cos(\alpha+\beta)=\cos\alpha\cos\beta-\sin\alpha\sin\beta;$$

$$\sin(\alpha-\beta)=\sin\alpha\cos\beta-\cos\alpha\sin\beta;\qquad \sin(\alpha+\beta)=\sin\alpha\cos\beta+\cos\alpha\sin\beta;$$

$$\cos\alpha\cos\beta=0{,}5[\cos(\alpha+\beta)-\cos(\alpha-\beta)];\qquad \sin\alpha\sin\beta=0{,}5[\cos(\alpha-\beta)-\cos(\alpha+\beta)].$$

### *Радиосигналы с амплитудной модуляцией*

Как уже отмечалось, первым передачу речи и музыки по радио осуществил в 1906 г. американский инженер Р. Фессенден. Модуляция несущего колебания осуществлялась путем включения микрофона, изменяющего затухание в цепи, связывающей передающую антенну и электромашинный генератор высокой частоты. Этот вид модуляции с 1920 г. стал основным в звуковом радиовещании в диапазонах низких, средних и высоких частот и сети аналогового вещания с амплитудной модуляцией.

Напомним, что в процессе амплитудной модуляции несущего колебания

$$u_{н}(t)=U_{н}\cos(\omega_0 t+\varphi_0)=U_{н}\cos\psi(t), \tag{2.112}$$

его амплитуда должна изменяться по закону:

$$U_{н}(t)=U_{н}+k_{A}e(t), \tag{2.113}$$

где $U_{н}$ — амплитуда несущей в отсутствие модуляции; $\omega_0$ — угловая частота; $\varphi_0$ — начальная фаза; $\psi(t)=\omega_0 t+\varphi_0$ — полная (текущая или мгновенная) фаза несущей; $k_{A}$ — безразмерный коэффициент пропорциональности; $e(t)$ — модулирующий сигнал.

Заметим, что функцию $U_{н}(t)$ в радиотехнике принято называть *огибающей* амплитудно-модулированного сигнала (АМ-сигнала).

Подставив (2.113) в (2.112), получим общую формулу АМ-сигнала:

$$u_{AM}(t)=U_{н}(t)\cos(\omega_0 t+\varphi_0)=[U_{н}+k_{A}e(t)]\cos(\omega_0 t+\varphi_0). \tag{2.114}$$

**Однотональная амплитудная модуляция.** Обратимся к простейшей амплитудной *однотональной* (от слова «тон» — звук одной частоты) или *гармонической* модуляции, когда модулирующий сигнал — гармоническое колебание

$$e(t)=E_0\cos(\Omega t+\theta_0), \tag{2.115}$$

где $E_0$ — амплитуда; $\Omega=2\pi/T_1=2\pi F$ — угловая частота модуляции; $F$ — циклическая частота модуляции; $T_1$ — период модуляции; $\theta_0$ — начальная фаза.

Физически процесс амплитудной модуляции осуществляется путем перемножения двух сигналов, поэтому, подчиняясь принципу суперпозиции, является линейной операцией. Подставив формулу (2.115) в соотношение (2.114), получим выражение для АМ-сигнала

$$u_{AM}(t)=[U_{н}+k_{A}E_0\cos(\Omega t+\theta_0)]\cos(\omega_0 t+\varphi_0). \tag{2.116}$$

Обозначив через $\Delta U = k_A E_0$ максимальное отклонение амплитуды АМ-сигнала от амплитуды несущей $U_н$ и проведя несложные выкладки, получим

$$u_{AM}(t) = U_н[1 + M\cos(\Omega t + \theta_0)]\cos(\omega_0 t + \varphi_0), \qquad (2.117)$$

где

$$M = k_A E_0 / U_н = \Delta U / U_н \qquad (2.118)$$

*— коэффициент* или *глубина амплитудной модуляции.*

Следует помнить, что однотональная модуляция (диаграмма АМ-сигнала) симметрична относительно оси времени.

Спектр АМ-сигнала. Применив в выражении (2.117) тригонометрическую формулу произведения косинусов, после несложных выкладок получим

$$u_{AM}(t) = U_н \cos(\omega_0 t + \varphi_0) + U_н M \cos(\Omega t + \theta_0)\cos(\omega_0 t + \varphi_0) =$$

$$= U_н \cos(\omega_0 t + \varphi_0) + \frac{MU_н}{2}\cos[(\omega_0 + \Omega)t + \varphi_0 + \theta_0] + \qquad (2.119)$$

$$+ \frac{MU_н}{2}\cos[(\omega_0 - \Omega)t + \varphi_0 - \theta_0].$$

Из формулы (2.119) видно, что при однотональной амплитудной модуляции спектр АМ-сигнала состоит из трех высокочастотных составляющих. Первая из них представляет собой исходное несущее колебание с постоянной амплитудой $U_н$ и частотой $\omega_0$. Вторая и третья составляющие характеризуют новые гармонические колебания, появляющиеся в процессе амплитудной модуляции и отражающие передаваемый сигнал.

Колебания с частотами $\omega_0 + \Omega$ и $\omega_0 - \Omega$ называются соответственно *верхней* (*upper sideband* — USB) и *нижней* (*lower sideband* — LSB) *боковыми составляющими.* Амплитуды боковых составляющих АМ-сигнала одинаковы, равны $MU_н/2$ и расположены симметрично относительно несущей частоты сигнала $\omega_0$. Впервые существование боковых полос частот в спектре амплитудно-модулированных колебаний установил в 1916 г. русский ученый Михаил Васильевич Шулейкин (1884–1939).

Заметим, что в общем случае представление однотонального АМ-сигнала в виде (2.119) *не является* разложением в ряд Фурье. Как ряд Фурье представление АМ-сигнала суммой трех гармонических функций можно рассматривать только тогда, когда для всех этих частот спектральных составляющих существует общий делитель. Это возможно, если отношение частот $\omega_0/\Omega$ является рациональной дробью.



М.В. Шулейкин

Реальная ширина спектра АМ-сигнала при однотональной модуляции

$$\Delta\omega_c = \Delta\omega_{AM} = 2\Omega = 4\pi F. \qquad (2.120)$$

Принцип однотональной амплитудной модуляции показан на рис. 2.28. Графики модулирующего сигнала $e(t)$ с $\theta_0 = 90°$, несущего колебания $u_н(t)$ с $\varphi_0 = 90°$ и АМ-сигнала $u_{АМ}(t)$ представлены на рис. 2.28, *а – в*, а на рис. 2.28, *г – е* — соответствующие им спектры $S_0(\omega)$, $S_н(\omega)$ и $S_{АМ}(\omega)$.

В отсутствие модуляции ($M = 0$) амплитуды боковых составляющих равны нулю и спектр АМ-сигнала переходит в спектр несущего колебания (составляющая $U_н$ на частоте $\omega_0$). При глубине модуляции $M \le 1$ амплитуда АМ-сигнала изменяется в пределах от минимального $U_{мин} = U_н(1 - M)$ до максимального $U_{макс} = U_н(1 + M)$ значения. Исключая постоянное значение $U_н$, получаем формулу, удобную для экспериментального определения коэффициента модуляции (отметим, что данное соотношение справедливо и для амплитуд токов АМ-сигнала, обозначаемых как $I_{макс}$ и $I_{мин}$):

$$M = \frac{U_{макс} - U_{мин}}{U_{макс} + U_{мин}}. \qquad (2.121)$$

Как правило, модулирующий сигнал (см. рис. 2.28, *а*) является двуполярным (знакопеременным). Из графика на рис. 2.28, *в* можно заметить, что амплитудная огибающая, которая будет выделена в процессе детектирования (демодуляции), в данном случае оказывается неправильной — она соответствует *модулю* исходного сигнала. Поэтому при реализации амплитудной модуляции к модулирующему сигналу предварительно добавляют постоянную составляющую $U_н$, чтобы сделать его однополярным: $U_н(t) = U_н + k_A e(t)$.

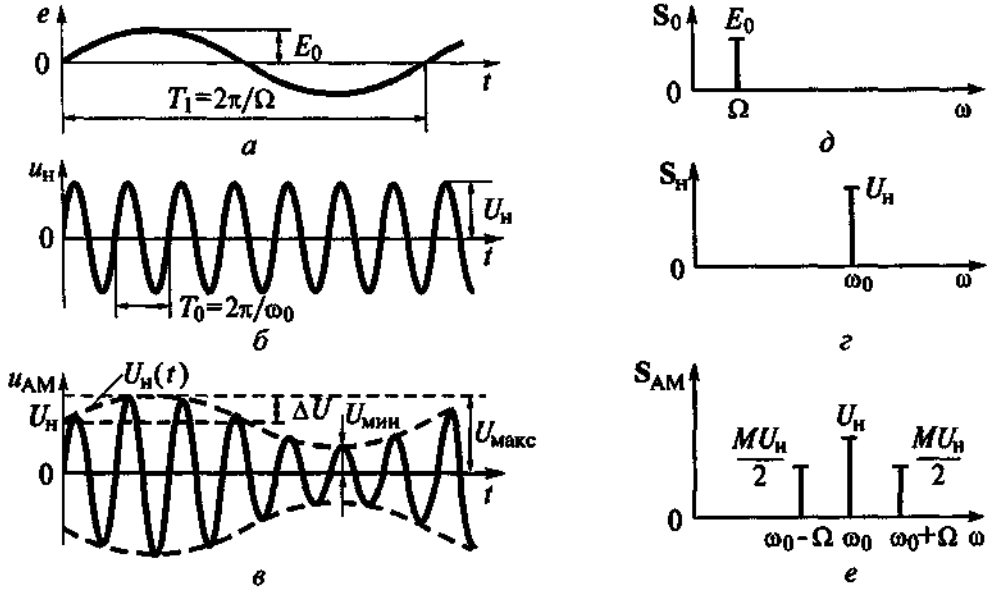



**Рис. 2.28.** Амплитудная модуляция:
*а* — модулирующий сигнал; *б* — несущее колебание; *в* — АМ-сигнал;
*г – е* — соответствующие спектры

Важное значение для верной передачи АМ-сигналов имеет коэффициент модуляции $M$. Если при модуляции получается, что $M > 1$, то возникают искажения, называемые *перемодуляцией* (рис. 2.29). Наличие таких искажений в АМ-сигнале может привести к потере передаваемой информации. Однако иногда этот режим может быть полезен на практике.




**Рис. 2.29.** Однотональный АМ-сигнал при перемодуляции

Чтобы понять, как из трех гармонических составляющих с *постоянной* амплитудой получается радиосигнал с *меняющейся* амплитудой, построим векторную диаграмму при однотональной модуляции несущего колебания. Для упрощения выкладок примем в соотношении (2.119) начальные фазы несущего колебания и модулирующего сигнала $\varphi_0 = 0$ и $\theta_0 = 0$. При необходимости они могут быть введены в окончательные соотношения. Тогда формула (2.119) отразится векторной диаграммой (рис. 2.30), на которой каждая из трех составляющих АМ-сигнала представлена соответствующим вектором.

На векторной диаграмме ось времени $t$ вращается по часовой стрелке с угловой скоростью $\omega_0$. Поэтому несущее колебание изображается на этой оси в виде вектора 0А длиной $U_н$. Поскольку принято, что угол $\varphi_0 = 0$, то ось времени совпала с вектором несущего колебания. Верхняя $U_{ВБ}$ и нижняя $U_{НБ}$ боковые составляющие изображаются на векторной диаграмме соответственно векторами АС и АД длиной $MU_н/2$ каждый. Они составляют с направлением вектора несущей 0А углы $\pm\Omega t$ и вращаются в противоположных направлениях с угловой скоростью $\Omega$. Равнодействующим вектором боковых составляющих АС и АД является *вектор модуляции* АВ. Так как векторы боковых составляющих вращаются в противоположных направлениях с одинаковой угловой скоростью и расположены симметрично относительно вектора 0А, то суммарный вектор 0В АМ-сигнала в любой момент времени совпадает с направлением вектора несущего колебания 0*А*. Длина вектора 0В АМ-сигнала будет периодически изменяться от максимальной $U_н(1 + M)$ до минимальной $U_н(1 - M)$ значений. Если при прохождении через электрические цепи нарушается равенство амплитуд боковых составляющих или симметрия их фаз по отношению фазы несущего, то возникает «качание» вектора модуляции относительно направления вектора несущей 0А. Это равносильно возникновению паразитной фазовой модуляции.

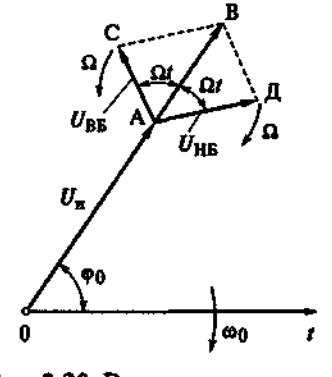



**Рис. 2.30.** Векторная диаграмма однотонального АМ-сигнала

Энергетические соотношения в АМ-сигнале. В соответствии с изменением амплитуды изменяется и средняя за пери-

од несущей частоты $T_0 = 2\pi/\omega_0$ мощность модулированного колебания. Пикам огибающей соответствует мощность в $(1 + M)^2$ раз больше мощности несущего колебания. При этом средняя мощность за период модуляции пропорциональна среднему квадрату амплитуды $U_н(t)$, и для ее определения необходимо применить предельный переход согласно формуле

$$P_{ср} = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} u^2(t)dt = \\ = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} \left\{ U_н \left[ 1 + M\cos(\Omega t + \theta_0) \right] \cos(\omega_0 t + \varphi_0) \right\}^2 dt = \frac{U_н^2}{2} + \frac{U_н^2 M^2}{4}. \quad (2.122)$$

Первое слагаемое в соотношении (2.122) не зависит от коэффициента модуляции и представляет собой мощность несущей (это мощность любого гармонического колебания), равную $0{,}5U_н^2$. Полезная мощность, заключенная в боковых частотах, представлена вторым слагаемым $0{,}25U_н^2M^2$.

Итак, средняя мощность однотонального АМ-сигнала за период модуляции $T_1 = 2\pi/\Omega$ превышает мощность несущего колебания в $(1 + 0{,}5M^2)$ раза. Нетрудно заметить, что доля мощности обеих боковых составляющих АМ-сигнала даже при 100%-ной гармонической модуляции ($M = 1$) равна лишь половине мощности несущего колебания. Поскольку полезная информация (передаваемое сообщение) заложена только в боковых составляющих АМ-сигнала, можно отметить неэффективность использования излучаемой мощности при амплитудной модуляции. Вместе с тем на принципах амплитудной модуляции пока еще действует значительное число вещательных систем, а также систем телевидения.

Отметим очевидное: *передача полного АМ-сигнала* (*несущей* и *двух боковых полос*) обеспечивает простое получение исходного сигнала на приеме. Для этого достаточно подать АМ-сигнал на демодулятор и с помощью фильтра выделить исходный сигнал. Оконечное передающее и приемное оборудование будет относительно простым, канальные фильтры — несложные, в приемном оборудовании нет необходимости использовать для демодуляции специальный детектор.

Попытки улучшить характеристики амплитудной модуляции привели к разработке нескольких ее модификаций. Учитывая структуру АМ-сигнала, формирование радиоканальных сигналов можно осуществлять путем передачи:

- двух боковых полос частот без несущей (балансная амплитудная модуляция);
- одной боковой полосы частот и несущей;
- одной боковой полосы частот без несущей;
- одной боковой полосы частот, несущей и части второй боковой полосы частот.

Коротко рассмотрим основные из этих методов для определения особенностей организации передачи информации при использовании каждого из них.

**Балансная амплитудная модуляция** (*balancing amplitude modulation*). Первое, что приходит в голову при размышлении на тему повышения энергетических характеристик амплитудной модуляции, — идея удалить бесполезное несущее

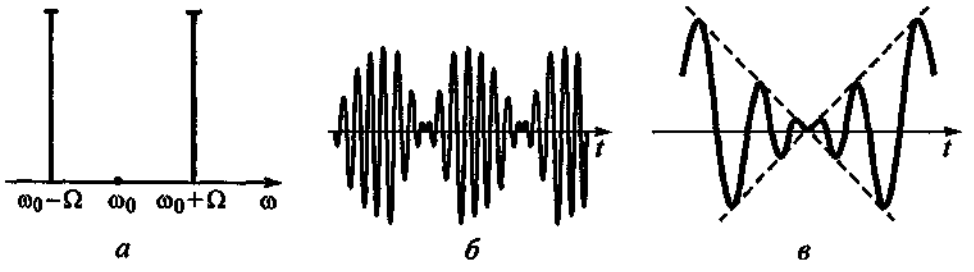



**Рис. 2.31.** Балансная модуляция:
*а* — спектр; *б* — осциллограмма; *в* — график перехода огибающей через нуль

колебание. Такой способ называется *балансной амплитудной модуляцией* (БАМ) или *амплитудной модуляцией с подавленной несущей* (АМ-ПН; *amplitude modulation with suppressed carrier*, AM-SC). Выражение для сигнала с балансной амплитудной модуляцией нетрудно получить из формулы (2.119), выбрав для упрощения выкладок фазовые углы $\varphi_0 = 0$ и $\theta_0 = 0$. Тогда

$$u_{\text{БАМ}}(t) = \frac{MU_{\text{н}}}{2}\cos(\omega_0 + \Omega)t + \frac{MU_{\text{н}}}{2}\cos(\omega_0 - \Omega)t. \qquad (2.123)$$

В случае балансной амплитудной модуляции (рис. 2.31) имеет место перемножение двух сигналов — модулирующего и несущего. Это явление и сигнал вида (2.123) называют *биениями* (*beat*) двух гармонических сигналов с одинаковыми амплитудами $MU_{\text{н}}/2$ и частотами, равными верхней и нижней боковым частотам однотонального АМ-сигнала (рис. 2.31, *а*). Получаемая при биениях разностная частота $\omega_6 = \omega_0 - \Omega$ называется *частотой биений* (*beat frequency*).

При анализе осциллограммы биений АМ-колебания (рис. 2.31, *б*) может показаться неясным, почему в спектре этого сигнала нет несущей частоты $\omega_0$, хотя совершенно очевидно наличие высокочастотного заполнения, изменяющегося во времени именно с этой частотой. Такое положение связано с тем, что при переходе огибающей биений через нуль (рис. 2.31, *в*) фаза высокочастотного заполнения скачком изменяется на 180°, поскольку функция огибающей $\cos\Omega t$ имеет разные знаки слева и справа от нуля. Если этот сигнал подать на высокодобротную колебательную систему (например, резонансный *LC*-контур), настроенную на частоту $\omega_0$, то выходной эффект будет очень мал, стремясь к нулю при возрастании добротности. Колебания в системе, возбужденные одним периодом биений, будут гаситься последующим периодом. Именно так с физических позиций принято рассматривать вопрос о реальном смысле спектрального состава сигнала с БАМ.

Ширина спектра АМ-сигнала с подавленной несущей такая же, как при обычной амплитудной модуляции (поскольку подавлена несущая частота, а боковые частоты остались на месте). Несложно показать, что и при многотональной балансной модуляции выражение АМ-сигнала с подавленной несущей содержит две симметричные группы верхних и нижних боковых колебаний.

Итак, амплитудная модуляция с подавленной несущей обладает определенными преимуществами по сравнению с обычной амплитудной модуляцией. Однако этот способ модуляции не получил широкого распространения, что связано с проблемами, возникающими при детектировании радиосигнала.

**Однополосная амплитудная модуляция** (*передача одной боковой полосы частот и несущей*). Большое значение для радиосвязи имело изобретение американским ученым Дж. Карсоном амплитудной модуляции с одной боковой полосой (ОБП; *Single Side Band*, SSB; часто *SSB*-сигналы), сделанное в 1915 г. В современных системах радиосвязи часто приходится экономить не только мощность, но и полосу занимаемых частот. Как уже отмечалось, спектры двух боковых полос АМ-сигнала являются зеркальным отражением друг друга относительно несущей частоты, т. е. они несут одну и ту же информацию. Поэтому одну из боковых полос можно удалить. В результате получается колебание с несущей и одной боковой полосой. В более общем случае под сигналами с одной боковой полосой, или сигналами *однополосной модуляции* (ОМ), понимают колебания, полученные при модуляции гармонической несущей частоты и отличающиеся тем, что их спектр (на положительных частотах) располагается по одну сторону (слева или справа) от несущей $\omega = \omega_0$. Сигналы с однополосной амплитудной модуляцией занимают полосу частот в два раза более узкую, чем обычный АМ-сигнал. Это обстоятельство и обусловливает большой интерес к системам связи с ОБП в тех случаях, когда экономия полосы частот канала является решающим фактором выбора системы сигналов.

По внешним характеристикам модулированный сигнал с одной боковой полосой и несущей напоминает обычный АМ-сигнал. В зависимости от того, какая боковая полоса сохраняется, говорят об однополосной модуляции с использованием *верхней* или *нижней* боковой составляющей (полосой). В частности, однотональный ОБП-сигнал с подавленной нижней боковой составляющей и начальными фазами несущей $\varphi_0 = 0$ и модулирующего колебания $\theta_0 = 0$ записывается в таком виде

$$u_{\text{ОБП}}(t) = U_{\text{н}} \cos\omega_0 t + \frac{MU_{\text{н}}}{2} \cos(\omega_0 + \Omega)t. \tag{2.124}$$

Проводя известные тригонометрические преобразования, получаем

$$u_{\text{ОБП}}(t) = U_{\text{н}} \cos\omega_0 t + \frac{MU_{\text{н}}}{2} \cos\Omega t \cos\omega_0 t - \frac{MU_{\text{н}}}{2} \sin\Omega t \sin\omega_0 t = \tag{2.125}$$

$$= U_{\text{н}} \left(1 + \frac{M}{2} \cos\Omega t\right) \cos\omega_0 t - \frac{MU_{\text{н}}}{2} \sin\Omega t \sin\omega_0 t.$$

Каждый из членов правой части последнего соотношения представляет собой произведение двух аналитических функций, одна из которых изменяется во времени медленно (отражает модулирующий сигнал), а другая — быстро (отражает несущее колебание).

Итак, сигнал с ОБП можно представить суммой двух АМ-сигналов с одинаковыми несущими, но сдвинутыми по фазе друг относительно друга на 90°. Амплитудными функциями АМ-сигналов являются модулирующий сигнал и его квадратурное дополнение. В зависимости от того, складываются эти сигналы или вычитаются (а точнее, какая из несущих опережает другую по фазе), формируется однополосный сигнал с верхней или нижней боковой полосой. По существу при однополосной модуляции происходит сдвиг спектра передаваемого сигнала в окрестность несущей частоты. В отличие от обычной амплитудной модуляции, каждая «половинка» спектра смещается в своем направлении: область положительных частот к $+\omega_0$, а область отрицательных частот к $-\omega_0$. При формировании сигнала с нижней боковой полосой спектр модулирующего сигнала инвертируется (зеркально поворачивается вдоль оси частот).

Учитывая, что «быстрые» сомножители находятся в квадратуре (сдвинуты по фазе на 90°), вычислим медленно изменяющуюся огибающую ОБП-сигнала:

$$U_{\text{ОБП}}(t) = U_{\text{н}}\sqrt{\left(1+\frac{M}{2}\cos\Omega t\right)^2 + \frac{M^2}{4}\sin^2\Omega t} = U_{\text{н}}\sqrt{1+M\cos\Omega t+\frac{M^2}{4}}. \quad (2.126)$$

Временна́я диаграмма огибающей ОБП-сигнала (кривая 1), рассчитанная по формуле (2.126) при глубине модуляции $M = 1$, показана на рис. 2.32. Здесь же для сравнения построена огибающая однотонального АМ-сигнала (кривая 2) с той же глубиной модуляции. Сравнение кривых показывает, что непосредственное детектирование ОБП-сигнала по огибающей будет сопровождаться значительными искажениями. Исходная информация на приеме будет образовываться от взаимодействия переданных боковой полосы частот и несущей. Однако для подавления одной из боковых полос необходимо использовать сложные в реализации канальные фильтры.

Другой, более эффективной с точки зрения энергетических показателей, разновидностью АМ-сигналов является однополосная амплитудная модуляция с подавленной несущей (ОБП-ПН):

$$u_{\text{ОБП-ПН}}(t) = \frac{MU_{\text{н}}}{2}\cos(\omega_0+\Omega)t. \quad (2.127)$$

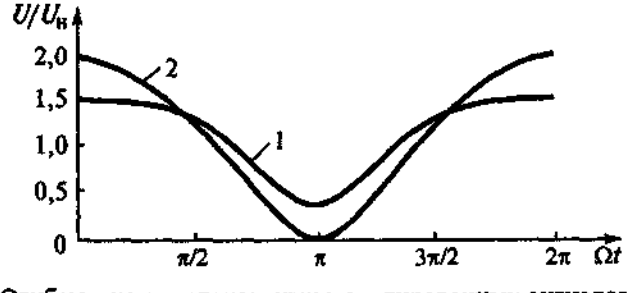



**Рис. 2.32.** Огибающие однотональных модулированных сигналов при $M = 1$:
1 — АМ-сигнала; 2 — ОБП-сигнала

Данный вид амплитудной модуляции представляет собой такое преобразование несущего колебания, при котором спектр радиосигнала полностью совпадает со спектром сообщения, перенесенным по оси частот в высокочастотную область нижней или верхней боковой полосы.

*Передача одной боковой полосы без несущей* обеспечивает минимально возможную ширину спектра канального сигнала, равную ширине спектра исходного сигнала, что позволяет наиболее экономно реализовать линейный спектр частот сигнальной полосы. *Передача одной боковой полосы частот, несущей и части второй боковой полосы частот* используется, когда спектр исходного сигнала начинается от очень низких частот, близких к нулю. К сигналам такого вида относятся телевизионный и факсимильный сигналы. При реализации этого метода передачи канальный фильтр должен иметь кососимметричную характеристику коэффициента передачи относительно несущей частоты.

В настоящее время однополосная модуляция с различным уровнем несущего колебания — от полного подавления до полного сохранения — нашла широкое применение в технике профессиональной радиосвязи.

На практике однотональные АМ-сигналы используются либо для учебных, либо для исследовательских целей. Реальный же модулирующий сигнал имеет сложный спектральный состав. Математически такой сигнал, состоящий из $N$ гармоник, можно представить тригонометрическим рядом

$$e(t) = \sum_{i=1}^{N} E_i \cos \Omega_i t. \qquad (2.128)$$

Здесь амплитуды гармоник сложного модулирующего сигнала $E_i$ произвольны, а их частоты образуют упорядоченный спектр $\Omega_1 < \Omega_2 < ... < \Omega_i < ... < \Omega_N$. В отличие от ряда Фурье частоты $\Omega_i$ не обязательно кратны друг другу.

Подставляя (2.128) в (2.114), после несложных преобразований, получим выражение АМ-сигнала с начальной фазой несущего $\varphi_0 = 0$

$$u_{\text{АМ}}(t) = U_{\text{н}}(1 + \sum_{i=1}^{N} M_i \cos \Omega_i t) \cos \omega_0 t, \qquad (2.129)$$

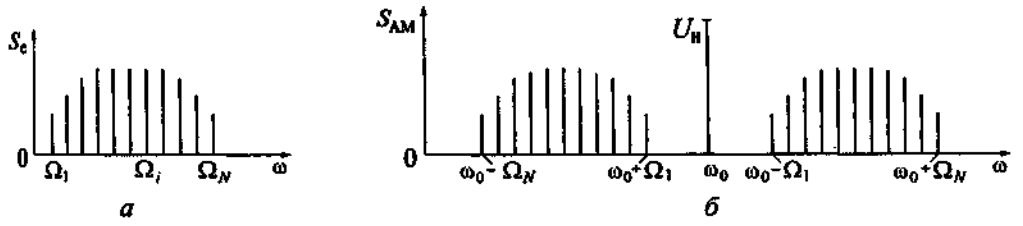
где

$$M_i = k_{\text{А}} E_i / U_{\text{н}}$$

— совокупность *парциальных (частичных) коэффициентов модуляции.*

Эти коэффициенты характеризуют влияние гармонических составляющих модулирующего сигнала на общее изменение амплитуды высокочастотного колебания. Воспользовавшись тригонометрической формулой произведения двух косинусов и проделав несложные преобразования, запишем (2.129) в виде

$$u_{\text{АМ}}(t) = U_{\text{н}} \cos \omega_0 t + \sum_{i=1}^{N} \frac{M_i U_{\text{н}}}{2} \cos(\omega_0 + \Omega_i)t + \sum_{i=1}^{N} \frac{M_i U_{\text{н}}}{2} \cos(\omega_0 - \Omega_i)t. \qquad (2.130)$$




*а* *б*

**Рис. 2.33.** Спектральные диаграммы при модуляции сложным сигналом: *а* — модулирующего сигнала; *б* — АМ-сигнала

Из формулы (2.130) видно, что в спектре сложного АМ-сигнала, наряду с несущим колебанием, содержатся группы верхних и нижних боковых составляющих, являющихся масштабными копиями спектра модулирующего сигнала и расположенных симметрично относительно частоты $\omega_0$. Отсюда следует важный вывод: *ширина спектра сложного АМ-сигнала равна удвоенному значению наивысшей частоты в спектре модулирующего сигнала* $\Omega_N$, т. е.

$$\Delta\omega_c = \Delta\omega_{AM} = 2\Omega_N. \qquad (2.131)$$

В качестве примера на рис. 2.33 показаны спектральные диаграммы при модуляции несущего сложным сигналом. Видно, что модулирующий сигнал $S_c(\omega)$ (рис. 2.33, *а*) в АМ-сигнале $S_{AM}(\omega)$ расположен по обе стороны от несущей $U_н$ (рис. 2.33, *б*).

**Пример 2.5.** Определить спектральный состав и записать аналитическое выражение АМ-сигнала

$$u_{AM}(t) = 20(1 + 0{,}5\cos 10^3 t)\cos 10^5 t.$$

Р е ш е н и е. Воспользовавшись соотношением (2.118), находим, что несущая частота $\omega_0 = 10^5$; частота модуляции $\Omega = 10^3$; боковые частоты $\omega_0 + \Omega = 1{,}01\cdot10^5$; $\omega_0 - \Omega = 0{,}99\cdot10^5$; амплитуда несущей $U_н = 20$ В; коэффициент модуляции $M = 0{,}5$. Следовательно, в соответствии с выражением (2.118) имеем:

$$u_{AM}(t) = 20\cos 10^5 t + 5\cos(1{,}01\cdot10^5 t) + 5\cos(0{,}99\cdot10^5 t).$$

Приведенные рассуждения справедливы не только для дискретного спектра, но и для любого вида спектра передаваемого сообщения. Можно считать, что спектр модулирующего сигнала заключен под огибающей заданного вида. В этом случае спектральная плотность передаваемого АМ-сигнала состоит из двух симметричных идентичных всплесков спектра модулирующего сигнала $S_c(\omega)$ относительно несущей частоты $\omega_0$. Значит для определения спектра амплитудно-модулированного колебания достаточно сдвинуть на частоту $\omega_0$ огибающую спектра исходного колебания.

Как уже отмечалось, исторически амплитудная модуляция была первым практически освоенным видом модуляции. Однако низкие энергетические показатели и ширина спектра, вдвое превышающая ширину спектра модулирующего сигнала, привели к тому, что сфера ее применения стала довольно узкой. В на-

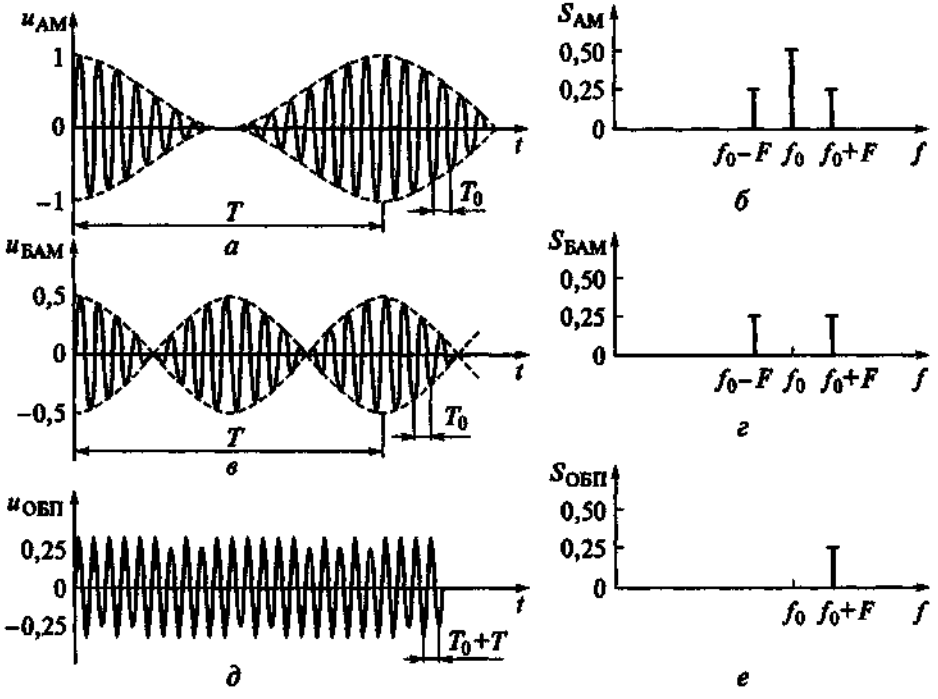



**Рис. 2.34.** Эпюры радиосигналов и их спектров с вариантами амплитудной модуляции: *а, б* — АМ-сигнала; *в, г* — БАМ-сигнала; *д, е* — ОБП-сигнала

стоящее время амплитудная модуляция применяется для радиовещания на сравнительно низких частотах (в диапазонах от километровых до декаметровых волн) и для передачи изображения в телевизионном вещании.

Сопоставление сигналов с вариантами амплитудной модуляции. На рис. 2.34 показаны эпюры радиосигналов и их спектров с вариантами однотональной амплитудной модуляции: АМ, БАМ и ОБП (на эпюрах $f_0 = 2\pi/\omega_0 = 1/T_0$ — циклическая частота несущей; $F = 2\pi/\Omega = 1/T$ — циклическая частота модулирующего гармонического сигнала).

Как видно из сопоставления рис. 2.34, *а, б* и рис. 2.34, *в, г*, устранение несущего колебания в исходном АМ-сигнале приводит к существенным изменениям огибающей модулированного колебания, которая перестала копировать модулирующий сигнал. Это обстоятельство усложняет детектирование — определение модулирующего сигнала при приеме балансно-модулированного колебания. Если убрать и нижнюю боковую частоту (получится модуляция с одной боковой полосой — ОБП), то спектр демонстрирует экономное использование отведенной полосы частот и мощности радиопередатчика системы передачи информации (рис. 2.34, *д, е*), но вместо модулированного колебания получится гармоническое с частотой, равной частоте верхней боковой $f_0 + F$. Для того чтобы продетектировать такое колебание, надо определить амплитуду и частоту модулирующего сигнала.

Наибольшие проблемы возникают при определении частоты модулирующего сигнала, поскольку надо знать частоту несущего колебания. Есть несколько вариантов решения этой проблемы. Например, можно непрерывно передавать в точку приема это несущее колебание с сильно уменьшенной амплитудой (в виде так называемого *пилот-сигнала*). Несущее колебание можно также передавать периодически, синхронизируя этими «всплесками» генератор несущей в точке приема (так передают информацию о частоте и фазе цветовой поднесущей в системах цветного телевидения). Однако сложность процесса детектирования окупается более эффективным использованием полосы рабочих частот.

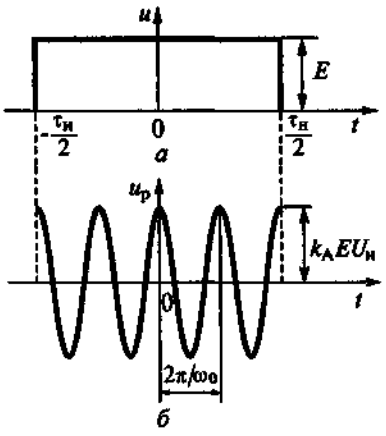



**Рис. 2.35.** Импульсные сигналы: *а* — видеоимпульс; *б* — радиоимпульс

**Радиосигналы с амплитудно-импульсной модуляцией.** В радиотехнике и теории связи широкое применение находят различные модулированные импульсные последовательности, среди которых обязательно следует выделить последовательности прямоугольных радиоимпульсов (рис. 2.35).

Вычислим спектр радиоимпульса, полученного в процессе амплитудной модуляции гармонического несущего колебания одиночным прямоугольным импульсом с амплитудой $E$ и длительностью $\tau_и$ (рис. 2.35, *а*). Как уже известно, спектральная плотность прямоугольного импульса определяется формулой (2.61) и имеет вид функции $\sin x/x$. Функцию, аналитически описывающую радиоимпульс прямоугольной формы, можно представить следующим образом:

$$u_р(t)=\begin{cases} k_A E U_н \cos\omega_0 t, & -\tau_и/2 \le t \le \tau_и/2; \\ 0, & t \le -\tau_и/2, \quad t \ge \tau_и/2, \end{cases} \qquad (2.132)$$

где $k_A$ — безразмерный коэффициент пропорциональности.

Очень важно определить спектральную плотность прямоугольного радиоимпульса и сравнить ее со спектральной плотностью видеоимпульса той же формы и тех же параметров. Спектральную плотность радиоимпульса вычислим, применив к (2.132) прямое преобразование Фурье (2.54). Для упрощения выкладок примем $k_A E U_н = U_р$ и запишем искомую спектральную плотность

$$S_р(\omega)=\int_{-\tau_и/2}^{\tau_и/2} U_р \cos\omega_0 t \cdot e^{-j\omega t}\,dt = \frac{U_р}{2}\left[\int_{-\tau_и/2}^{\tau_и/2} e^{-j(\omega-\omega_0)t}\,dt + \int_{-\tau_и/2}^{\tau_и/2} e^{-j(\omega+\omega_0)t}\,dt\right]. \qquad (2.133)$$

В этом соотношении первый интеграл определяет спектральную плотность исходного видеоимпульса с амплитудой $E = 1$ при частоте $\omega - \omega_0$, а второй — ту

же спектральную плотность, но при частоте $\omega + \omega_0$. Поэтому последнее выражение можно записать в следующей форме:

$$S_p(\omega) = \frac{U_p}{2}\left[S(\omega - \omega_0) + S(\omega + \omega_0)\right]. \quad (2.134)$$

Подставляя значение спектральной плотности прямоугольного импульса из соотношения (2.61) в (2.134), получим искомую спектральную плотность радиоимпульса (рис. 2.35, *б*):

$$S_p(\omega) = \frac{U_p \tau_и}{2}\left[\frac{\sin\frac{(\omega - \omega_0)\tau_и}{2}}{\frac{(\omega - \omega_0)\tau_и}{2}} + \frac{\sin\frac{(\omega + \omega_0)\tau_и}{2}}{\frac{(\omega + \omega_0)\tau_и}{2}}\right]. \quad (2.135)$$

На рис. 2.36 в качестве наглядного примера и для сравнения показаны диаграммы спектральных плотностей соответственно модулирующего видеоимпульса и аналогичного ему по огибающей радиоимпульса. Как нетрудно заметить из приведенных графиков, спектральная плотность радиоимпульса полностью повторяет по форме спектральную плотность модулирующего импульса.

Основное отличие спектральной плотности видеоимпульса (рис. 2.36, *а*) от спектральной плотности радиоимпульса той же формы — наличие сдвига спектральной плотности радиоимпульса по оси частот по обе стороны от начала координат на значение частоты несущей $\pm\omega_0$ (рис. 2.36, *б*).

**Полярная модуляция в стереофоническом вещании**. В СССР в 1939 г. наш соотечественник инженер А.И. Косцов изобрел еще один вид аналоговой амплитудной модуляции сигналов, названный *полярной модуляцией* (ПМ; *polar modulation*). Первая система звукового стереофонического радиовещания с полярной модуляцией сигналов была разработана русским инженером Л.М. Кононовичем в 1960 г. Радиосигналы не являются колебаниями с совершенно новым видом модуляции. Это скорее демонстрация применения разновидности амплитудной модуляции для решения конкретной технической задачи. Речь пойдет о реализации стереофонии в радиовещании метрового и дециметрового диапазонов и в различных устройствах звукового сопровождения (телевидении, видеоконференцсвязи и т. п.).

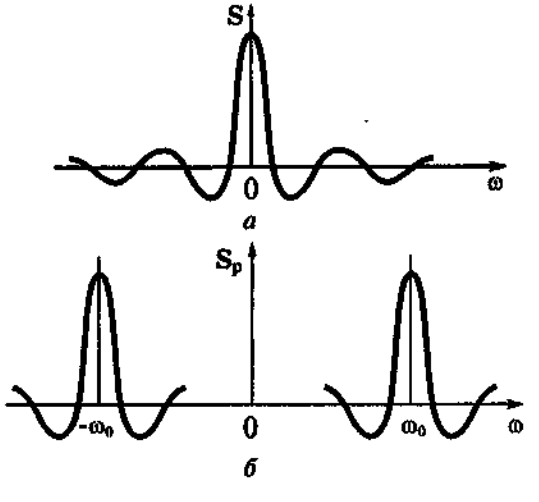



**Рис. 2.36.** Спектральные диаграммы:
*а* — видеоимпульса; *б* — радиоимпульса

Известно, что в радиовещании в последние годы наряду с монофоническими (от греч. *mónos* — один, единственный) все чаще используются стереофонические (от греч. *stereós* — объемный, пространственный) системы. Стереофоническое воспроизведение звука предусматривает раздельное формирование и передачу колебаний как минимум от двух микрофонов — левого и правого. По существу стереофоническое вещание дает представление о пространственном местонахождении источника звука и его перемещениях.

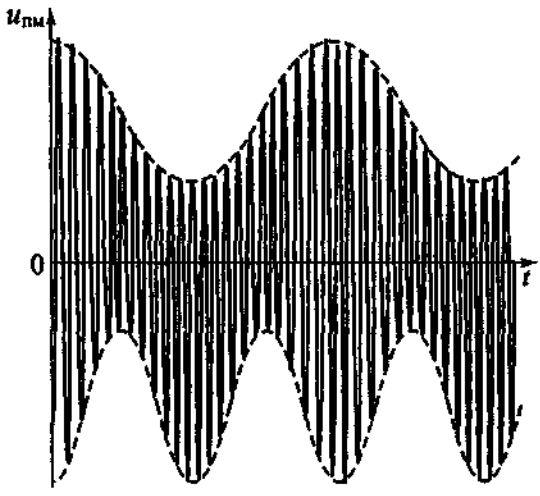



**Рис. 2.37.** Стереофонический сигнал с полярной модуляцией

Субъективное качество восприятия человеком стереофонической программы значительно выше, чем традиционной монофонической. Достаточно хороший стереоэффект получается при двухканальной передаче звука. В принципе стереофонический радиоканал можно получить, объединив два монофонических канала, по одному из которых передается сигнал $l(t) = E_l \cos\Omega_l t$ от левого источника звука, а по другому сигнал $r(t) = E_r \cos\Omega_r t$ от правого источника. Однако при этом не достигается *совместимость* стереофонической системы радиовещания с монофоническими приемниками, которые, возможно, будут находиться в эксплуатации еще сравнительно долгое время.

Совместимое с монофоническим стереофоническое радиовещание создают, используя несколько модифицированный вид амплитудной модуляции сигналов, называемый *полярной модуляцией*. Типичная осциллограмма сигнала с полярной модуляцией представлена на рис. 2.37. Как видно, получающийся сигнал (его называют *композитным стереосигналом*) благодаря наличию низкочастотных составляющих обладает интересным свойством: его верхняя (положительная) и нижняя (отрицательная) огибающие *не совпадают* по частоте. Более того, каждая из них с точностью до постоянной составляющей соответствует сигналу одного из стереоканалов: верхняя (положительные значения) огибающая — это левый канал, а нижняя (отрицательные значения) — правый. Такая структура передаваемого сигнала позволяет в радиоприемнике легко выделить из него сигналы стереоканалов. Для осуществления совместимого с монофоническим вещанием при стереовещании необходимо передавать сигналы $l(t)$ (левого) и $r(t)$ (правого) каналов по одной линии связи одновременно. При разработке системы такого вещания накладывается еще и требование совместимости с уже имеющимися монофоническими приемниками. Поэтому для стереовещания модифицируется низкочастотный *модулирующий* сигнал, который поступает на

вход модулятора передатчика (используется угловая модуляция). В стереофонической системе сигналы звука правого $l(t)$ и левого $r(t)$ каналов преобразуются в суммарно-разностном преобразователе в суммарные и разностные колебания. Низкочастотная составляющая сигнала модуляции, лежащая в звуковом диапазоне, для совместимости с монофоническими приемниками должна представлять монофонический сигнал, т. е. сумму сигналов левого и правого каналов:

$$u_м(t) = l(t) + r(t).$$

В области более высоких (ультразвуковых) частот модулирующего сигнала с помощью амплитудной модуляции передается дополнительный сигнал, позволяющий впоследствии выделить из полученной смеси сигналы $l(t)$ и $r(t)$ по отдельности. Поскольку монофонический сигнал — это сумма двух каналов, этим дополнительным компонентом, позволяющим восстановить исходные сигналы двух каналов, естественно надо выбрать их разность:

$$u_р(t) = l(t) - r(t).$$

Разностный сигнал модулирует по амплитуде вспомогательное высокочастотное колебание $u_п(t) = U_п\cos\omega_п t$, где $U_п$ и $\omega_п$ — амплитуда и частота. Сигнал $u_п(t)$ служит в данном случае поднесущим колебанием.

Сумма модулированной разностным сигналом поднесущей и монофонического сигнала $u_м(t)$ образует *полярно-модулированное* колебание. Для формирования такого сигнала используют *суммарно-разностный метод.*

Действительно, нужное свойство передаваемого колебания будет реализовано, если скомбинировать монофонический $u_м(t)$ и разностный $u_р(t)$ сигналы следующим образом:

$$u_{ПМ}(t) = U_п\left[M\frac{u_м(t)}{2} + \left(1 + M\frac{u_р(t)}{2}\right)\cos\omega_п t\right] = \tag{2.136}$$

$$= U_п\left[M\frac{l(t)+r(t)}{2} + \left(1 + M\frac{l(t)-r(t)}{2}\right)\cos\omega_п t\right],$$

где $M$ — коэффициент амплитудной модуляции, одинаковый в обоих каналах.

Легко проверить, что если $\cos\omega_п t = 1$, то $u_{ПМ}(t) = U_п[1 + Ml(t)]$, т. е. верхняя (положительной полярности) огибающая полярно-модулированного колебания действительно отражает сигнал $l(t)$ левого канала. Если же $\cos\omega_п t = -1$, то напряжение $u_{ПМ}(t) = U_п[1 - Mr(t)]$; значит, нижняя огибающая соответствует правому каналу $r(t)$. В отсутствие перемодуляции огибающие имеют разные знаки и могут быть без труда разделены в радиоприемнике.

В частности, выделяя, а затем складывая и вычитая колебания $u_м(t)$ и $u_р(t)$ можно восстановить сигналы левого и правого каналов:

$$l(t) = \frac{u_м(t) + u_р(t)}{2}; \qquad r(t) = \frac{u_м(t) - u_р(t)}{2}. \tag{2.137}$$

Использование поднесущей частоты. Необходимо заметить, что в показанном на рис. 2.37 виде сигнал для передачи не используется. Дело в том, что из-за наличия несущего колебания размах сигнала (разность между его максимальным и минимальным значениями) сильно увеличивается по сравнению с монофоническим сигналом, что технически нежелательно. Поэтому несущее колебание частично или полностью подавляют, так что разностный сигнал передается в режиме *перемодуляции* (кстати, это полезное свойство перемодуляции).

Непосредственно сигналы с полярной модуляцией в вещании нельзя применять, поскольку они содержат низкочастотную (звуковую) составляющую, пропорциональную полусумме колебаний левого и правого каналов. Поэтому переходят к двухступенчатому процессу. Вначале формируют сигнал с полярной модуляцией, выбрав в качестве частоты $\omega_n$ так называемую *поднесущую частоту*, лежащую значительно выше верхней границы передаваемого звукового диапазона. Потом проводят обычную частотную модуляцию несущего, частота которого соответствует метровому или дециметровому диапазону волн. В качестве модулирующего колебания используют полярно-модулированный сигнал, полученный на первом этапе. При детектировании такого колебания в обычном монофоническом приемнике полуразностный сигнал не воспроизводится, поскольку его спектр расположен в окрестности поднесущей частоты, т.е. за пределом звукового диапазона. Полусуммарный сигнал поступает на выход приемника и обеспечивает вполне приемлемое качество звучания. Таким образом, требования совместимости систем вещания оказываются выполненными.

Стандарт стереофонического радиовещания Российской Федерации и зарубежных стран отличаются друг от друга. Принятый у нас стандарт устанавливает частоту поднесущего колебания равной 31,25 кГц. Для эффективного использования радиоканала в передатчике поднесущее колебание ослабляют по амплитуде в 5 раз, а в приемнике за счет специальных схемных решений восстанавливают этот сигнал до исходного уровня. Такая мера связана с тем, что по стандарту девиация частоты (т. е. изменение частоты) не может превышать 50 кГц. Если чрезмерно использовать этот ресурс для передачи вспомогательного поднесущего колебания, то возможно ухудшение качества воспроизведения звуковых колебаний и, прежде всего, снижение их громкости.

Американский и западноевропейский стандарты решают эту задачу несколько по-другому. Частота поднесущего колебания выбирается в этих стандартах равной 38 кГц. Это колебание подавляется в передатчике полностью, а вместо него в состав модулирующего сигнала вводят так называемый *пилот-тон* (*pilot tone*) на частоте 19 кГц. В приемнике частота пилот-тона удваивается, и за счет этого происходит восстановление поднесущего колебания.

### *Радиосигналы с угловой модуляцией*

Обратимся к аналоговым модулированным сигналам, полученным путем изменения по закону передаваемого сообщения в несущем колебании (2.112) частоты $\omega_0$, или начальной фазы $\varphi_0$. Поскольку в этих случаях происходит *линейное влияние* на аргумент функции cos, а сам аргумент гармонического коле-


Э. Армстронг

бания $\psi(t) = \omega_0 t + \varphi_0$ определяет мгновенное значение фазы (фазового угла), такие сигналы получили название *сигналов с угловой модуляцией* (УМ; *angle modulation*). Если в несущем колебании изменяется частота $\omega_0$, то имеем дело с *частотной модуляцией* (ЧМ), если же изменяется начальная фаза $\varphi_0$ — *фазовой модуляцией* (ФМ). По существу различие между фазовой и частотной модуляцией заключается лишь в том, как именно мгновенная фаза $\psi(t)$ связана с модулирующим сигналом. Чтобы оценить связь частотной модуляции с фазовой, введем понятие *мгновенной частоты* (*instantaneous frequency*), равной производной от полной фазы по времени $\omega(t) = d\psi/dt$.

### Частотная модуляция

Передачу сигналов методом частотной модуляции изобрел в 1902 г. К. Эретом (США). Начало широкого практического применения ЧМ в радиовещании положил в 1935 г. Эдвин Армстронг (Edwin Armstrong; 1890–1954; США).

При частотной модуляции (*frequency modulation*; FM) мгновенное значение несущей частоты $\omega(t)$ связано с модулирующим сигналом $e(t)$ зависимостью

$$\omega(t) = \omega_0 + k_ч e(t), \tag{2.138}$$

здесь $k_ч$ — размерный коэффициент пропорциональности между частотой и напряжением, рад/(В·с).

Рассмотрим *однотональную частотную модуляцию*, когда модулирующий сигнал — гармоническое колебание $e(t) = E_0 \cos\Omega t$, у которого для упрощения начальная фаза $\theta_0 = 0$. Пусть и начальная фаза несущего колебания $\varphi_0 = 0$. При необходимости начальные фазы $\theta_0$ и $\varphi_0$ легко можно ввести в окончательные соотношения. Полную фазу ЧМ-сигнала в любой момент времени $t$ определим путем интегрирования мгновенной частоты, выраженной через формулу (2.138),

$$\psi(t) = \int_0^t \omega(t)dt = \int_0^t (\omega_0 + k_ч E_0 \cos\Omega t)dt = \omega_0 t + \frac{\omega_{дч}}{\Omega}\sin\Omega t = \omega_0 t + m_ч \sin\Omega t. \tag{2.139}$$

В формуле (2.139)

$$\omega_{дч} = k_ч E_0 \tag{2.140}$$

— максимальное отклонение частоты от значения $\omega_0$, или *девиация частоты* (*frequency deviation*) при частотной модуляции;

$$m_ч = \omega_{дч}/\Omega = k_ч E_0/\Omega, \tag{2.141}$$

— максимальное отклонение от текущей фазы $\omega_0 t$ или *девиация фазы* несущего колебания называется *индексом частотной модуляции* (*index of frequency modulation*). Данный парамер определяет интенсивность колебаний начальной фазы радиосигнала.

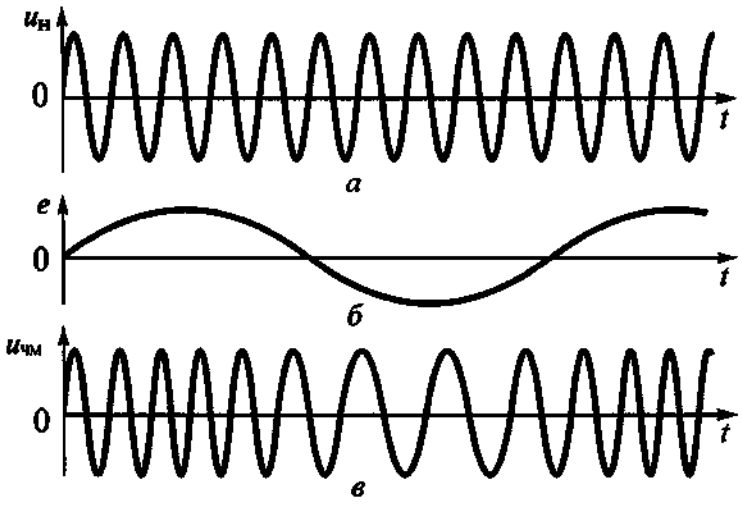



**Рис. 2.38.** Частотная однотональная модуляция:
*а* — несущее колебание; *б* — модулирующий сигнал; *в* — ЧМ-сигнал

С учетом соотношений (2.112) и (2.139) частотно-модулированный сигнал запишется в следующем виде:

$$u_{ЧМ}(t) = U_н\cos\psi(t) = U_н\cos(\omega_0 t + m_ч\sin\Omega t). \qquad (2.142)$$

На рис. 2.38 представлены упрощенные временны́е диаграммы сигналов при частотной модуляции. Несущее колебание $u_н(t)$ с начальной фазой $\theta_0 = 90°$ показано на рис. 2.38, *а*, модулирующий сигнал $e(t)$ с начальной фазой $\varphi_0 = 90°$ — на рис. 2.38, *б*, и полученный в процессе частотной модуляции ЧМ-сигнал $u_{ЧМ}(t)$ — на рис. 2.38, *в*.

Нетрудно заметить, что по форме временна́я диаграмма ЧМ-сигнала напоминает сжатые и растянутые меха русской гармошки. Форма и уровень «сжатия» или «растягивания» несущего колебания зависят от значения *индекса частотной модуляции* $m_ч$.

**Спектр ЧМ-сигнала при однотональной модуляции.** Используя известные тригонометрические преобразования, запишем (2.142) следующим образом (здесь и далее индекс у коэффициента модуляции $m_ч$ опущен, т. е. $m_ч = m$)

$$u_{ЧМ}(t) = U_н\cos(\omega_0 t + m\sin\Omega t) =$$

$$= U_н\cos(m\sin\Omega t)\cos\omega_0 t - U_н\sin(m\sin\Omega t)\sin\omega_0 t. \qquad (2.143)$$

Проанализируем выражение (2.143) отдельно для малых ($m << 1$) и больших ($m > 1$) индексов модуляции.

Спектр ЧМ-сигнала при $m << 1$ (такую угловую модуляцию называют *узкополосной*). В этом случае имеют место приближенные равенства:

$$\cos(m\sin\Omega t) \approx 1; \qquad \sin(m\sin\Omega t) \approx m\sin\Omega t. \qquad (2.144)$$

Подставив формулы (2.144) в выражение (2.143), после несложных математических преобразований получим (при начальных фазах модулирующего и несущего колебаний $\theta_0 = 0$ и $\varphi_0 = 0$):

$$u_{\text{ЧМ}}(t) = U_{\text{н}} \cos\omega_0 t - U_{\text{н}} m \sin\Omega t \cdot \sin\omega_0 t = \\ = U_{\text{н}} \cos\omega_0 t + \frac{mU_{\text{н}}}{2}\cos(\omega_0 + \Omega)t - \frac{mU_{\text{н}}}{2}\cos(\omega_0 - \Omega)t. \quad (2.145)$$

Сравнение формул (2.145) и (2.119) показывает, что по аналитической записи спектр ЧМ-сигнала при однотональной модуляции напоминает спектр АМ-сигнала и также состоит из несущего колебания и двух боковых составляющих с частотами $(\omega_0 + \Omega)$ и $(\omega_0 - \Omega)$, причем и амплитуды их рассчитываются аналогично (только вместо коэффициента амплитудной модуляции $M$ в формуле для ЧМ-сигнала фигурирует индекс угловой модуляции $m$). Но есть и принципиальное отличие, превращающее амплитудную модуляцию в частотную, знак минус перед одной из боковых составляющих.

Не вдаваясь в особые подробности отметим: чтобы превратить однотональный АМ-сигнал в однотональный ЧМ-сигнал, достаточно в первом изменить на 180° начальную фазу одной из боковых частот. Еще один способ превратить АМ-сигнал в ЧМ-сигнал — изменить на 90° начальную фазу спектральной составляющей несущей. На рис. 2.39 показана спектральная диаграмма для однотонального ЧМ-сигнала при индексе модуляции $m \ll 1$. Отметим очевидное: ширина спектра ЧМ-сигнала, как и в спектра при АМ-сигнала, равна $2\Omega$.

Построим векторную диаграмму ЧМ-сигнала (рис. 2.40) аналогично тому, как это было сделано для однотональной амплитудной модуляции. На диаграмме показано, как поворот фазы вектора АД нижней боковой составляющей $U_{\text{НБ}}$ на 180° влияет на вектор результирующего колебания 0В, который изменяет свое направление относительно вектора несущей 0А. Такое изменение направления вектора 0В характерно для угловой модуляции и это отличает ее от амплитудной. В то же время изменение направления вектора АД на 180° не влияет на вектор модуляции АВ, который всегда перпендикулярен вектору несущей 0А (для сравнения направление вектора АД нижней боковой составляющей при АМ-сигнале обозначено штриховой линией — вектор АД′).

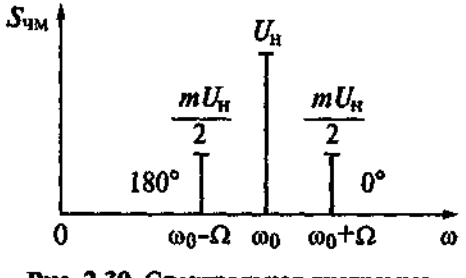



**Рис. 2.39.** Спектральная диаграмма ЧМ-сигнала при $m \ll 1$

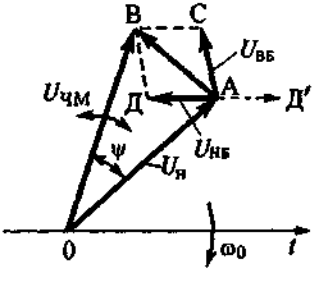



**Рис. 2.40.** Векторная диаграмма ЧМ-сигнала при $m \ll 1$


Ф. Бессель

Итак, вектор результирующего ЧМ-колебания 0В изменяется по фазе, т. е. с течением времени «качается» вокруг центрального положения. При этом вектор изменяется и по амплитуде, хотя теоретически этого не должно быть. Этот нюанс связан с тем, что выражение (2.145), по которому производится построение векторной диаграммы, является приближенной, поскольку формула (2.144) имеет достаточно приближенный характер. Однако при $m << 1$ изменения амплитуды вектора результирующего колебания 0В настолько малы, что ими можно пренебречь и модуляцию рассматривать как чисто частотную.

Спектр ЧМ-сигнала при $m > 1$. Этот случай представляет основной практический интерес, поскольку при больших $m$ помехоустойчивость передачи сигнала существенно выше, чем при амплитудной модуляции. Здесь при расчетах оказывается удобным аппарат функций Ф. Бесселя (*F. Bessel*; 1784–1846; немецкий математик и астроном). Из математики известно, что эти функции (функции Бесселя являются собственными решениями дифференциального уравнения мембраны) можно косвенно определить так

$$\cos(m\sin\Omega t) = J_0(m) + 2\sum_{n=1}^{\infty} J_{2n}(m)\cos 2n\Omega t, \qquad (2.146)$$

$$\sin(m\sin\Omega t) = 2\sum_{n=0}^{\infty} J_{2n+1}(m)\sin(2n+1)\Omega t, \qquad (2.147)$$

где $J_n(m)$ — функция Бесселя 1-го рода $n$-го порядка.

В теории функций Бесселя доказывается, что функции с положительными и отрицательными индексами связаны между собой формулой

$$J_{-n}(m) = (-1)^n J_n(m). \qquad (2.148)$$

Ряды (2.146) и (2.147) подставим в формулу (2.143), а затем заменим произведение косинусов и синусов полусуммами косинусов соответствующих аргументов. Тогда, с учетом (2.148), получим следующее выражение для ЧМ-сигнала

$$u_{\text{ЧМ}}(t) = U_{\text{н}} J_0(m)\cos\omega_0 t + \sum_{n=1}^{\infty} U_{\text{н}} J_n(m)\cos(\omega_0 + n\Omega)t +$$

$$+ \sum_{n=1}^{\infty} (-1)^n U_{\text{н}} J_n(m)\cos(\omega_0 - n\Omega)t. \qquad (2.149)$$

Итак, спектр ЧМ-сигнала с однотональной модуляцией при индексе модуляции $m > 1$ состоит из множества высокочастотных гармоник: несущего колебания и бесконечного числа боковых составляющих с частотами $\omega_0 + n\Omega$ и $\omega_0 - n\Omega$, расположенными попарно и симметрично относительно несущей частоты $\omega_0$.

При этом, исходя из (2.148), можно отметить, что начальные фазы боковых колебаний с частотами $\omega_0 + n\Omega$ и $\omega_0 - n\Omega$ совпадают, если $n$ — четное число, и отличаются на 180°, если $n$ — нечетное.

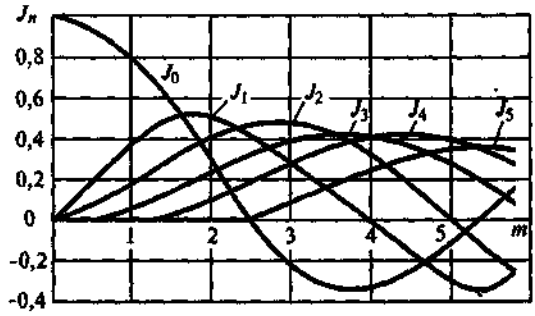



**Рис. 2.41.** Графики функций Бесселя

Теоретически спектр ЧМ-сигнала (так же и ФМ-сигнала) бесконечен, однако в реальных случаях он ограничен. Для детального анализа и построения спектральных диаграмм необходимо знать поведение функций $J_n(m)$ при различных $m$ в зависимости от $n$. На рис. 2.41 приведены графики нескольких функций Бесселя. Отметим, что чем больше индекс функции Бесселя, тем протяженнее область аргументов, при которых функция очень мала. Поэтому считают, что, начиная с номера $n \geq m + 1$, значения функций Бесселя становятся весьма малыми и практическая ширина спектра сигналов с угловой модуляцией

$$\Delta\omega_c = \Delta\omega_{ум} = 2(m + 1)\Omega. \qquad (2.150)$$

ЧМ- и ФМ-сигналы, применяемые на практике в радиотехнике и связи, имеют индекс модуляции $m >> 1$, поэтому

$$\Delta\omega_c = \Delta\omega_{ум} = 2m\Omega = 2\omega_д. \qquad (2.151)$$

Полоса частот ЧМ-сигнала с однотональной модуляцией равна удвоенной девиации частоты и не зависит от частоты модуляции. Спектр ЧМ-сигнала с угловой модуляцией при негармоническом модулирующем сигнале определить трудно. Но он всегда сложнее, чем спектр АМ-сигнала при том же модулирующем сигнале. Ширина спектра также больше, чем при амплитудной модуляции. Спектр простейшего ЧМ-сигнала при $m = 3$ показан на рис. 2.42.

### Фазовая модуляция

В ФМ-сигнале полная фаза несущего колебания изменяется пропорционально модулирующему сигналу

$$\psi(t) = \omega_0 t + k_ф e(t), \qquad (2.152)$$

где $k_ф$ — размерный коэффициент пропорциональности, рад/В.




**Рис. 2.42.** Спектр простейшего ЧМ-сигнала

При однотональной модуляции фаза несущего колебания определяется следующим соотношением

$$\psi(t) = \omega_0 t + k_\phi E_0 \cos\Omega t. \qquad (2.153)$$

Из (2.153) следует, что, как и в случае частотной модуляции, полная фаза несущего колебания при фазовой модуляции изменяется по гармоническому закону. Максимальное отклонение фазы несущего колебания от начальной фазы характеризует *индекс фазовой модуляции*

$$m_\phi = k_\phi E_0. \qquad (2.154)$$

Подставляя формулы (2.153) и (2.154) в соотношение для несущей (2.112), получим следующую запись ФМ-сигнала

$$u_{\Phi M}(t) = U_н \cos(\omega_0 t + k_\phi E_0 \cos\Omega t) = U_н \cos(\omega_0 t + m_\phi \cos\Omega t). \qquad (2.155)$$

Дифференцируя формулу (2.153) по времени, находим мгновенную частоту ФМ-сигнала

$$\omega(t) = \psi(t)/dt = \omega_0 - m_\phi \Omega \sin\Omega t = \omega_0 - \omega_{д\phi} \sin\Omega t, \qquad (2.156)$$

где $\omega_{д\phi} = m_\phi \Omega = k_\phi E_0 \Omega$ — максимальное отклонение частоты от значения несущей $\omega_0$, т. е. *девиация частоты* при фазовой модуляции.

**Пример 2.6.** ФМ-сигнал с амплитудой $U_н = 5$ В и несущей частотой $f_0 = 200$ МГц модулирован однотональным гармоническим колебанием с частотой $F = 20$ кГц при индексе модуляции $m = 10$. Записать выражение для ФМ-сигнала, определить пределы, в которых изменяется частота, и рассчитать ширину спектра.

Р е ш е н и е. Воспользовавшись формулой (2.155), запишем ФМ-сигнал

$$u_{\Phi M}(t) = 5\cos[4\pi \cdot 10^8 t + 10\cos(4\pi \cdot 10^4 t)].$$

Определим девиацию циклической частоты

$$F_д = \omega_д/(2\pi) = mF = 0{,}2 \text{ МГц}.$$

Значит, при фазовой модуляции мгновенная частота изменяется в пределах:

$$f_{min} = 100 - 0{,}2 = 99{,}8 \text{ МГц};$$

$$f_{max} = 100 + 0{,}2 = 100{,}2 \text{ МГц}.$$

Ширина спектра ФМ-сигнала (2.150) будет

$$\Delta f_{\Phi M} = \Delta\omega_{\Phi M}/(2\pi) = 2(m+1)F = 440 \text{ кГц}.$$

В сводной табл. 2.2 показано, как связаны с модулирующим однотональным сигналом различные характеристики модулированного колебания при фазовой и частотной модуляции. Наглядное представление о законах изменения частоты и фазы при частотной и фазовой однотональной модуляции дают графические построения, представленные на рис. 2.43.

Таблица 2.2. **Сигналы с угловой модуляцией**

| Закон модуляции | $e(t) = E_0\cos\Omega t$ | |
|---|---|---|
| Вид сигнала | ЧМ | ФМ |
| Аналитическая запись | $u_{ЧМ}(t) = U_н\cos(\omega_0 t + m_ч\sin\Omega t)$ | $u_{ФМ}(t) = U_н\cos(\omega_0 t + m_ф\cos\Omega t)$ |
| Изменение частоты | $\omega(t) = \omega_0 + \omega_{дч}\cos\Omega t$ | $\omega(t) = \omega_0 - \omega_{дф}\sin\Omega t$ |
| Девиация частоты | $\omega_{дч} = k_ч E_0$ | $\omega_{дф} = k_ф E_0\Omega$ |
| Индекс модуляции | $m_ч = k_ч E_0/\Omega$ | $m_ф = k_ф E_0$ |
| Отклонение фазы | $\varphi(t) = m_ч\sin\Omega t$ | $\varphi(t) = m_ф\cos\Omega t$ |

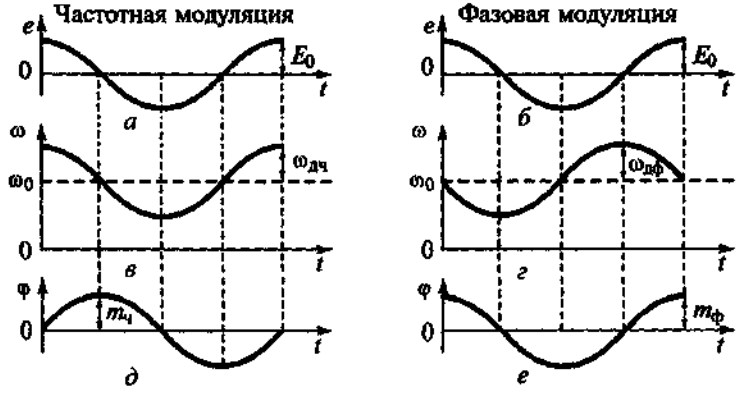



**Рис. 2.43.** Графики изменения частоты и фазы при угловой однотональной модуляции:
*а, б* — модулирующий сигнал; *в, г* — частота; *д, е* — фаза

Выражения (2.145), (2.155) и приведенные на рис. 2.43 соответствующие им графики показывают, что при однотональной угловой модуляции невозможно определить, является ли сигнал частотно- или фазомодулированным? Различия между этими достаточно близкими видами однотональной угловой модуляции проявляются только при изменении значений амплитуды $E_0$ или частоты $\Omega$ модулирующего сигнала $e(t)$. При частотной модуляции *девиация частоты* $\omega_{дч}$ *пропорциональна амплитуде* $E_0$ и *не зависит от частоты* $\Omega$ модулирующего сигнала $e(t) = E_0\cos\Omega t$. *Индекс же модуляции* $m_ч$ *прямо пропорционален амплитуде* $E_0$ и *обратно пропорционален частоте* $\Omega$ модулирующего сигнала. При фазовой модуляции девиация частоты $\omega_{дф}$ изменяется пропорционально амплитуде и частоте модулирующего сигнала. Индекс модуляции $m_ф$ пропорционален амплитуде $E_0$ и не зависит от частоты $\Omega$ модулирующего сигнала.

Из проведенного данных сигналов анализа можно сделать еще два важнейших вывода:

- если пропустить модулирующий сигнал через идеальное *дифференцирующее* устройство, а затем подать его на *частотный модулятор*, получится радиосигнал с *фазовой* модуляцией (верхняя ветвь на рис. 2.44);

• если пропустить модулирующий сигнал через *идеальный интегратор* и подать его на *фазовый модулятор*, получим ЧМ-сигнал (нижняя ветвь на рис. 2.44).

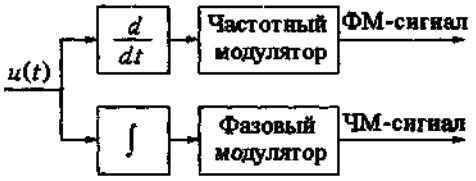



**Рис. 2.44.** Формирование сигналов с угловой модуляцией

Спектр ЧМ- и ФМ-сигналов при произвольном модулирующем сигнале. В отличие от амплитудной модуляции, при угловой модуляции получить простое аналитическое выражение для спектра радиосигнала при произвольном модулирующем сигнале не удается. Даже попытка рассмотреть спектр при двухтональном модулирующем сигнале делает аналитические выкладки намного сложнее, чем при однотональной модуляции.

**Сравнение помехоустойчивости радиосистем с амплитудной и угловой модуляцией.** Следует отметить, что радиосигналы с угловой модуляцией имеют ряд важных преимуществ перед амплитудно-модулированными колебаниями.

1. Поскольку при угловой модуляции амплитуда модулированных колебаний не несет в себе никакой информации и не требуется ее постоянства (в отличие от амплитудной модуляции), то практически любые вредные нелинейные изменения амплитуды радиосигнала в процессе осуществления связи не приводят к заметному искажению передаваемого сообщения.

2. Постоянство амплитуды радиосигнала при угловой модуляции позволяет полностью использовать энергетические возможности генератора несущей частоты, который работает при неизменной средней мощности колебаний.

Все сказанное справедливо лишь при следующем условии: амплитуда полезного сигнала на входе частотного детектора приемника должна значительно превышать среднее квадратическое значение напряжения шума. В противном случае неизбежен так называемый *пороговый эффект* и может оказаться, что ЧМ-система при малом отношении сигнал/шум на входе будет функционировать хуже, чем аналогичная система с АМ-сигналами. Кроме того, для реализации отмеченных преимуществ ЧМ- и ФМ-колебаний необходимо отводить конкретному радиосигналу слишком широкую полосу частот, значительно превышающую ширину спектра модулирующей функции.

Заканчивая краткий обзор некоторых вопросов оценки аналоговых модулированных сигналов, отметим очень важный в практическом отношении вопрос о сравнительных характеристиках двух наиболее распространенных видов модуляции. Широкополосность ЧМ- и ФМ-сигналов приводит к гораздо большей помехоустойчивости колебаний с угловой модуляцией по сравнению с АМ-сигналами. Однако именно широкополосность обусловливает их применимость для целей радиосвязи лишь на очень высоких частотах — в диапазонах метровых и более коротких волн. Заметим также, что в аналоговых радиорелейных и спутниковых системах передачи информации, у которых уровень помех в линии весьма значителен, применяются частотная или фазовая модуляция с индексами модуляции $m > 1$ как наиболее помехоустойчивые виды модуляции.

**Выбор несущих частот при аналоговой модуляции.** Реальная полоса пропускания канала связи не должна существенно превосходить ширину спектра радиосигнала, чтобы передаче полезной информации не мешали сигналы других радиостанций и различные внутренние и внешние помехи. В частности, при использовании амплитудной модуляции несущая частота более чем в 10 раз должна превосходить максимальную частоту $\Omega_N$ передаваемого сигнала. Это обусловлено необходимостью:

- легкого разделения несущей и модулирующих частот при детектировании АМ-радиосигналов в приемнике;
- малости относительной полосы частот занимаемой модулированным колебанием для излучения всех ее спектральных составляющих простейшей передающей антенной.

Для того чтобы разместить все возрастающее количество радиовещательных и телевизионных станций в диапазонах с заданным отношением максимальной $f_{0max}$ и минимальной $f_{0min}$ несущих частот, требуется увеличивать несущие частоты излучения.

**Пример 2.7.** Ширина спектра АМ-сигнала $\Delta f_{AM} = \Delta f_{треб} = \Delta\Omega_{AM}/(2\pi) = 9$ кГц, а отношение несущих частот $f_{0max}/f_{0min} = 10$. Определить, сколько радиостанций может одновременно работать в заданной полосе частот.

Р е ш е н и е. При $f_{0min} = 30$ кГц в этом диапазоне можно разместить

$$n = \frac{f_{max} - f_{min}}{\Delta\Omega/(2\pi)} = \frac{9 f_{min}}{\Delta f_{треб}} \approx 30 \text{ радиостанций.}$$

При $f_{0min} = 300$ кГц можно разместить уже 300 радиостанций.

Так как в данном спектральном диапазоне могут быть размещены несущие частоты только конечного числа радиостанций, не мешающих друг другу, существуют международные соглашения и соответствующее законодательство Российской Федерации, определяющее распределение несущих частот между разными видами систем радиосвязи и радиовещания, и для выхода в эфир новой радиостанции требуется разрешение Государственной комиссии по радиочастотам Федерального агентства связи.

### *Сигналы с линейной частотной модуляцией*

Сигналы с внутриимпульсной частотной модуляцией относятся к особому классу модулированных сигналов, широко применяющихся в радиолокации, системах связи, измерительной технике и других областях. Они привлекли внимание специалистов прежде всего в связи с поисками способов сжатия импульсных радиолокационных сигналов для целей повышения разрешающей способности РЛС. Эти сигналы отличаются от обычных радиоимпульсов (отрезков несущего колебания) тем, что их высокочастотное заполнение имеет переменную частоту. Чаще всего используется внутриимпульсная частотная модуляция с линейным законом изменения мгновенной частоты во времени.

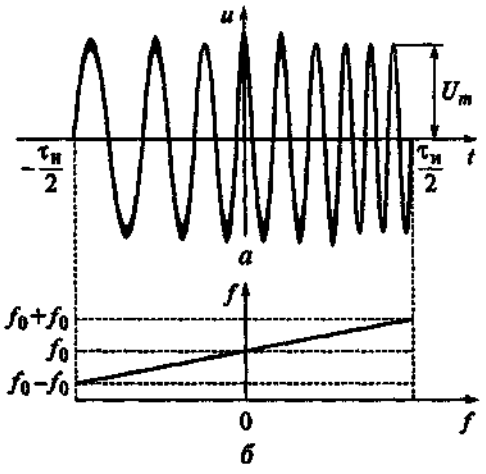



**Рис. 2.45.** Радиоимпульс с ЛЧМ:
*а* — графическое представление сигнала; *б* — закон изменения частоты

**Принцип создания сигналов с линейной частотной модуляцией.** При модуляции частоты несущего колебания по негармоническому закону определение спектра сигнала резко усложняется. Проведем анализ часто применяемого в радиолокации радиоимпульса с огибающей прямоугольной формы, у которого частота заполнения (несущая) линейно нарастает от начала импульса к его концу (рис. 2.45).

Конкретизируя математическую модель сигнала с переменной частотой заполнения, предположим, что его длительность равна $\tau_и$, причем середине импульса соответствует точка $t = 0$, а мгновенная частота изменяется по линейному закону (рис. 2.45, *а*):

$$\omega(t) = \omega_0 + \alpha t. \qquad (2.157)$$

Здесь $\omega_0$ — несущая частота немодулированного сигнала $u_н(t) = U_н\cos(\omega_0 t + \varphi_0)$; $\alpha$ — параметр [$с^{-2}$], характеризующий скорость изменения (нарастания) частоты модулированного колебания.

Нетрудно подсчитать, что за время, равное длительности импульса, девиация частоты (рис. 2.45, *б*) составит:

$$\omega_д = 2\pi f_д = \alpha\tau_и. \qquad (2.158)$$

Для упрощения анализа примем $\varphi_0 = 0$, поскольку наличие этого сдвига несущественно. Тогда интегрирование (2.157) даст мгновенную (полную) фазу сигнала в любой момент времени по длительности импульса:

$$\psi(t) = \omega_0 t + \alpha t^2/2. \qquad (2.159)$$

Итак, *радиоимпульсом с линейной частотной модуляцией* (*ЛЧМ-сигналом, ЛЧМ-импульсом*) назовем сигнал, описываемый следующими соотношениями:

$$u_{\text{ЛЧМ}}(t)=\begin{cases}0, & t<-\dfrac{\tau_{\text{и}}}{2};\\ U_m\cos\left(\omega_0 t+\dfrac{\alpha t^2}{2}\right), & -\dfrac{\tau_{\text{и}}}{2}\le t\le\dfrac{\tau_{\text{и}}}{2};\\ 0, & t>\dfrac{\tau_{\text{и}}}{2}.\end{cases} \qquad (2.160)$$

Отметим очень полезное свойство ЛЧМ-сигналов. Пусть некоторое электронное устройство осуществляет временну́ю задержку входного сигнала, величина которой зависит от частоты колебаний. Если с ростом частоты время задержки уменьшается, то при определенных условиях, подавая на вход устройства импульс достаточно большой длительности, можно «сжать» его во времени. Этот эффект обусловлен тем, что на выходе устройства задержки как низкочастотные составляющие, относящиеся к началу импульса, так и более высокочастотные, находящиеся в его конце, будут появляться одновременно.

Спектр ЛЧМ-импульса прямоугольной формы. Используя аналитическую модель (2.160) и формулу Эйлера, запишем выражение спектральной плотности одиночного ЛЧМ-импульса:

$$\begin{aligned}\mathrm{S}(\omega)&=U_{\text{и}}\int_{-\tau_{\text{и}}/2}^{\tau_{\text{и}}/2}\cos(\omega_0 t+\alpha t^2/2)\,\mathrm{e}^{-j\omega t}dt=\\ &=\frac{U_{\text{и}}}{2}\int_{-\tau_{\text{и}}/2}^{\tau_{\text{и}}/2}\mathrm{e}^{j\left[(\omega_0-\omega)t+\alpha t^2/2\right]}dt+\frac{U_{\text{и}}}{2}\int_{-\tau_{\text{и}}/2}^{\tau_{\text{и}}/2}\mathrm{e}^{-j\left[(\omega_0+\omega)t+\alpha t^2/2\right]}dt.\end{aligned} \qquad (2.161)$$

Первый интеграл в правой части (2.161) описывает спектральную плотность с резко выраженным максимумом в области положительных частот, близких к $\omega_0$. Второй интеграл соответствует всплеску спектральной плотности, сосредоточенной вблизи частоты $\omega = -\omega_0$. На практике интерес представляет случай, когда эффект перекрытия отмеченных спектров, концентрирующихся при положительных и отрицательных частотах, невелик и близок к нулю. Это связано с тем, что девиация частоты $\omega_{\text{д}}$ за время длительности импульса очень мала по сравнению с несущей частотой $\omega_0$:

$$\omega_{\text{д}}=\alpha\tau_{\text{и}}<<\omega_0.$$

Поэтому в формуле (2.161) достаточно вычислить только первый интеграл, дающий спектральную плотность при частотах $\omega > 0$. Аргумент экспоненциальной функции в первом слагаемом формулы (2.161) целесообразно дополнить до полного квадрата разности. Тогда после несложных преобразований получим

$$\mathrm{S}(\omega)=\frac{U_{\text{и}}}{2}\mathrm{e}^{-j\frac{(\omega-\omega_0)^2}{2\alpha}}\int_{-\tau_{\text{и}}/2}^{\tau_{\text{и}}/2}\mathrm{e}^{j\frac{\alpha}{2}\left(t-\frac{\omega-\omega_0}{\alpha}\right)^2}dt. \qquad (2.162)$$

Перейдем от переменной $t$ к новому аргументу $x$, выполнив замену

$$\sqrt{\alpha}\left(t-\frac{\omega-\omega_0}{\alpha}\right)=\sqrt{\pi}x. \tag{2.163}$$

После вычислений запишем

$$\mathrm{S}(\omega)=\frac{U_{\text{н}}}{2}\sqrt{\frac{\pi}{\alpha}}\mathrm{e}^{-j\frac{(\omega-\omega_0)^2}{2\alpha}}\int_{-X_1}^{X_2}\mathrm{e}^{j\frac{\pi x^2}{2}}dx, \tag{2.164}$$

где пределы интегрирования

$$X_1=\frac{\frac{\alpha\tau_{\text{н}}}{2}+(\omega-\omega_0)}{\sqrt{\pi\alpha}};\qquad X_2=\frac{\frac{\alpha\tau_{\text{н}}}{2}-(\omega-\omega_0)}{\sqrt{\pi\alpha}}. \tag{2.165}$$

Для дальнейшего анализа ЛЧМ-сигнала используем известные интегралы О. Френеля (O. Fresnel; французский физик; 1788–1827):

$$C(x)=\int_0^x\cos\frac{\pi y^2}{2}dy;\qquad S(x)=\int_0^x\sin\frac{\pi y^2}{2}dy. \tag{2.166}$$

Подставив формулы (2.165) и (2.166) в (2.164), получим окончательное выражение для спектральной плотности ЛЧМ-сигнала

$$\mathrm{S}(\omega)=\frac{U_{\text{н}}}{2}\sqrt{\frac{\pi}{\alpha}}\mathrm{e}^{-j\frac{(\omega-\omega_0)^2}{2\alpha}}\left\{C(X_1)+C(X_2)+j[S(X_1)+S(X_2)]\right\}. \tag{2.167}$$

Представим соотношение (2.167) в показательной форме

$$\mathrm{S}(\omega)=|\,\mathrm{S}(\omega)\,|\mathrm{e}^{j\Psi(\omega)},$$

где

$$\left|\mathrm{S}(\omega)\right|=\frac{U_{\text{н}}}{2}\sqrt{\frac{\pi}{\alpha}\left\{[C(X_1)+C(X_2)]^2+[S(X_1)+S(X_2)]^2\right\}} \tag{2.168}$$

– амплитудный спектр (модуль спектральной плотности);

$$\Psi(\omega)=\Psi_1(\omega)+\Psi_2(\omega)=-\frac{(\omega-\omega_0)^2}{2\alpha}+\operatorname{arctg}\frac{S(X_1)+S(X_2)}{C(X_1)+C(X_2)} \tag{2.169}$$

– фазовый спектр;

$$\Psi_1(\omega)=-(\omega-\omega_0)^2/(2\alpha). \tag{2.170}$$

Составляющую фазового спектра ЛЧМ-сигнала

$$\Psi_2(\omega)=\operatorname{arctg}\frac{S(X_1)+S(X_2)}{C(X_1)+C(X_2)} \tag{2.171}$$

называют *остаточным фазовым членом*.

**ЛЧМ-сигналы с большой базой.** Анализ полученных выражений показывает, что характер частотной зависимости модуля и фазы спектральной плотности одиночного прямоугольного ЛЧМ-импульса связан с следующим безразмерным числом

$$B = f_д \tau_и = \frac{\alpha \tau_и^2}{2\pi}, \qquad (2.172)$$

равным произведению девиации частоты $f_д$ на длительность импульса $\tau_и$ и называемым *базой* ЛЧМ-сигнала.

Упомянутые особенности ЛЧМ-сигналов связаны с рядом обстоятельств.

Во-первых, амплитудный спектр здесь практически постоянен в пределах полосы частот шириной $(-\omega_д/2 \ldots \omega_д/2)$ с центром в точке $\omega_0$. Соответствующие графики амплитудного спектра и остаточного фазового члена, построенные по формулам (2.168) и (2.171), представлены на рис. 2.46.

Во-вторых, осцилляции амплитудного спектра существенно уменьшаются с увеличением базы ЛЧМ-сигнала вплоть до теоретически полного исчезновения.

Несложный анализ (2.168) показывает, что на центральной частоте амплитудного спектра $\omega_0$

$$|S(\omega_0)| = U_и \sqrt{\frac{\pi}{2\alpha}}. \qquad (2.173)$$

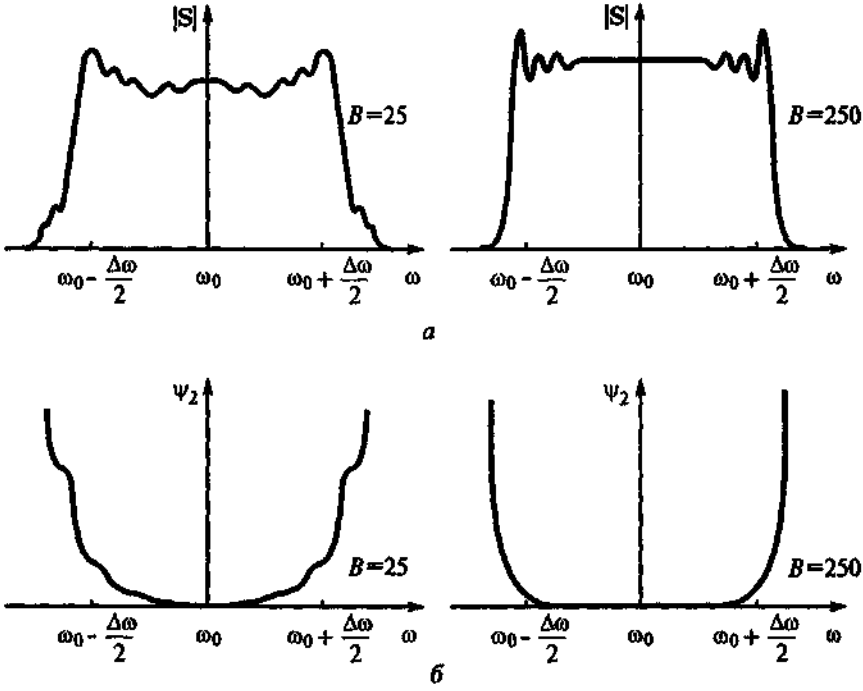



**Рис. 2.46.** Спектральная плотность ЛЧМ-сигнала при различных значениях базы:
*а* — амплитудный спектр; *б* — частотная зависимость остаточного фазового члена

Таким образом, амплитудный спектр ЛЧМ-сигнала с большой базой

$$|\mathbf{S}(\omega)| = \begin{cases} 0, & \omega < \omega_0 - \dfrac{\omega_д}{2}; \\ U_и\sqrt{\dfrac{\pi}{2\alpha}}, & \omega_0 - \dfrac{\omega_д}{2} \le \omega \le \omega_0 + \dfrac{\omega_д}{2}; \\ 0, & \omega > \omega_0 + \dfrac{\omega_д}{2}. \end{cases} \qquad (2.174)$$

Энергетический спектр такого сигнала, определяемый согласно (2.106) как

$$\mathbf{W}_и(\omega) = |\mathbf{S}(\omega)|^2 = \frac{\pi}{2\alpha}U_и^2, \qquad (2.175)$$

также постоянен в полосе частот ($\omega_0 - \omega_д/2$, $\omega_0 + \omega_д/2$) и становится практически равным нулю вне этой полосы.

**Пример 2.8.** Определить параметры спектра прямоугольного ЛЧМ-импульса, у которого амплитуда $U_и = 10$ В, длительность $\tau_и = 5$ мкс, циклическая частота несущей немодулированного сигнала $f_0 = 5$ ГГц, девиация частоты несущей $f_д = 50$ МГц.

Р е ш е н и е. База анализируемого ЛЧМ-сигнала $B = f_д\tau_и = 50\cdot10^6\cdot5\cdot10^{-6} = 250$. Из соотношения (2.158) вычисляем скорость нарастания частоты ЛЧМ-импульса $\alpha = 2\pi f_д/\tau_и = 6{,}28\cdot50\cdot10^6/(5\cdot10^{-6}) = 6{,}28\cdot10^{13}$ с$^{-2}$. Согласно (2.175) энергетический спектр $\mathbf{W}_и(\omega) = 3{,}14\cdot100/(12{,}56\cdot10^{13}) = 2{,}5\cdot10^{-12}$. Поскольку база сигнала достаточно велика, то амплитудный и энергетический спектры расположены в полосе частот от $f_0 - f_д/2 = 4{,}975$ ГГц до $f_д + f_д/2 = 5{,}025$ ГГц.

А в т о к о р р е л я ц и о н н а я ф у н к ц и я Л Ч М - с и г н а л а. Для упрощения анализа положим, что база ЛЧМ-сигнала достаточно велика, и поэтому его энергетический спектр равномерен и расположен лишь в полосе $-\omega_д/2 \ldots \omega_д/2$ вокруг несущей частоты $\omega_0$. Тогда, согласно (2.107), АКФ ЛЧМ-сигнала:

$$B_{ЛЧМ}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathbf{W}_и(\omega)e^{j\omega\tau}d\omega = \frac{U_и^2}{2\alpha}\int_{\omega_0-\omega_д/2}^{\omega_0+\omega_д/2}\cos\omega\tau d\omega = \frac{U_и^2}{2\alpha}\int_{\omega_0-\alpha\tau_и/2}^{\omega_0+\omega_д/2}\cos\omega\tau d\omega = \frac{U_и^2\tau_и}{2}\frac{\sin(\alpha\tau_и\tau/2)}{\alpha\tau_и\tau/2}\cos\omega_0\tau. \qquad (2.176)$$

Типичный график нормированной АКФ прямоугольного ЛЧМ-импульса $R(\tau) = B_{ЛЧМ}(\tau)/B_{ЛЧМ}(0)$ показан на рис. 2.47. Формула (2.176) устанавливает следующее важное свойство ЛЧМ-сигнала — ширина главного лепестка огибающей АКФ обратно пропорциональна девиации частоты несущей $f_д$. Это объясняется тем, что огибающая АКФ ЛЧМ-сигнала первый раз обращается в нуль при сдвиге сигнала относительно его копии на интервал времени $\tau = 2\pi(\alpha\tau_и) = 1/f_д$.

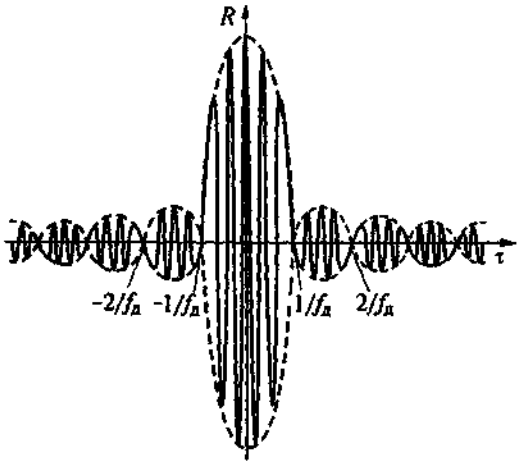



**Рис. 2.47.** График нормированной АКФ ЛЧМ-импульса

Применяемые в РЛС и других специальных радиосистемах передачи информации ЛЧМ-сигналы характеризуются значительной девиацией частоты, поэтому главный лепесток АКФ получается весьма узким, что без сомнения полезно. Однако с точки зрения корреляционных свойств ЛЧМ-сигналам присущ известный недостаток — высота двух первых симметричных боковых лепестков АКФ достаточно велика, составляя 0,212 от высоты центрального лепестка. При значительном уровне шумов радиоприемника это может привести к ошибочному определению временно́го положения импульса при его обработке. Если же энергетические потери в устройстве сжатия ЛЧМ-импульса малы, то амплитуда выходного сигнала может значительно превысить уровень шумов. Это увеличивает надежность обнаружения приемником РЛС слабых отраженных от цели сигналов.

Важной особенностью ЛЧМ-сигнала с большой базой является возможность *управлять* ее шириной $B \approx f_д \tau_и$ при постоянной длительности импульса, изменяя *девиацию* частоты. В силу этого свойства ЛЧМ-сигнал играет особую роль в радиоизмерительной технике: он используется в таких важных приборах, как генераторы измерителей АЧХ и анализаторы спектра. В радиотехнике понятие базы применяют по отношению к разнообразным сигналам. При $B >> 1$ сигнал называют *сложным*, при $B \approx 1$ — *простым*. На практике почти всегда выполняется условие, что $B >> 1$. Спектр ЛЧМ-сигналов с большой базой имеет ряд специфических особенностей и такие сигналы наиболее востребованы.

Отметим, что ЛЧМ-сигналы находят широкое применение в радиолокации. ЛЧМ-импульсы сжимаются оптимальным фильтром во времени, что дает возможность увеличивать разрешающую способность РЛС по дальности без укорочения длительности импульсов (кроме того, можно разделить сигналы от близко расположенных друг от друга целей). Это, в свою очередь, позволяет увеличивать энергию зондирующих импульсов, от которой зависит дальность действия. Увеличивать энергию зондирующих сигналов за счет повышения импульсной мощности передатчика не всегда практически возможно, поскольку у РЛС импульсная мощность может исчисляться мегаваттами. Генерировать и подводить к антенне сигналы такой мощности сложно технически. И еще одно преимущество ЛЧМ-сигналов перед простыми радиоимпульсами — при ограниченной пиковой мощности РЛС появляется возможность увеличивать дальность ее действия не за счет повышения пиковой мощности, а путем увеличения длительности импульса без ухудшения разрешающей способности по дальности.

## 2.5. Сигналы с импульсной и цифровой модуляцией

При *импульсной модуляции* в качестве несущего колебания (поднесущего) используют различные периодические импульсные последовательности, один из параметров которого изменяют по закону передаваемого сообщения (рис. 2.48).

### *Импульсная модуляция*

Теоретической основой импульсной модуляции служит *теорема Котельникова* (*теорема отсчетов*). Упрощенно теорему можно интерпретировать так: *произвольный сигнал u(t), спектр которого ограничен некоторой верхней частотой* $F_в$, может быть передан своими отсчетами (*или полностью восстановлен по последовательности своих отсчетных значений*), следующими с интервалом

$$T = \Delta t = \frac{1}{2F_в}. \qquad (2.177)$$

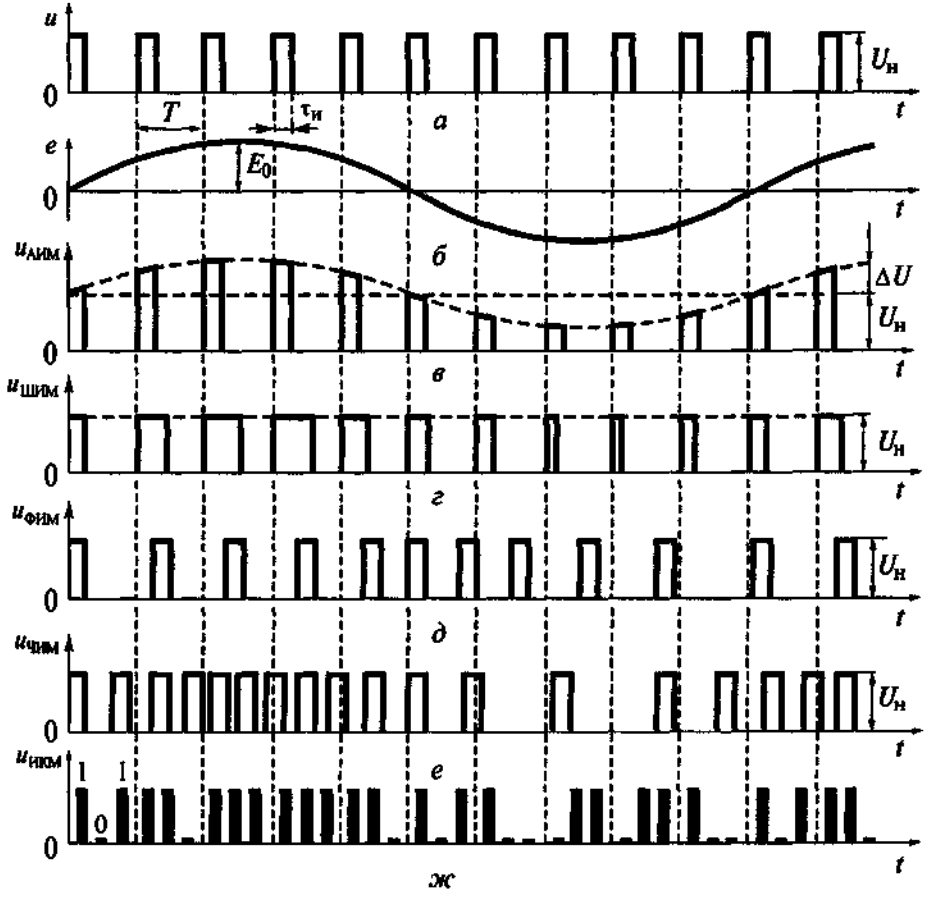



**Рис. 2.48.** Импульсная модуляция:
*а* — периодическая последовательность исходных импульсов;
*б* — модулирующий сигнал; *в* — АИМ; *г* — ШИМ; *д* — ФИМ; *е* — ЧИМ; *ж* — ИКМ

Заметим, что в радиотехнике и связи при представлении импульсных, дискретных и цифровых сигналов часто период обозначают как $\Delta t = T$.

Как правило, достаточно большие временны́е интервалы между полезными импульсами используют для передачи полезных импульсов от других источников сообщений, т. е. для осуществления многоканальной передачи сигналов с временны́м разделением (уплотнением) каналов.

Положим, что в качестве поднесущего колебания в системе связи с импульсной модуляцией сигналов используется периодическая последовательность прямоугольных импульсов с амплитудой $U_н$, длительностью $\tau_и$ и периодом повторения $T$ (рис. 2.48, *а*); соответственно частота следования импульсов несущей $F_н = 1/T$. Для наглядности математических выкладок выберем в качестве модулирующего сигнала (отражающего передаваемое сообщение) гармоническое колебание $e(t) = E_0\cos\Omega t$ (причем $\Omega = 1/T_0$), у которого для упрощения принята начальная фаза $\theta_0 = 90°$ (рис. 2.48, *б*).

Импульсную модуляцию сигналов в зависимости от выбора изменяемого параметра модулируемой последовательности импульсов делят на такие виды:

- *амплитудно-импульсную* (АИМ; *pulse amplitude modulation — PAM*), когда по закону передаваемого сообщения изменяется амплитуда прямоугольных импульсов исходной последовательности (рис. 2.48, *в*);
- *широтно-импульсную* (ШИМ; (*pulse-duration modulation — PDM* или *pulse-width modulation* — PWM), при изменении по закону передаваемого сообщения длительности (ширины) прямоугольных импульсов исходной последовательности (рис. 2.48, *г*);
- *фазоимпульсную* (ФИМ; *pulse-position modulation — PPM*), или *времяимпульсную* (ВИМ), если по закону передаваемого сообщения изменяется временно́е положение импульсов в последовательности (рис.2.48, *д*); ФИМ отличается от ВИМ методом синхронизации — при ФИМ сдвиг фазы импульса изменяется не относительно синхронизирующего импульса, а относительно некоторой условной фазы;
- *частотно-импульсную* (ЧИМ; *pulse-frequency modulation* — PFM), при изменении по закону передаваемого сообщения частоты следования прямоугольных импульсов (рис. 2.48, *е*);
- *импульсно-кодовую* (ИКМ; *Pulse Code Modulation* — PCM) — вид дискретной модуляции (*цифровой манипуляции — keying*), когда аналоговый первичный сигнал часто кодируется сериями импульсов и превращается в цифровой код — последовательность стандартных импульсов (единиц) и пауз (нулей), имеющих одинаковую длительность, применяется наиболее широко в современных системах связи.

Последний вид импульсной модуляции, упрощенно рассмотренный в разд. 2.1, представлен на рис. 2.48, *ж*. Заметим, что в одном периоде $\Delta t$ зачастую интервалы между соседними кодовыми посылками отсутствуют, как показано на рис. 2.3, *г*. Необходимо отметить два основных метода преобразования аналоговых сигналов в цифровую форму — импульсно-кодовую и дельта-модуляцию

(ДМ), которые традиционно относят к видам модуляции, хотя таковыми они, по сути, не являются, поскольку не связаны, как это характерно для всех остальных видов модуляции, с изменениями параметров несущей частоты. Оба вида модуляции были изобретены в 40-х гг. и приобрели исключительно большое значение в конце XX в., когда началась глобальная цифровизация систем связи и вещания. При ИКМ преобразования аналогового сигнала в цифровую форму осуществляют в два этапа. На первом этапе этот сигнал модулирует по амплитуде последовательность импульсов, следующих с частотой, определяемой теоремой Котельникова и равной $2F_в$, где $F_в$ — верхняя частота спектра сигнала. На втором этапе диапазон возможных уровней сигнала разбивается на $2^n$ интервалов и определяется, в каком из интервалов находится уровень каждого из модулированных импульсов. В результате каждый импульс преобразуется в $n$-значную бинарную кодовую комбинацию, соответствующую данному интервалу. Для увеличения динамического диапазона сигналов, преобразуемых в цифровую форму, применяются *компандеры*, а при обратном преобразовании — *экспандеры*. Метод ИКМ нашел широкое применение в современных системах связи при передаче звуковых сигналов и сигналов телевидения.

Частота следования импульсов несущей $F_н$ в импульсных системах связи определяется максимальной частотой первичного сигнала (здесь модулирующего $F_{max} = \Omega$): $F_н \geq 2\Omega$. Действительно, в импульсных системах радиосвязи передаются лишь дискретные (во времени) отсчеты первичного сигнала $e(t)$. Согласно теореме Котельникова частота дискретизации $F_д \geq 2\Omega$. Поэтому частоту дискретизации $F_д$ и можно выбрать в качестве частоты следования импульсов $F_н$.

Отметим, что иногда в специальных системах передачи информации используется частный вариант ИКМ — *счетно-импульсная модуляция* (СИМ; *pulse-accounting modulation — PAM*), при которой информационным параметром является число импульсов в кодовой группе.

**Амплитудно-импульсная модуляция.** В качестве примера, позволяющего оценить характеристики импульсно-модулированных колебаний, рассмотрим АИМ-сигнал и определим его спектр при модуляции периодической последовательности прямоугольных импульсов гармоническим колебанием $e(t) = E_0\cos\Omega t$.

С аналитической точки зрения процедуру получения АИМ-сигнала $u_{АИМ}(t)$ удобно рассматривать как непосредственное умножение непрерывного передаваемого сигнала $u(t)$ на вспомогательную последовательность $y(t)$ прямоугольных видеоимпульсов единичной амплитуды (см. гл.8)

$$u_{АИМ}(t) = u(t)y(t). \qquad (2.178)$$

Представим последовательность прямоугольных импульсов $u(t)$, имеющих амплитуду $U_н$, длительность $\tau_н$ и период повторения $T$, тригонометрическим рядом Фурье (2.36). Введем в соотношение (2.117) в качестве несущего колебания $u_н(t) = U_н\cos\omega_0 t$ обобщенную функцию $u(t)$, описывающую последовательность прямоугольных импульсов. Тогда АИМ-сигнал можно записать в виде

$$u_{АИМ}(t) = (1 + M\cos\Omega t)\,u(t). \qquad (2.179)$$

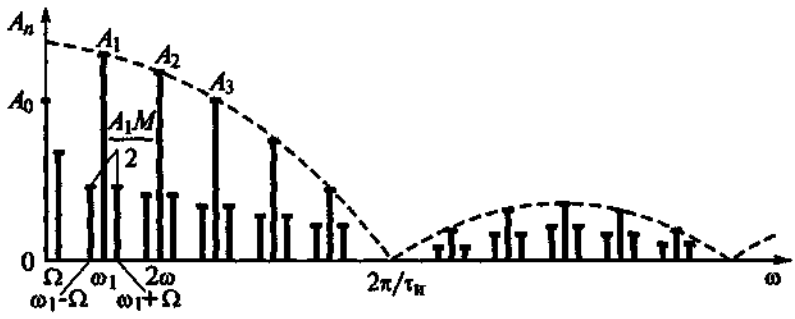



**Рис. 2.49.** Спектр сигнала при амплитудно-импульсной модуляции

В данном соотношении параметр $M = \Delta U/U_н$ — коэффициент (глубина) модуляции прямоугольных импульсов. Подставляя значение функции $u(t)$ из (2.36) в (2.179), после несложных преобразований для АИМ-сигнала получим

$$u_{АИМ}(t) = (1 + M\cos\Omega t)A_0 + (1 + M\cos\Omega t)\sum_{n=1}^{\infty} A_n \cos(n\omega_1 t - \varphi_n) =$$

$$= A_0 + A_0 M\cos\Omega t + \sum_{n=1}^{\infty} A_n \cos(n\omega_1 t - \varphi_n) + \tag{2.180}$$

$$+ \sum_{n=1}^{\infty} \frac{A_n M}{2}\cos[(n\omega_1 + \Omega)t - \varphi_n] + \sum_{n=1}^{\infty} \frac{A_n M}{2}\cos[(n\omega_1 - \Omega)t - \varphi_n].$$

Из соотношения (2.180) следует, что при однотональной амплитудно-импульсной модуляции последовательности прямоугольных видеоимпульсов спектр АИМ-сигнала достаточно сложен и содержит постоянную составляющую $A_0$, гармонику $A_0M$ частоты Ω модулирующего колебания и высшие гармонические составляющие $A_n$ частоты следования импульсов несущей $n\omega_1$, около каждой из которых симметрично попарно расположены боковые составляющие с частотами $n\omega_1 + \Omega$ и $n\omega_1 - \Omega$ (рис. 2.49).

Основные разновидности АИМ-сигналов. Сигналы с амплитудно-импульсной модуляцией подразделяются на два основных вида: сигнал первого рода — АИМ-I и сигнал второго рода — АИМ-II (рис. 2.50).

Пусть имеется исходная последовательность прямоугольных импульсов с амплитудой $U_н$, длительностью $\tau_и$ и периодом следования $T$ (рис. 2.50, *а*), являющаяся переносчиком сигнала (импульсной несущей). Мгновенное значение амплитуды импульсов сигнала АИМ-I зависит от мгновенного значения модулирующего колебания $e(t)$ (рис. 2.50, *б*), а амплитуда импульсов сигнала АИМ-II определяется только значением модулирующего колебания в тактовых точках (рис. 2.50, *в*). Тактовые моменты могут совпадать с началом импульса, любой точкой его середины или концом. Поэтому при АИМ-II несущая последовательность импульсов характеризуется еще одним параметром — положением импульсов относительно тактовых точек.

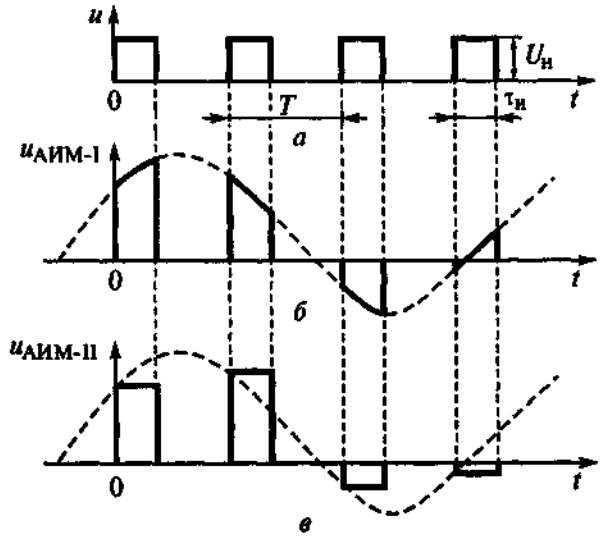



**Рис. 2.50.** Формирование АИМ-сигналов:
*а* — импульсная несущая; *б* — АИМ-I; *в* — АИМ-II

Различие между модулированными сигналами видов АИМ-I и АИМ-II оказывается существенным, если длительность несущих импульсов $\tau_и$ сравнима с периодом их следования $T$. Для оценки эффективности методов АИМ для передачи сообщений по каналу радиосвязи необходимо знать полосу частот используемых сигналов.

Сигналы типа АИМ-I при простейшем, однотональном модулирующем сигнале, определяемом формулой (2.180), на практике в системах передачи информации используются лишь для настройки.

Оценим и проанализируем спектр импульсного сигнала с АИМ-I (на положительных частотах) при некотором реальном узкополосном модулирующем колебании (рис. 2.51). Пусть функция $e(t)$ — модулирующий сигнал сложной формы, имеющий спектральную плотность $\mathbf{E}(\omega)$, расположенную в некоторой полосе частот $0 \le \omega \le \Omega_m$ (рис. 2.51, *а*). В этом случае аналитическое выражение для спектра получаемого АИМ-сигнала вида (2.180) можно записать в обобщенной форме

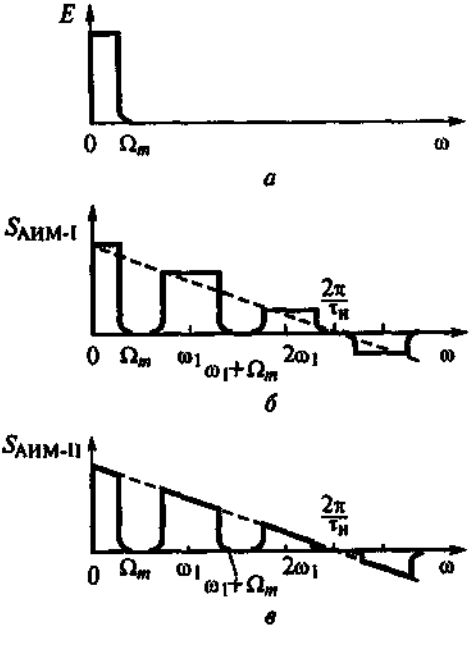



**Рис. 2.51.** Спектры сигналов с АИМ:
*а* — передаваемого;
*б* — АИМ-I; *в* — АИМ-II

(оно приводится в специальной литературе, однако его нетрудно вывести студентам самостоятельно)

$$\mathbf{S}(\omega)_{\text{АИМ-I}} = \frac{\tau_и}{T}\left[\mathbf{E}(\omega) + \sum_{n=1}^{\infty}\mathbf{U}_0(n\omega_1)\mathbf{E}(\omega + n\omega_1)\right], \qquad (2.181)$$

где $\mathbf{U}_0(n\omega_1)$ — дискретные значения спектра одиночного элемента (импульса) несущей $u(t)$.

Спектр одиночного элемента импульсной несущей $\mathbf{U}_0(\omega)$ можно вычислить достаточно просто практически при любой его форме. Так, для прямоугольного импульса с амплитудой $U_н$ и длительностью $\tau_и$ спектр определяют по формуле

$$\mathbf{U}_0(\omega) = U_н\tau_и\frac{\sin(\omega\tau_и/2)}{\omega\tau_и/2}. \qquad (2.182)$$

Из (2.181) и (2.182) следует, что спектр сигнала АИМ-I с несущей в виде последовательности прямоугольных импульсов содержит с точностью до постоянного множителя $\tau_и/T$ спектр модулирующего сигнала $e(t)$ и бесконечное множество боковых полос сигнала около каждой гармоники $n\omega_1$ несущей (рис. 2.51, *б*).

При произвольной форме импульсов несущей $u(t)$ сигнал АИМ-II

$$u(t)_{\text{АИМ-II}} = \sum_{n=-\infty}^{\infty}e(nT)u_0(t-nT), \qquad (2.183)$$

где $e(nT)$ — амплитуда передаваемого сигнала (отсчета) в тактовой точке $t = nT$; $u_0(t)$ — одиночный элемент (импульс) импульсной несущей $u(t)$.

Импульсную несущую можно представить суммой

$$u(t) = \sum_{n=-\infty}^{\infty}u_0(t-nT). \qquad (2.184)$$

При определении спектра сигнала АИМ-II воспользуемся соотношением (2.183), которое в математике и основах теории цепей называют *импульсной сверткой* сигналов (функций) во времени. В то же время из математики также известно, что свертке во времени функции $u_0(t)$ и последовательности отсчетов первичного сигнала $\{e(nT)\} = e_д(t)$ в частотной области соответствует произведение их спектров, т. е.

$$\mathbf{S}(\omega)_{\text{АИМ-II}} = \mathbf{U}_0(\omega)\mathbf{E}_д(\omega), \qquad (2.185)$$

где $\mathbf{E}_д(\omega)$ — спектр последовательности отсчетов первичного сигнала $e(t)$.

Спектр последовательности отсчетов первичного сигнала легко получить из спектра сигнала АИМ-I путем предельного перехода (при длительности отсчета $\tau_и > 0$ и его амплитуды $U_н \to \infty$), когда площадь каждого отсчетного импульса составляет $\tau_и U_н = 1$. Для прямоугольных импульсов спектр последовательности отсчетов первичного сигнала

$$\mathbf{E}_{\text{д}}(\omega) = \lim_{\tau_{\text{и}} \to 0} \frac{\tau_{\text{и}} U_{\text{и}}}{T} \sum_{n=-\infty}^{\infty} \frac{\sin(n\omega_1 \tau_{\text{и}}/2)}{n\omega_1 \tau_{\text{и}}/2} [\mathbf{E}(\omega - n\omega_1) + \mathbf{E}(\omega + n\omega_1)] = \\ = \frac{1}{T} \sum_{n=-\infty}^{\infty} [\mathbf{E}(\omega - n\omega_1) + \mathbf{E}(\omega + n\omega_1)]. \tag{2.186}$$

Объединяя $\mathbf{U}_0(\omega)$ и $\mathbf{E}_{\text{д}}(\omega)$ в произведение, получим спектр сигнала АИМ-II (рис. 2.51, *в*) с несущей в виде последовательности прямоугольных импульсов

$$\mathbf{S}(\omega)_{\text{АИМ-II}} \approx \frac{\tau_{\text{и}}}{T} \frac{\sin(\omega\tau_{\text{и}}/2)}{\omega\tau_{\text{и}}/2} \left[ \mathbf{E}(\omega) + \sum_{n=-\infty}^{\infty} \mathbf{E}(\omega \pm n\omega_1) \right]. \tag{2.187}$$

Анализ формулы (2.187) показывает, что спектр сигнала АИМ-II так же, как и АИМ-I (сравни рис. 2.51, *б* и 2.51, *в*), состоит из спектра $\mathbf{E}(\omega)$ модулирующего первичного сигнала $e(t)$ и бесчисленного множества его боковых полос около каждой гармоники импульсной несущей $n\omega_1$. Однако в отличие от АИМ-I, здесь перед суммой стоит частотно-зависимый множитель $\{[\sin(\omega\tau_{\text{и}}/2)]/(\omega\tau_{\text{и}}/2)\}$, описывающий спектр отдельного элемента импульсной несущей. Это говорит о наличии амплитудно-частотных искажений всех спектральных составляющих, включая $\mathbf{E}(\omega)$. По существу при АИМ-II имеет место аппроксимация непрерывной линии передаваемого аналогового сигнала ступенчатой (ступенчатая аппроксимация), поэтому и появляются амплитудно-частотные искажения. Из спектральной диаграммы сигнала АИМ-II видно, что степень амплитудно-частотных искажений определяется значением $\tau_{\text{и}}$. При $\tau_{\text{и}} \to 0$ амплитудно-частотные искажения уменьшаются и сигнал АИМ-II практически совпадает с сигналом АИМ-I. Но в то же время доля мощности полезной составляющей в спектре сигнала как АИМ-I, так и АИМ-II при $\tau_{\text{и}} \to 0$ снижается, что, естественно, сказывается на помехозащищенности принимаемого сигнала и дальности действия систем передачи информации.

В случаях, когда требуется передать сообщение на большое расстояние, сигналами с импульсной модуляцией модулируют, в свою очередь, высокочастотное колебание несущей частоты. Полученные подобным образом радиоимпульсы затем излучаются антенной в свободное пространство. Заметим, что помехоустойчивость АИМ-сигналов является очень низкой, поскольку любая помеха, изменяющая амплитуду импульса, искажает форму огибающей импульсной последовательности, а следовательно, и сигнал в приемнике.

**Широтно-импульсная модуляция.** С точки зрения помехоустойчивости систем передачи информации из импульсно-модулированных сигналов (кроме сигналов с цифровой модуляцией) более эффективны сигналы с ШИМ. Так же, как и АИМ, ШИМ бывает первого и второго рода. При ШИМ-I длительность импульсов определяется значениями модулирующего сигнала в моменты возникновения переднего фронта и среза (заднего фронта) импульса, а при ШИМ-II — в тактовых (отсчетных) точках. Если длительность немодулированного импульса $\tau_{\text{и}} << \Delta t$, то разница между ШИМ-I и ШИМ-II незначительна. Поэтому ШИМ

классифицируют как одно- и двухстороннюю модуляцию. При односторонней ШИМ изменяется временно́е положение переднего или заднего фронта импульса, а при двухсторонней — сдвигаются и фронт и срез импульса. В основном применяют одностороннюю ШИМ-I (ОШИМ-I), при которой длительность импульса при модуляции гармоническим сигналом $e(t) = E_0\cos\Omega t$

$$\tau_ш = \tau_и + \Delta\tau_{max}\cos\Omega t,$$

где $\tau_ш$ — среднее значение длительности импульса при ШИМ; $\Delta\tau_{max}$ — максимальное отклонение фронта импульса.

Анализ свойств сигналов с ОШИМ-I показывает, что такие колебания по сравнению с АИМ-сигналами обладают большей помехоустойчивостью. Это обстоятельство связано с тем, что при ОШИМ-I (или любом другом ШИМ) для устранения влияния помех можно применять двухстороннее ограничение амплитуд импульсов. ШИМ в основном применяется в телеметрии.

**Фазоимпульсная модуляция.** Более помехоустойчивым видом импульсной модуляции является фазоимпульсная или времяимпульсная модуляция. Этот вид модуляции представляет собой разновидность временно́й импульсной модуляции, при которой импульсы периодической последовательности, сохраняя свою форму и фиксированную длительность, смещаются во времени относительно тактовых точек. Если величина сдвига импульсов определяется только амплитудой модулирующего сигнала и не зависит от частоты модуляции, то имеет место ФИМ. Таким образом, информационным параметром ВИМ-сигнала является временно́е положение импульсов в пределах тактовых интервалов.

При фазоимпульсной модуляции длительность импульса постоянна и в отсутствие модуляции может быть у́же, чем при ШИМ. Поэтому средняя мощность сигнала при ФИМ меньше, чем при ШИМ. Следовательно, и средняя мощность передатчиков при ФИМ меньше, чем при ШИМ. При одинаковой средней мощности сигнала ФИМ позволяет увеличить амплитуды импульсов и тем самым повысить на входе приемника отношение сигнал/шум. Поскольку при ШИМ передаются и короткие импульсы, то полоса пропускания тракта передачи оказывается более широкой, чем при передаче импульсов с ФИМ.

### *Квадратурная модуляция*

Квадратурную (*quadrature*) модуляцию осуществляют путем передачи по каналу связи в одной и той же полосе частот двух модулированных сигналов, несущие колебания которых ортогональны и квадратурны (их частоты равны, а фазы сдвинуты на 90°, что и поясняет смысл слова «квадратурный»). Временны́е диаграммы, поясняющие квадратурную модуляцию, приведены рис. 2.52.

Ранее были проанализированы случаи, когда амплитуда и начальная фаза несущего гармонического колебания подвергались модуляции по отдельности. Однако если изменять эти два параметра одновременно, то можно будет передавать сразу два сигнала модулированных по амплитуде $U_н(t)$ и фазе $\psi(t)$

$$u_{мод}(t) = U_н(t)\cos[\omega_0 t + \psi(t)]. \qquad (2.188)$$

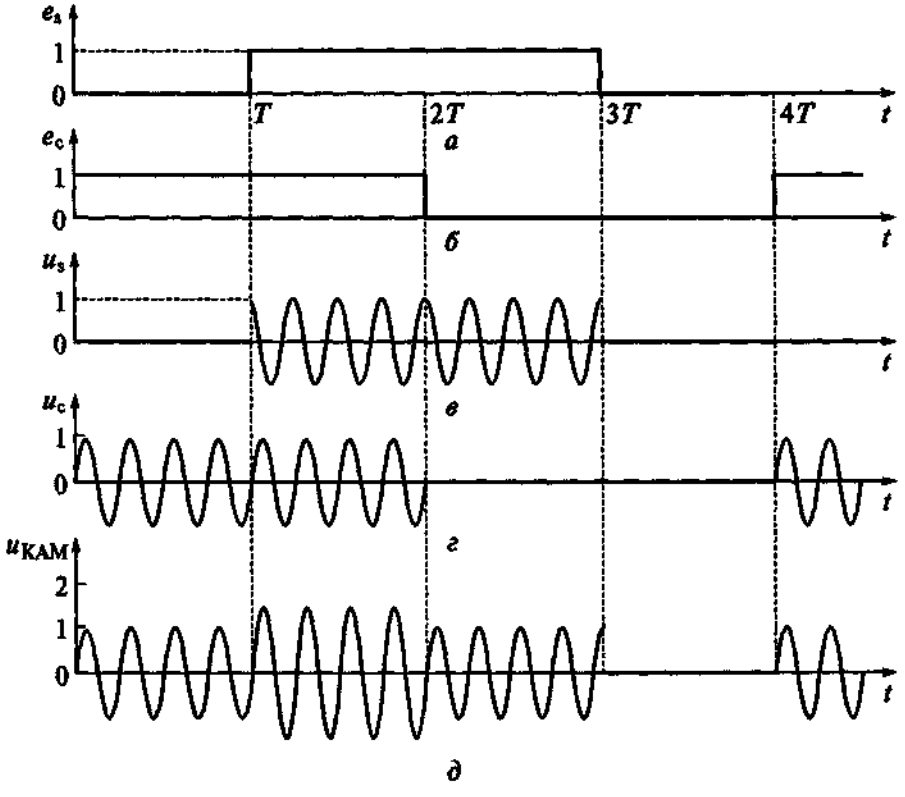



**Рис. 2.52.** Временны́е диаграммы сигналов с квадратурной модуляцией:
*а* — исходный синусный; *б* — исходный косинусный; *в* — квадратурный синусный;
*г* — квадратурный косинусный; *д* — квадратурный амплитудно-модулированный

Такую модуляцию следовало бы назвать просто *амплитудно-фазовой* и очевидно аналоговой. Однако два модулирующих сигнала модулируют совершенно разные параметры несущего колебания — амплитуду и фазу (используется двумерный характер гармонического несущего колебания). Можно сделать ситуацию более «симметричной», слегка преобразовав форму представления рассматриваемого сигнала. Для этого воспользуемся тригонометрической формулой представления косинуса суммы двух углов и преобразуем выражение (2.188) к виду

$$u_{мод}(t) = U_н(t)\cos\psi(t)\cos\omega_0 t - U_н(t)\sin\psi(t)\sin\omega_0 t. \qquad (2.189)$$

Теперь исходный радиосигнал оказался представленным в виде суммы двух АМ-колебаний. Их несущие единичной амплитуды — $\cos\omega_0 t$ и $\sin\omega_0 t$ — сдвинуты по фазе на 90° друг относительно друга, а амплитудными функциями являются соответственно $U_н(t)\cos\psi(t)$ и $U_н(t)\sin\psi(t)$, преобразованные в два битовых потока с уровнями 1 и 0 (часто используются уровни +1 и –1). Косинусная составляющая такого сигнала называется *синфазной*, а синусная — *квадратурной*.

Преобразуем данные функции в импульсную форму с единичной амплитудой и длительностью $2T$, где $T$ — интервал (скорость передачи), в несколько раз

больший, чем $T_0 = 2\pi/\omega_0$ (рис. 2.52). Обозначим АМ-колебания в (2.189) в виде $e_c(t) = U_н(t)\cos\psi(t)$ и $e_s(t) = -U_н(t)\sin\psi(t)$ (рис. 2.52, *а*, *б*) и используем их как новую пару модулирующих сигналов (вместо амплитуды и начальной фазы), тогда

$$u_{мод}(t) = e_c(t)\cos\omega_0 t + e_s(t)\sin\omega_0 t = u_c(t) + u_s(t). \qquad (2.190)$$

Из формулы (2.190) следует, что в общем случае полученный сигнал можно рассматривать как сумму *квадратурных* колебаний:
*косинусного*

$$u_c(t) = e_c(t)\cos\omega_0 t = U_н\cos\psi(t)\cos\omega_0 t;$$

*синусного*

$$u_s(t) = e_s(t)\sin\omega_0 t = U_н\sin\psi(t)\sin\omega_0 t,$$

каждое из которых модулировано только по амплитуде (рис. 2.52, *в*, *г*).

Нетрудно также заметить, что закон амплитудной модуляции для косинусного колебания определяется «медленной» функцией $\cos\psi(t)$, а синусного — «медленной» функцией $\sin\psi(t)$. Но уже было установлено, что для определения спектра АМ-колебания достаточно сдвинуть на частоту $\omega_0$ спектр огибающей амплитуд передаваемого сообщения. Следовательно, для нахождения спектра колебания $u_{мод}(t)$, определяемого выражением (2.189), необходимо сначала найти спектры функций $\cos\psi(t)$ и $\sin\psi(t)$, т. е. спектры огибающих квадратурных колебаний. Перенос этих спектров на частоту $\omega_0$ можно затем осуществить как и при обычной амплитудной модуляции.

Модулирующие сигналы при этом оказываются совершенно равноправными. Как видно, модуляция является сложной: в моменты $T$ и $2T$ (и далее периодически) могут возникать (и возникают) как скачки амплитуд, так и скачки фаз. Это позволяет увеличить количество кодируемых в единицу времени бит и при этом существенно повысить помехоустойчивость их передачи по каналу связи. В настоящее время число кодируемых информационных бит на одном бодовом интервале может достигать 8-9.

Упрощенно квадратурную модуляцию можно трактовать следующим образом. Передаваемый сигнал разделяют на два независимых битовых потока — $e_c(t)$ и $e_s(t)$. Сигнал $e_c(t)$ с уровнями 1 и 0 модулирует косинусоидальное несущее колебание, а сигнал $e_s(t)$ аналогичным образом — синусоидальное несущее колебание. Затем два модулированных колебания (по существу это и АМ-сигнал и ФМ-сигнал) суммируются, образуя единое квадратурно-модулированное колебание. Такое представление исходного сигнала (рис. 2.52, *д*) называют *квадратурным*, а данный способ модуляции — *квадратурной модуляцией*, или точнее, *квадратурной амплитудной модуляцией* (КАМ; *quadrature amplitude modulation, QAM*). Результирующий сигнал называют *двухполосным сигналом с подавлением несущей* (*double-sideband suppressed-carrier*; *DSB-SC*), поскольку полоса радиочастот вдвое больше полосы немодулированного сигнала и не содержит выделенной несущей. Итак, квадратурная амплитудная модуляция представляет сочетание АМ и ФМ, в результате чего образуется квадратурный ФМ-сигнал.

Получение различных видов модуляции с помощью квадратурного модулятора обеспечивают подачей на его входы биполярных АИМ-сигналов $e_c(t)$ и $e_s(t)$, квантованных на разное число уровней и симметричных относительно нуля. В «вырожденном» случае, т. е. когда на один из входов подан нуль напряжения, а на другой двоичная последовательность с относительными уровнями ±1, работает только один канал, и модулятор превращается из квадратурного в обычный балансный. На выходе формируется одномерный ФМ-сигнал с изменением фазы на 180°, переносящий 1 бит/символ. При подаче двоичных АИМ-сигналов в оба канала модулятора по каждому из каналов передается 1 бит/символ, а общая скорость передачи составляет 2 бит/символ. В результате образуется сигнал ФМ-4, но формально относящийся к широкому классу *квадратурной амплитудно-фазовой модуляции* (КАФМ). Поэтому квадратурную модуляцию называют еще и *многопозиционной амплитудно-фазовой манипуляцией* (АФМ; *amplitude phase keying — APK*). И наконец, данный вид модуляции можно считать двухмерной амплитудной манипуляцией, откуда еще одно название — *квадратурная амплитудная манипуляция* (КАМн; *quadrature amplitude shift keying — QASK*).

В теории связи доказано, что КАМ-сигнал (2.190) можно представить как

$$u_{\text{мод}}(t) = u_{\text{КАМ}}(t) = \sqrt{e_c^2(t) + e_s^2(t)}\cos\left\{\omega_0 t - \operatorname{arctg}\frac{e_c(t)}{e_s(t)} + \frac{\pi}{2}\left[\operatorname{sign}\left(e_c(t) - 1\right)\right]\right\}. \quad (2.191)$$

Спектр сигнала с квадратурной модуляцией. Для квадратурной модуляции получить спектральную функцию не составляет труда. Поскольку КАМ-сигнал $u_{\text{мод}}(t) = u_{\text{КАМ}}(t)$ представляет собой сумму двух АМ-сигналов (2.190), то его можно записать так

$$\mathbf{S}_{\text{КАМ}}(\omega) = \frac{1}{2}\left[\mathbf{S}_c(\omega + \omega_0) + \mathbf{S}_c(\omega - \omega_0)\right] + j\frac{1}{2}\left[\mathbf{S}_s(\omega + \omega_0) - \mathbf{S}_s(\omega - \omega_0)\right], \quad (2.192)$$

где $\mathbf{S}_c(\omega)$ и $\mathbf{S}_s(\omega)$ — соответственно спектральные плотности косинусной $u_c(t)$ и синусной $u_s(t)$ составляющих квадратурно-модулированного колебания.

Итак, аналогично тому, что происходит при амплитудной модуляции, спектры модулирующих сигналов при КАМ «раздваиваются» и смещаются в окрестности несущей частоты $\pm\,\omega_0$. Если спектры модулирующих сигналов $u_c(t)$ и $u_s(t)$ занимают одну и ту же полосу частот (так обычно и бывает), то они будут перекрываться и после сдвига в область несущей частоты. Однако при этом спектр, соответствующий синусной несущей, дополнительно умножается на мнимый коэффициент $\pm j$. Именно эта операция дает возможность разделить квадратурные составляющие модулированного колебания при приеме КАМ-сигнала.

Ширина спектра квадратурно-модулированного колебания, равная $1/T$, в два раза меньше ширины спектра обычного аналогового амплитудно-модулированного колебания, модулированного единым сигналом сообщения со скоростью передачи $1/T$.

Но возможно ли все-таки детектирование и разделение двух модулирующих сигналов $e_c(t)$ и $e_s(t)$? Это принципиально возможно, а практически и технически

обеспечивается с помощью синхронных детекторов, так как квадратурные несущие ортогональны, т. е. их среднее (по времени) произведение равно нулю. В синхронном детекторе квадратурно-модулированное колебание умножается на косинусоидальный сигнал и результат перемножения усредняется во времени. В результате подавляется квадратурная компонента $e_c(t)$ и выделяется огибающая синфазной составляющей $e_c(t)$. Аналогичным образом выделяется и огибающая квадратурной компоненты модулированного колебания $e_s(t)$.

Квадратурный модулятор является универсальным устройством, которое может быть использовано для получения сигнала линейно-модулированной несущей с двумя боковыми полосами, включая фазовую и амплитудно-фазовую модуляции несущей.

Квадратурную модуляцию используют в системах аналогового цветного телевидения *NTSC* и *PAL* при модуляции цветовой поднесущей двумя цветоразностными сигналами. Кроме того, в большинстве систем цифрового телевидения и демодуляторы сигналов также строят по квадратурным схемам. В известных современных системах цифрового цветного телевидения и телевидения высокой четкости применяют многоуровневую амплитудную модуляцию с частично подавленной нижней боковой полосой: (8-, 16-*VSB*), четырехпозиционную квадратурную фазовую модуляцию (см. далее) и квадратурную амплитудно-фазовую модуляцию (16-, 64-, 256-*QAM*).

### *Цифровая модуляция*

Цифровая модуляция — процесс преобразования цифровых символов в сигналы, совместимые с характеристиками канала связи. Главная идея цифровой модуляции состоит в том, что каждому возможному значению передаваемого символа ставятся в соответствие некоторые параметры аналогового несущего колебания. Способ цифровой или импульсной модуляции, когда параметры несущего колебания меняются скачкообразно, называют *манипуляцией*. При низкочастотной модуляции (*baseband modulation*) эти сигналы обычно имеют вид импульсов заданной формы.

При цифровой модуляции используют чаще всего дискретные последовательности двоичных символов — двоичных кодов. Простейшим линейным кодом, широко применяемым на практике, является код *NRZ* (*Non Return to Zero*) — невозвращающийся в нуль код.

Существуют две разновидности этого кода — униполярный и биполярный NRZ-коды. В униполярном NRZ-коде логической единице соответствует прямоугольный импульс положительной полярности, а логическому нулю — нулевое напряжение (пауза). В биполярном NRZ-коде логической единице соответствует прямоугольный импульс положительной полярности, а логическому нулю — прямоугольный импульс отрицательной полярности. Положительное или отрицательное напряжение на выходе кодера сохраняется неизменным в течение длительности символа, что и определяет термин «невозвращающийся в нуль» код. Длительность импульсов и пауз в NRZ-кодах равна длительности одного символа (бита) информации.

Наиболее известны следующие виды цифровой модуляции:

- амплитудная манипуляция (АМн; иначе ИКМ-АМ, или цифровая амплитудная модуляция — ЦАМ; *amplitude shift keying — ASK*);
- фазовая манипуляция (ФМн; иначе ИКМ-ФМ, или цифровая фазовая модуляция — ЦФМ; *phase shift keying — PSK*);
- частотная манипуляция (ЧМн; иначе ИКМ-ЧМ, или цифровая частотная модуляция — ЦЧМ; *frequency shift keying — FSK*) и ее разновидность минимальная частотная манипуляция (МЧМн);
- квадратурная амплитудная манипуляция.

Сейчас (часто в радиосвязи и других видах связи, и модемах) используют и другие виды цифровой модуляции.

Типичный подход при осуществлении передачи дискретной последовательности символов сообщения состоит в следующем. Каждому из возможных значений символа сопоставляется некоторый набор параметров несущего колебания. Эти параметры поддерживаются постоянными в течение интервала $T$, т. е. до прихода следующего символа. Фактически это означает преобразование последовательности чисел $\{n_k\}$ в ступенчатый сигнал $s_n(t)$ с использованием кусочно-постоянной интерполяции:

$$s_n(t) = f(n_k),\ kT \le t \le (k+1)T.$$

Здесь $f$ — некоторая функция преобразования. Полученный сигнал $s_n(t)$ далее используют в качестве модулирующего сигнала обычным способом.

Для передачи цифрового первичного сигнала по каналу связи применяют различные несущие. Рассмотрим гармоническую несущую.

При цифровой модуляции закодированный первичный аналоговый сигнал $e(t)$, представляющий собой последовательность кодовых символов $\{e_n\} = e_n^{(k)}$ ($n = 0, 1, 2, 3, \ldots$ — порядковый номер символа; $k \in 0$; $(m-1)$ — номер позиции кода; $m$ — основание кода, т. е. число различных его элементов), преобразуется в последовательность элементов (посылок) сигнала $\{u_n(t)\}$ путем воздействия кодовых символов на высокочастотное несущее колебание $u_н(t)$. Обычно посредством модуляции частота или фаза несущего в радиоимпульсе изменяется по закону, определяемому цифровым кодом.

На рис. 2.53 приведены формы цифрового сигнала (в теории связи его называют *канальным сигналом*) при двоичном коде для различных видов дискретной или цифровой модуляции. При этом для наглядности использованы униполярный и биполярный NRZ-коды, показанные на рис. 2.53, *а*, *б* соответственно.

Битовому символу «1» при ИКМ-АМ (рис. 2.53, *в*) соответствует передача несущего колебания в течение времени $\tau_н$ (длительность посылки), символу «0» — отсутствие колебания (пауза) на таком же временно́м интервале.

В случае применения ИКМ-ЧМ (рис. 2.53, *г*) передача несущего с частотой $f_0$ соответствует символу «1», а передача колебания с частотой $f_1$ — символу «0».

При двоичной ИКМ-ФМ (рис. 2.53, *д*) фаза несущей меняется на 180° при каждом переходе символов от 1 к 0 и от 0 к 1.

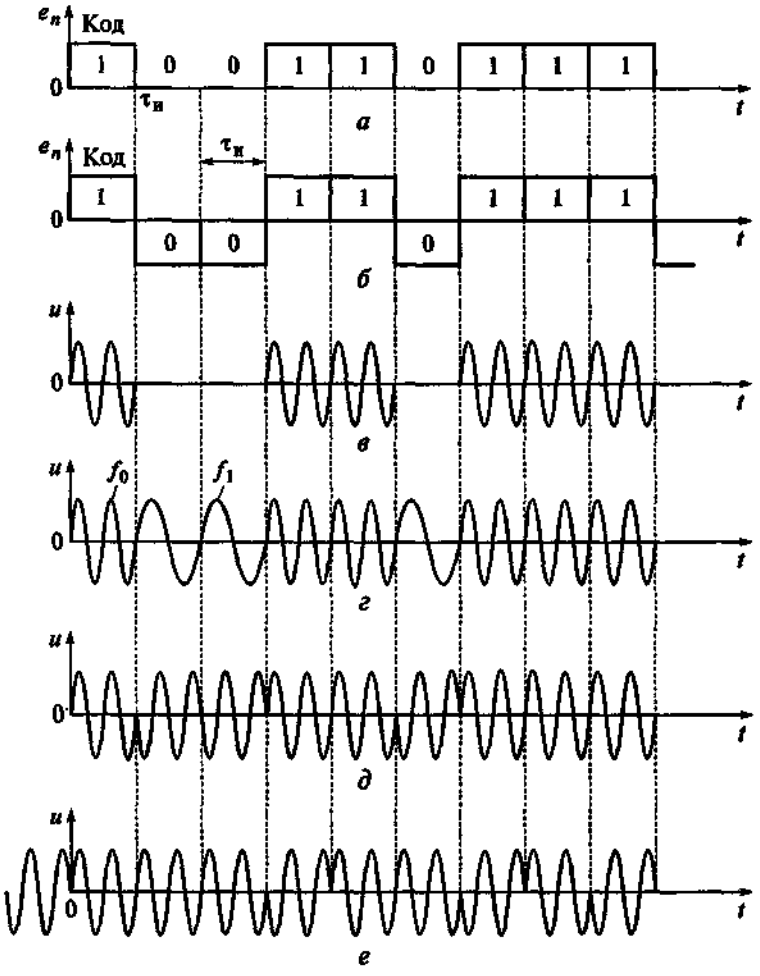



**Рис. 2.53.** Формы сигналов при различных видах цифровой модуляции двоичным кодом:
*а* — униполярный код; *б* — биполярный код; *в* — ИКМ–АМ;
*г* — ИКМ–ЧМ; *д* — ИКМ–ФМ; *е* — ОФМ

В модемах систем сотовой связи (например, с МДВР) применяют дискретную систему сигналов с *относительной фазовой* (*дифференциальной*; *фазоразностной*) *манипуляцией* (ОФМ; *differential phase shift keying — DPSK*), часто называемой *многопозиционной амплитудно-фазовой манипуляцией* (рис. 2.53, *е*).

В более широком смысле цифровую модуляцию следует рассматривать как преобразование последовательности кодовых символов 0, 1, ..., $m - 1$ в определенные отрезки гармонического высокочастотного сигнала $u_i(t)$, где $i = 0, 1, ..., m - 1$ — номер передаваемого символа. При этом вид гармонического высокочастотного сигнала $u_i(t)$ может быть произвольным. На практике его выбирают таким, чтобы удовлетворить требованиям, предъявляемым к системе радиосвязи (в частности, по полосе частот и скорости передачи информации), и чтобы сигналы могли быть выделены из воздействующих помех.

Цифровой сигнал $e_{ц}(t)$ при последовательной передаче символов сообщения $e_n^{(k)}$, следующих с равным тактовым интервалом $\Delta t$, опишем выражением

$$e_{ц}(t) = \sum_{n=0}^{\infty} v(t - t_n, e_n^{(k)}), \qquad (2.193)$$

где $t_n$ — момент появления $n$-го символа; $v(t, e_n^{(k)})$ — форма элементарного импульсного сигнала, представленного символом $e_n^{(k)}$.

Цифровой сигнал $e_{ц}(t)$ (2.193) образуется как линейная комбинация одинаковых элементов $v(t)$. Причем этот сигнал чаще всего является *изохронным*, т. е. отдельные кодовые символы появляются с равным тактовым интервалом $\Delta t$. В этом случае $t_n = n\Delta t$, и можно записать

$$e_{ц}(t) = \sum_{n=0}^{\infty} e_n^{(k)} v(t - n\Delta t). \qquad (2.194)$$

Как правило, в системах передачи дискретных сообщений используются двоичные коды ($m = 2$) и поэтому $\Delta t = \tau_{и}$ (см. рис. 2.53, *а*).

**Цифровая амплитудная манипуляция.** При такой модуляции скачкообразно меняется *амплитуда* несущего колебания, и она является частным случаем квадратурной манипуляции. Положим, что передается сигнал, отражаемый $M$ возможных используемых уровней амплитудной манипуляции. Символы, подлежащие передаче, должны принимать целочисленные значения, лежащие в диапазоне 0 ... $M$–1. Символу 0 соответствует значение амплитуды, равное –1, а символу $M$–1 — значение амплитуды, равное 1. Остальные уровни равномерно распределены между этими значениями. Строго говоря, в данном случае может меняться не только амплитуда, но и фаза несущего колебания (отрицательные амплитудные множители соответствуют изменению фазы на 180°).

В качестве примера на рис. 2.54 представлена временна́я диаграмма реального сигнала с 8-позиционной АМн, содержащего весь возможный набор символов. На диаграмме хорошо виден скачок фазы в середине диаграммы сигнала. Нетрудно также заметить, что фазы посылок в первой и второй половинах графика сигнала, имеющих одинаковые амплитуды, отличаются на 180° .

Канальный сигнал при ЦАМ с учетом (2.116) и (2.194) при $\varphi_0 = 0$ имеет вид

$$u_{ЦАМ}(t) = \left[ U_{н} + k_{АМ} \sum_{n=0}^{\infty} e_n^{(k)} v(t - n\Delta t) \right] \cos\omega_0 t, \qquad (2.195)$$

где $U_{н}$ и $\omega_0$ — соответственно амплитуда и частота несущего колебания.

Спектр АМн-сигнала содержит несущую и две боковые полосы, каждая из которых повторяет спектр первичного сигнала $e_{ц}(t)$. Если в формуле (2.195) несущая подавлена и $U_{н} = 0$, то имеем сигнал с *цифровой балансной амплитудной модуляцией* (ЦБАМ), или ЦАМ *без несущей*. В рассматриваемом случае элементарные сигналы $v(t)$ имеют вид прямоугольных импульсов, но они могут быть и

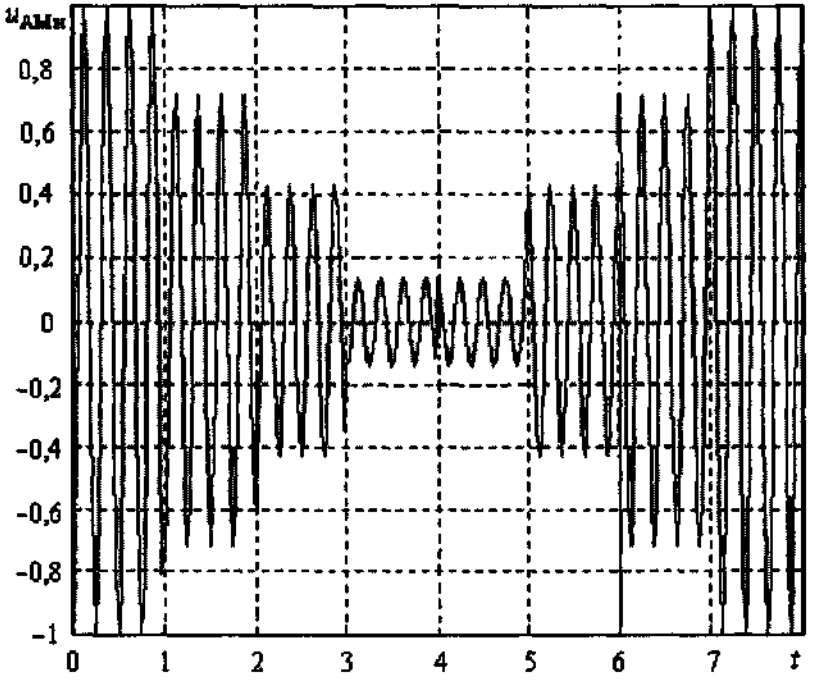



**Рис. 2.54**. Диаграмма сигнала с 8-позиционной амплитудной манипуляцией

другой формы, выбираемой из соображения ограниченности полосы частот канала связи. Вид сигнала ЦАМ, реализации модулятора и детектора упрощаются при использовании прямоугольных импульсов $\nu(t)$ с амплитудой $U_н = 1$ и длительностью $\Delta t = \tau_н$.

Демодуляция АМн-сигнала может выполняться теми же методами, что и в случае квадратурной манипуляции (путем умножения на несущее колебание). Однако наличие всего лишь двух возможных значений начальной фазы несущей, отличающихся друг от друга на 180°, делает возможной реализацию автоматической подстройки начальной фазы с помощью петли и *фазовой автоподстройки частоты* (ФАПЧ; *Phase-locked loop — PLL*; см. гл. 9).

**Цифровая фазовая манипуляция**, при которой скачкообразно меняется фаза несущего колебания. Фазовая манипуляция, предложенная Г. Найквистом в 1928 г., долгое время не находила применения из-за явления «обратной работы», возникающего вследствие невозможности восстановления на приеме опорного колебания, строго синфазного с несущей частотой принимаемого сигнала.

Канальный сигнал при ЦФМ можно записать в таком виде:

$$u_{ЦФМ}(t) = U_н \cos\left[\omega_0 t + k_{ФМ}\sum_{n=0}^{\infty} e_n^{(k)} \nu(t - n\Delta t)\right] =$$

$$= U_н \cos\left[k_{ФМ}\sum_{n=0}^{\infty} e_n^{(k)} \nu(t - n\Delta t)\right]\cos\omega_0 t - U_н \sin\left[k_{ФМ}\sum_{n=0}^{\infty} e_n^{(k)} \nu(t - n\Delta t)\right]\sin\omega_0 t. \qquad (2.196)$$

На практике цифровую фазовую манипуляцию применяют при небольшом числе возможных значений начальной фазы — как правило, 2, 4 или 8. Кроме того, при приеме сигнала сложно определить *абсолютное* значение начальной фазы; значительно проще измерить *относительный* фазовый сдвиг между двумя соседними символами. Поэтому обычно используется *фазоразностная* манипуляция (часто относительная фазовая манипуляция, дифференциальная фазовая манипуляция; *differential phase shift keying — DPSK*).

При ОФМ в зависимости от значения информационного элемента изменяется только фаза сигнала при неизменной амплитуде и частоте, при этом фазу канального сигнала отсчитывают не от некоторого эталона, а от фазы предыдущего элемента. Например, символ «0» передается отрезком синусоиды с начальной фазой предшествующего элемента сигнала, а символ «1» — таким же отрезком с начальной фазой, отличающейся от начальной фазы предшествующего элемента на 180°. При ОФМ передача сообщения начинается с посылки одного не несущего передаваемой информации элемента, который служит лишь опорным (эталонным) сигналом для сравнения фазы последующего элемента (см. рис. 2.53, *д*). Каждому информационному биту ставится в соответствие не абсолютное значение фазы, а ее изменение относительно предыдущего значения.

В цифровом телевидении для передачи по спутниковым трактам и в наземном телевещании при тяжелых условиях приема используется двукратная, или *четырехфазная ОФМ* (ОФМ-4; другое название — *квадратурная относительная фазовая модуляция* — КОФМ; англ. — *Quadrature phase shift keying — QPSK*; кстати, ОФМ-4 совпадает с КАМ-4). Модуляция *QPSK* обеспечивает необходимый компромисс между скоростью передачи информации и помехоустойчивостью системы и применяется как самостоятельно, так и в комбинациях с другими методами. Этот вид модуляции основан на передаче четырех сигналов, каждый из которых несет информацию о двух битах (дибите) исходной двоичной последовательности. Обычно используется два набора фаз: в зависимости от значения дибита (00, 01, 10 или 11) фаза сигнала может измениться на 0, 90, 180, 270 или 45, 135, 225, 315° соответственно. При этом, если число кодируемых бит более трех (8 позиций поворота фазы), резко снижается помехоустойчивость ОФМ. Потому для высокоскоростной передачи данных ОФМ использовать не рекомендуется.

**Межсимвольные искажения**. Несмотря на то что амплитуды гармонических составляющих в спектре последовательности прямоугольных импульсов уменьшаются с ростом частоты (см. рис. 2.16, *а*), этот спад все же довольно медленный. Наиболее неприятным результатом искажений импульсов в каналах связи является то, что в приемнике переходный процесс от одного импульса не завершается к моменту прихода следующего (рис. 2.55). Импульсы на выходе канала накладываются друг на друга, искажаясь еще больше. Прямоугольный импульс (рис. 2.55, *а*), поданный на вход канала связи в результате искажений, обусловленных ограниченностью полосы пропускания канала и межсимвольной интерференции, может иметь на выходе амплитуду меньше, чем у входного импульса, что уменьшает отсчетное значение, повышает чувствительность к шу-

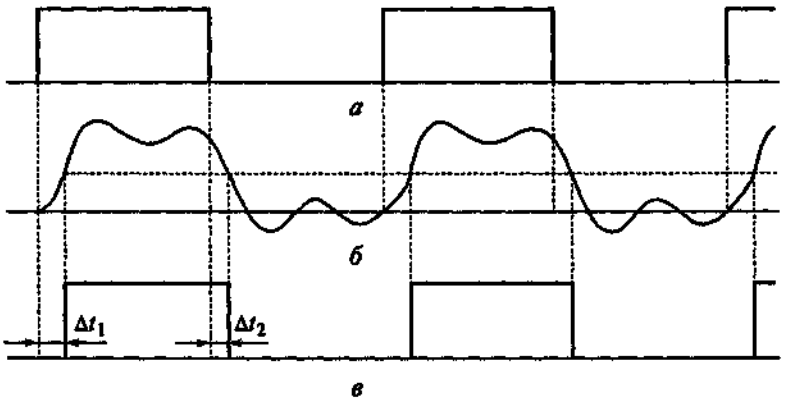



**Рис. 2.55.** Импульсы при межсимвольной интерференции:
*а* — передаваемые; *б* — искаженные; *в* — восстановленные

мам и помехам и увеличивает вероятность ошибки в определении отсчетного уровня (рис. 2.55,*б*). Длительность выходного импульса, оцениваемая на уровне половины максимального значения, также отклоняется от заданной величины (такие временны́е отклонения приводят к *краевым искажениям* принятого импульса, показанным на рис. 2.55, *в* — см. сдвиги восстановленного импульса на интервалы $\Delta t_1$ и $\Delta t_2$). Взаимные искажения, возникающие в результате наложения импульсов, называют *межсимвольной интерференцией* (МСИ; *Inter symbol interference — ISI*) или межсимвольными искажениями. При приеме таких сигналов необходимо не только восстанавливать их форму, но определять их временно́е положение. Для передачи импульсов без искажений необходима бесконечная полоса частот канала связи. Чтобы искажения были малозаметны, граничное значение полосы частот должно быть во много раз больше величины, обратной длительности импульса. Однако все реальные каналы имеют конечную полосу пропускания, что приводит к искажениям формы импульсов.

Для оценки влияния межсимвольных искажений в случае передаваемых последовательностей двоичных символов прибегают к *глазковым диаграммам*.

Глазковая диаграмма — изображение, полученное в результате измерения отклика приемника на заданные низкочастотные сигналы (импульсы или символы). По существу любая глазковая диаграмма представляет собой типичный аналог «осциллограммы» демодулированного непрерывного сигнала, построенную при длительности «прямого хода развертки», равной одному символьному такту $T$ ($-T/2 \ldots T/2$), и бесконечном «времени послесвечения экрана». В точках оптимальной дискретизации линии на такой диаграмме образуют узкие пучки — *глазковую диаграмму*, свободное пространство между которыми по форме напоминает *раскрытый глаз*. При снятии глазковой диаграммы на вертикальные пластины осциллографа подается отклик приемника на случайную последовательность импульсов, а на горизонтальные — пилообразное напряжение символьной частоты. Другими словами, горизонтальная временна́я развертка осциллографа устанавливается равной длительности символа (импульса).

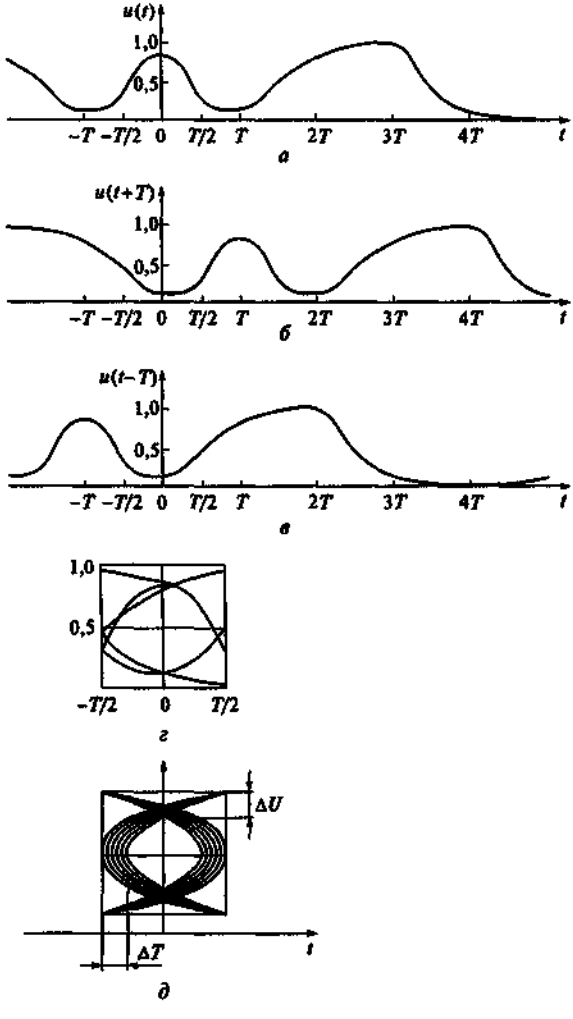



**Рис. 2.56.** Построение глазковой диаграммы:
*а* — принятые импульсы; *б – в* — смещения принятых импульсов; *г* — временна́я диаграмма смещенных отрезков принятых сигналов; *д* — глазковая диаграмма

В течение каждого сигнального такта $T$ очередной сигнал накладывается на семейство кривых в интервале $(0, T)$. Построение глазковой диаграммы показано на рис. 2.56, где она получена при двоичной биполярной передаче. Выходной сигнал (рис. 2.56, *а*) соответствует длинной последовательности двоичных символов, передаваемых с периодом $T$. Фиксируя сигналы на интервале $-T/2 \ldots T/2$ при разных, но кратных периоду, смещениях принятого сигнала (рис. 2.56, *б*, *в*) и складывая их, получают временну́ю диаграмму (рис. 2.56, *г*). Увеличивая длину последовательности символов, составленной из разных сочетаний передаваемых двоичных знаков, можно получить полную глазковую диаграмму (рис. 2.56, *д*).

Поскольку символы поступают из случайного источника, они могут быть как положительными, так и отрицательными, и отображение послесвечения

электронного луча позволяет видеть изображение, имеющее форму глаза. Ширина открытия глаза указывает время, когда должна быть проведена выборка сигнала. Оптимальное время взятия выборки соответствует максимально раскрытому глазу, что дает максимальную защиту от воздействия помех. Если в системе не используют фильтрацию, т. е. если передаваемым информационным импульсам соответствует бесконечная полоса, то отклик системы дает импульсы идеальной прямоугольной формы. В этом случае диаграмма будет выглядеть уже не как глаз, а как прямоугольник. Диапазон разностей амплитуд, обозначенный $\Delta U$, является мерой искажения, вызванного межсимвольной интерференцией. Диапазон разностей времен перехода через нуль, обозначенный $\Delta T$, определяет краевые искажения относительно исходной ширины $-T/2 \ldots T/2$ (мера неустойчивой синхронизации). Глазковую диаграмму используют для качественной оценки степени межсимвольной интерференции. «Отрытые» глаза указывает на качество системы связи. По мере закрытия глаза межсимвольная интерференция увеличивается, а по мере открытия — уменьшается.

Искажения формы принятых импульсов, зафиксированные глазковой диаграммой, зависят как от граничного значения полосы частот канала связи и формы его частотной характеристики, так и от скорости передачи данных. Чем уже полоса и чем выше скорость передачи, тем больше искажения импульса. Какие искажения признать допустимыми и какую установить скорость передачи данных зависит от требований, предъявляемых к системе связи, и от уровня помех в канале связи. Если, например, принять, что для минимизации искажений полоса канала связи должна быть в 10 раз больше частоты первого нуля в амплитудном спектре прямоугольного импульса (см. рис. 2.16, *а*), то за время $T$ в полосе $\Delta f = 10/T$ будет передан один двоичный символ. Можно ввести удельную скорость передачи данных, т. е. скорость передачи данных в расчете на единицу полосы частот. В приведенном примере $v = 1/T \approx 0{,}1$(бит/с)/Гц, что является весьма малой величиной. Однако прямое увеличение частоты может привести к значительным межсимвольным искажениям.

Обычно в системах связи с ЦФМ в качестве кодовых символов используются прямоугольные импульсы $v(t)$ единичной амплитуды и длительности $\Delta t = \tau_и$. Поэтому вид канального сигнала, схемная реализация модулятора и детектора существенно упрощаются. В этом случае отсутствует межсимвольная интерференция, и вместо (2.196) при $e_n = \pm 1$ имеем

$$u_{\text{ЦФМ}}(t) = U_н \cos(\Delta\theta)\cos\omega_0 t - U_н\left[\sum_{n=0}^{\infty} e_n^{(k)} v(t - n\Delta t)\right]\sin(\Delta\theta)\sin\omega_0 t, \qquad (2.197)$$

где $\Delta\theta = k_{\text{ФМ}}$ — разность фаз для двух позиций кода (при этом девиация фазы несущего колебания $\Delta\varphi = \Delta\theta/2$).

Спектральный состав ЦФМ-сигнала (2.197) не отличается от ЦАМ-сигнала (2.195). Отметим, что если разность фаз при двухпозиционном коде $\Delta\theta = \pi$ (используются разнополярные сигналы), несущая в спектре ЦФМ-сигнала исчезает, когда символы с различными значениями появляются с равной вероятностью.

**Цифровая частотная манипуляция.** При частотной манипуляции каждому возможному значению передаваемого символа сопоставляется своя частота несущей, т. е. в течение любого символьного интервала передается гармоническое колебание с частотой, соответствующей текущему символу. При этом возможны различные способы, различающиеся выбором начальной фазы отдельных синусоидальных посылок.

Первый способ характерен тем, что все посылки, соответствующие одному закону изменения передаваемого символа, имеют одинаковую начальную фазу, а значит, являются идентичными. При этом можно заранее сформировать наборы отсчетов для всех возможных дискретных символов. Тогда осуществление частотной манипуляции сводится к последовательной передаче заранее рассчитанных последовательностей отсчетов, соответствующих поступающим символам. Однако если используемые частоты манипуляции не кратны символьной скорости, сформированный таким образом ЧМн-сигнал будет содержать разрывы (скачки) на стыках символов. Вследствие этого спектр сигнала будет иметь всплески на частотах, кратных символьной скорости.

Второй способ связан с непрерывной генерацией колебаний всех $m$ необходимых частот, формируемых, как правило, синтезатором частот и подключением этих сигналов в соответствии с поступающими символами. При этом в течение каждого символьного интервала передается гармоническое колебание с определенной частотой, соответствующей текущему символу. При данном способе возможно появление скачков на стыках символов. В результате этого возникает паразитная амплитудная модуляция сигнала, и его *пик-фактор* (отношение пиковой, т. е. максимальной, и средней мощностей сигнала; *peak-factor*) увеличивается. С целью сужения спектра и сохранения минимального пик-фактора канального сигнала необходимо обеспечить непрерывность изменения мгновенной фазы сигнала. Вместе с тем, поскольку начальные фазы посылок меняются от символа к символу, скачки возникают не на всех стыках, и их величина оказывается различной. В результате возникающие из-за скачков всплески спектра в данном случае менее заметны.

Третий способ отличается тем, что поступающие для передачи сигнальные символы управляют скоростью линейного нарастания текущей фазы, а частотно-манипулированный сигнал формируется путем вычисления косинуса этой фазы. При этом фазовая функция, а значит, и сам ЧМн-сигнал оказываются непрерывными (не имеющими скачков). Данный способ сложнее в реализации, но он дает наиболее компактный спектр сигнала и находит наибольшее применение в системах связи. ЧМн-сигнал, полученный таким образом, называется *частотно-манипулированным сигналом с непрерывной фазой* (*continuous phase frequency shift keying — CPFSK*).

Частотная манипуляция весьма помехоустойчива, поскольку помехи телефонного канала искажают в основном амплитуду, а не частоту сигнала. Однако при ЧМн неэкономно расходуется ресурс полосы частот телефонного канала. Поэтому этот вид модуляции применяется в низкоскоростных протоколах, позволяющих осуществлять связь по каналам с низким отношением сигнал/шум.

Цифровую частотную модуляцию с непрерывной фазой обозначим ЦЧМНФ. В системах ЦЧМНФ с кодовыми символами в виде прямоугольных импульсов единичной высоты мгновенная частота сигнала меняется по закону

$$\omega(t) = \omega_0 + k_{\text{ЧМ}} \sum_{n=0}^{\infty} e_n^{(k)} v(t - n\Delta t), \tag{2.198}$$

а канальный сигнал

$$u_{\text{ЦЧМНФ}}(t) = U_{\text{н}} \cos\left[\omega_0 t + k_{\text{ЧМ}} \sum_{n=0}^{\infty} e_n^{(k)} v(t - n\Delta t) + \varphi_0\right], \tag{2.199}$$

где $\varphi_0$ — начальная фаза несущего колебания; $k_{\text{ЧМ}}$ — девиация частоты.

Представим сигнал (2.199) на отрезке времени $[0, \Delta t]$ при передаче $i$-й позиции символа в виде

$$u_i(t) = U_{\text{н}} \cos(\omega_i t + \varphi_{0,k}), \tag{2.200}$$

где $\omega_i = \omega_0 + (2\pi/\Delta t)$ — частота, соответствующая $i$-й позиции символа; $\varphi_{0,k}$ — начальная фаза к данному ($k$-му) тактовому интервалу; $k = 1, 2, 3, \ldots$

В теории связи доказано, что при осуществлении ЦЧМНФ ортогональность сигналов (2.200) обеспечивается тогда, когда частотный сдвиг

$$\Delta f_{\min} = \frac{\Delta\omega_{\min}}{2\pi} = \frac{1}{2\Delta t}. \tag{2.201}$$

Теоретические исследования также установили, что оптимальные результаты в формировании ЦЧМ-сигнала достигаются при периоде следования модулирующих сигналов $\Delta t = \tau_{\text{и}}$, где $\tau_{\text{и}}$ — длительность бита или символа (элементарного сигнала — прямоугольного импульса или паузы).

Цифровую двухпозиционную (бинарную) ЧМн с непрерывной фазой и частотным сдвигом (2.201) называют частотной *манипуляцией с минимальным* (частотным) *сдвигом* (ММС; *Minimum Shift Keying — MSK*).

Нетрудно показать, что индекс частотной модуляции в этой цифровой системе связи $m = 0{,}5$. Определим данный индекс как отношение девиации частоты $k_{\text{ЧМ}} = \Delta\omega/2$ (максимальное отклонение от средней частоты) к частоте модулирующего сигнала $\Omega = \pi/\Delta t$

$$m = \frac{k_{\text{ЧМ}}}{\Omega} = \frac{\Delta\omega\Delta t}{2\pi} = \Delta f \Delta t. \tag{2.202}$$

С учетом соотношений (2.201) и (2.202) получаем, что индекс модуляции для сигналов с ММС действительно равен 0,5. Отсутствие скачков фазы в системах с ЦЧМНФ благоприятно сказывается на форме амплитудного спектра передаваемого сигнала. При $m = 0{,}5$ амплитудный спектр ЦЧМНФ-сигнала весьма узок и сосредоточен вблизи частоты несущей, однако при $m > 1$ амплитудный спектр становится более широким.

Рассмотрим упрощенно принцип формирования ЧЧМ-сигналов. Для систем ЧЧМ с индексом частотной модуляции $m$ = 0,5 широко используется квадратурный метод модуляции со сдвигом модулирующих функций. Его действие сводится к сглаживанию закона изменения фазы $\varphi(t)$ при частотной манипуляции. Математически доказано, что огибающая энергетического спектра ЧЧМ-сигналов достаточно резко убывает по мере удаления от несущей при увеличении степени сглаживания закона изменения фазы. Значит, можно повысить эффективность ЧЧМ путем сужения рабочей полосы. Однако, как уже было показано, беспредельно сужать рабочую полосу не удается — оптимальные результаты в формировании ЧЧМ-сигнала достигаются при периоде следования модулирующих сигналов $\Delta t = \tau_и$.

При осуществлении ММС для передачи логической 1 (импульса) фазу несущей $\varphi(t)$ линейно увеличивают таким образом, чтобы за время длительности такта $\Delta t$ приращение фазы $\Delta\varphi$ составляло $\pi/2$; передаче же логического 0 (паузы) должно соответствовать линейное уменьшение фазы несущей на $\pi/2$. Это наименьшее возможное изменение фазы несущей, распознаваемое в приемнике при данном типе модуляции, которую можно трактовать как *частотно-фазовую*. Зная величину изменения фазы за один такт $\Delta\varphi = \pi/2$, нетрудно рассчитать *девиацию частоты* $\Delta f$. Учитывая, что $\Delta\varphi = 2\pi\Delta f\Delta t$, находим $\Delta f = 0{,}25/\Delta t$.

**Многопозиционные сигналы**. Эффективность систем передачи цифровых сообщений можно существенно повысить путем использования многопозиционных (многоуровневых) сигналов, которые можно применять при большой мощности сигнала без риска увеличить вероятность ошибки при определении значения принимаемого сигнала. Увеличение числа позиций, или уровней, позволяет увеличить удельную скорость модуляции, но лишь за счет увеличения мощности излучаемого колебания. То же самое можно сказать и о выборе корректирующих кодов. Выбор сигналов и кодов в этих случаях является определяющим для построения высокоэффективных *кодемов* (согласованных между собой кодеков и модемов).

В последнее время большой интерес проявляется к многопозиционным сигналам с АФМ, которые можно реализовать схемой квадратурной модуляции. В системах связи широко используют многопозиционные цифровые сигналы с АФМ (ЦАФМ), когда начальная фаза несущей принимает $m$ значений. В системах АФМ в течение интервала передачи одного элемента фаза и амплитуда принимают значения, выбранные из ряда возможных дискретных значений амплитуд и фаз. Каждая комбинация значений амплитуды и фазы отображает один из многопозиционных сигналов группового сигнала. Сигналы АФМ можно формировать, например, путем многоуровневой амплитудной и фазовой модуляции двух квадратурных (сдвинутых на $\pi/2$) колебаний несущей частоты. При многопозиционном сигнале сдвиг фазы между сигналами равен $\pi/m$.

Формирование четырехпозиционного сигнала показано на рис. 2.57. Пары соседних значений двоичных данных (длительность каждого символа $\tau_и$) передаваемого первичного сигнала $u_1(t)$ (рис. 2.57, *а*) определяют один из четырех

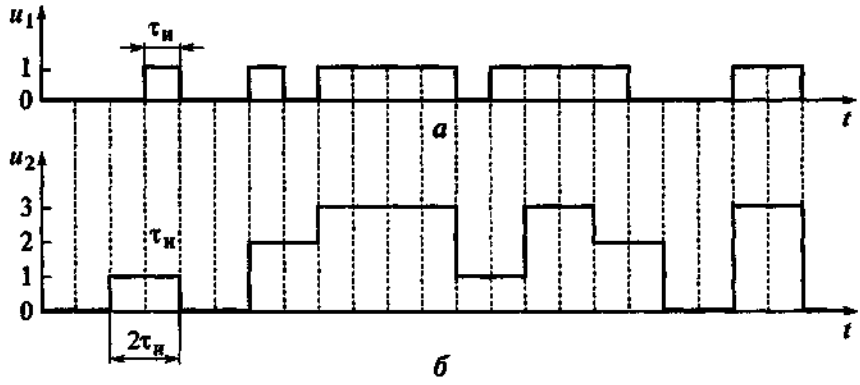



**Рис. 2.57.** Формирование четырехпозиционного сигнала:
*а* — передаваемый первичный сигнал; *б* — четырехпозиционный сигнал

уровней, который занимает сигнал $u_2(t)$ (рис. 2.57, *б*). Пара двоичных символов 00 соответствует уровню (амплитуде) 0, пара 01 — уровню 1, пара 10 — уровню 2 и пара 11 — уровню 3. Сигнал $u_2(t)$ меняется в 2 раза реже, чем исходный $u_1(t)$, для его передачи требуется в 2 раза меньшая полоса частот, следовательно, использование четырехпозиционного сигнала позволяет увеличить удельную скорость передачи в 2 раза. Но надо помнить, что применение многопозиционных сигналов связано со значительным увеличением их мощности.

**Дифференциальная импульсно-кодовая модуляция.** Как известно, для речевого сигнала более вероятны низкочастотные составляющие спектра. Это означает, что мгновенные значения дискретных отсчетов сигнала в соседних точках дискретизации с большой вероятностью мало отличаются друг от друга. Благодаря этому можно довольно точно *предсказывать* сигнал, формируя линейную комбинацию предыдущих его значений. Эта разность преобразуется в цифровую форму. Остаточная ошибка предсказания оказывается малой, и для ее квантования можно использовать меньшее число уровней. Поэтому можно вместо кодирования и дальнейшей передачи отсчетов передавать по тракту связи кодированные значения разности соседних отсчетов, по которым на приемной стороне восстанавливаются значения отсчетов сигнала. Такой метод передачи сигналов называют *дифференциальной импульсно-кодовой модуляцией* (ДИКМ; *differential pulse code modulation — DPCM*; изобрел К. Катлер; США).

Дельта-модуляция (ДМ) представляет собой вариант ДИКМ, где для кодирования разностного сигнала используется только один бит. Этот бит служит для того, чтобы увеличить или уменьшить оценочный уровень.

**Манипуляция с частотным уплотнением и ортогональными несущими.** Новым способом модуляции, предложенным сравнительно недавно, является *частотное уплотнение с ортогональными несущими* (ЧУОН; *Orthogonal frequency division multiplexing — OFDM*). Подобно квадратурной модуляции, этот способ использует ортогональные несущие, но в отличие от квадратурной модуляции частоты этих несущих не являются одинаковыми, они расположены в

некотором диапазоне, отведенном для передачи данных путем модуляции. Частоты несущих соответствуют уравнению $e_n(t) = \cos[2\pi(f_0 + n/T_s)t]$, где $f_0$ — начало частотного интервала, в котором производится частотное уплотнение, $n$ — номер несущей, находящийся в диапазоне от 0 до $(N–1)$, т. е. всего несущих $N$; $T_s$ — длительность интервала передачи одного символа.

### Представление модулированных сигналов амплитудно-фазовыми диаграммами. Сигнально-кодовые конструкции

В разд. 2.1 отмечалось, что удобным и наглядным представлением модулированного колебания является амплитудно-фазовая диаграмма (*amplitude phase diagram*; *сигнальная конструкция — signal structure*; *сигнальное созвездие — signal constellation*; иногда используют устаревший термин «пространственная диаграмма»; *space diagram*), представляемая в полярной системе координат.

Строят амплитудно-фазовую диаграмму следующим образом (рис. 2.58).

Пусть имеются временны́е диаграммы импульсов с несущим колебанием, отражающие единичный $e_1 = e_1(t) = \cos(\omega_0 t + \varphi_0)$ и нулевой $e_2 = e_2(t) = 0$ передаваемые двоичные символы (рис. 2.58, *а*, *б*). Отложим по горизонтальной оси диаграммы фазу (от 0 до 360°), а по вертикальной — амплитуду передаваемого сигнала. Теперь для каждого двоичного сигнала у нас есть два параметра — амплитуда и фаза, т. е. две координаты. Значит, всю информацию о сигналах $e_1$ и $e_2$ можно отобразить соответствующими векторами (впрочем, один из которых в данном случае с нулевой амплитудой) на комплексной плоскости — вещественная и мнимая оси обозначены соответственно как Re и Im (рис. 2.58, *в*).

Из геометрии известно, что при вращении комплексного вектора с частотой исследуемого колебания его проекция на действительную ось дает вещественный сигнал. Для упрощения диаграммы, особенно при отображении сигналов современных многопозиционных видов модуляции, обычно изображают только конечные точки векторов, исходящих из начала координат, а сами векторы, как правило, опускают. В этом случае можно перейти к диаграмме, на которой два колебания вместо векторов представлены точками $e_1$ и $e_2$ (рис. 2.58, *г*).

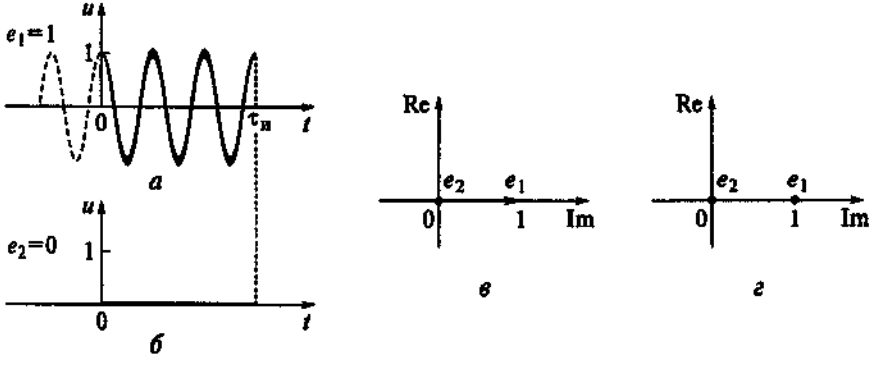



**Рис. 2.58.** К построению амплитудно-фазовой диаграммы сигналов:
*а, б* — временны́е диаграммы; *в* — векторные диаграммы; *г* — сигнальные точки

Часто опускают и сами оси Re и Im, подразумевая, что они проходят через центр симметрии фигуры. Такую картинку называют точечной *амплитудно-фазовой диаграммой* модулированных сигналов. Из этой диаграммы легко определяется распределение амплитуд сигналов в любом фазовом интервале.

Пользуясь геометрической трактовкой, каждый сигнал любого вида модуляции можно представить вектором. Причем, концы сигнальных векторов отображают в виде точек, координаты которых определяются значениями амплитуды (по вертикальной оси — действительной оси Re) и фазы (по горизонтальной оси — мнимой оси Im; фаза может изменяться от 0 до 360°) сигнала. По существу эти точки представляются комплексными числами. Каждая точка называется *сигнальной*. Совокупность сигнальных точек (векторов) передаваемых символов и образует амплитудно-фазовую диаграмму, или *сигнальное созвездие*. Форма сигнального созвездия соответствует виду модуляции, а расстояния между точками характеризуют помехоустойчивость. Не вдаваясь особенно в подробности, скажем, что помехоустойчивость тем выше, чем больше расстояние $d$ между ближайшими точками созвездия на комплексной плоскости. На практике используют созвездия, содержащие от четырех до нескольких тысяч точек.

Такой способ является очень удобным при представлении и более сложных модулированных колебаний (рис. 2.59). Квадратурно-модулированное колебание, изображенное на рис. 2.52, *д*, представлено с помощью амплитудно-фазовой диаграммы в векторной форме на рис. 2.59, *а*. Как видно, пространство комплексной плоскости используется не слишком эффективно — занят только один квадрант. На рис. 2.58, *б* показана КАМ с четырьмя одинаковыми фазовыми сдвигами в 90° (КАМ-4; *QAM*-4), вектор которого при модуляции двух квад-

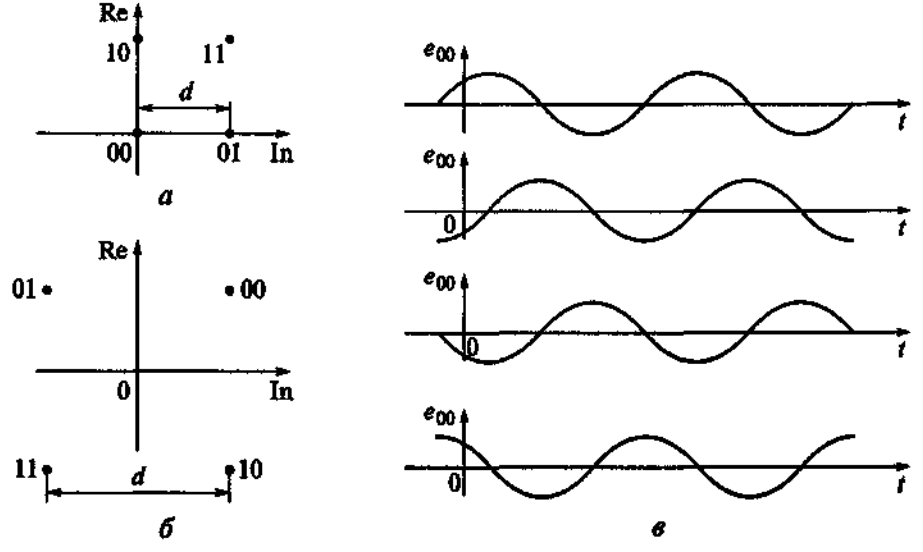


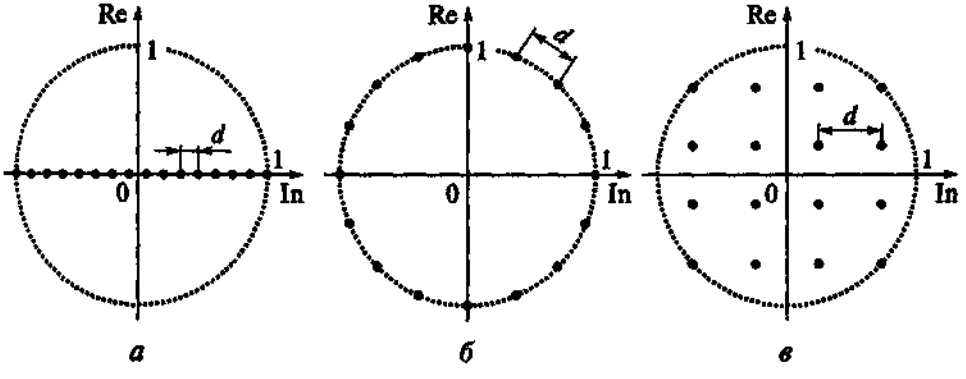

**Рис. 2.59.** Расположение сигнальных точек на амплитудно-фазовых диаграммах:
*а* — в квадратурной ЦФМ; *б* — в 4-позиционной ЦФМ;
*в* — временные диаграммы сигналов, соответствующих четырем положениям вектора 4-ЦФМ

ратурных компонентов занимает также 4 точки, но уже в четырех квадрантах, что повышает помехоустойчивость системы модуляции. Временные диаграммы сигналов, соответствующих этим положениям вектора, показаны на рис. 2.59, *в*. Амплитуды сигналов равны, поэтому модуляцию называют *четырехпозиционной цифровой фазовой модуляцией* (КФМ-4 — квадратурным КФМ-сигналом) или это и есть упоминаемая ранее QPSK. Поскольку в каждом канале осуществляется амплитудная манипуляция, этот вид модуляции называют также QASK.

И еще о помехоустойчивости систем цифровой и импульсной модуляции. В частности, отметим, что за счет использования двумерного характера гармонического несущего колебания квадратурная манипуляция обеспечивает большую помехоустойчивость (т. е. меньшую вероятность ошибки приема символа), чем АМн и ФМн. Как уже указывалось, помехоустойчивость модулированных сигналов тем выше, чем больше расстояние $d$ между ближайшими точками созвездия на комплексной плоскости. Сравним для примера помехоустойчивости 16-позиционных амплитудной, фазовой и квадратурной манипуляций (рис. 2.60).

На практике применяют различные многопозиционные системы ЦФМ (они, кстати, относятся к системам с нелинейной модуляцией; заметим, что системы сигналов с амплитудной модуляцией считаются линейными).

Для повышения качества передачи сообщений (что называют *минимизацией средней вероятности ошибочного приема в канале связи с шумом*) надо подбирать такие сигнальные последовательности, для которых сигнальные точки разрешенных кодовых комбинаций находятся друг от друга на максимально возможном расстоянии. Задача согласования модулированного сигнала с радиоканалом решается, в частности, оптимизацией формы и числа точек двумерного сигнального созвездия. В частности, сигнал QAM, переносящий $n$ бит/символ, т. е. имеющий $2n$ точек сигнального созвездия, обладает следующим интересным свойством. Если $n$ — целое четное число, то сигнальное созвездие представляет простое отображение двух независимых квадратурных каналов и обладает квадратной формой, а точки созвездия имеют координаты в виде нечетных чисел. Если число $n$ — число нечетное, то созвездие имеет крестообразную форму при рас-




*а* *б* *в*

**Рис. 2.60.** Сравнение помехоустойчивости различных 16-позиционных манипуляций:
*а* — амплитудной; *б* — фазовой; *в* — квадратурной

положении точек в узлах той же прямоугольной координатной сетки, что и для четного $n$. Минимальное относительное расстояние между любыми двумя точками созвездия равно двум. При увеличении порядка созвездия (числа бит на символ) на единицу надо повышать мощность передатчика на 3 дБ (в 2 раза).

**Сигнально-кодовые конструкции (треллис-модуляция).** Многопозиционные сигналы с так называемой *плотной упаковкой* (например, АФМ, КАМ, ЦФМ и др.) обеспечивают высокую удельную скорость передачи информации за счет снижения помехоустойчивости. С другой стороны, применение помехоустойчивого решетчатого кодирования (вид сверточного кодирования, см. гл. 9) позволяет повысить помехоустойчивость при определенном снижении удельной скорости. Каждый из этих способов дает выигрыш по одному показателю в обмен на ухудшение другого. Вместе с тем во многих случаях важным является одновременное повышение как помехоустойчивости, так и скорости передачи информации. Решение этой задачи возможно при использовании ансамблей многопозиционных сигналов совместно с помехоустойчивым кодированием. Очевидно, что при этом необходимо сформировать такие сигнальные последовательности, точки которых в многомерном пространстве плотно упакованы (для уменьшения отношения сигнал/шум, а значит создания высокой частотной эффективности) и достаточно разнесены (чтобы обеспечить высокую энергетическую эффективность). Такие сигнальные последовательности, построенные на базе многопозиционных сигналов и помехоустойчивых кодов, называют *сигнально-кодовыми конструкциями* (СКК; *signal code structure — SCS*; в иностранной технической литературе СКК носят название *треллис-модуляции — Trellis Coded Modulation*; *TCM*).

Выбранная определенным образом комбинация многопозиционного сигнала и помехоустойчивого кода позволяет повысить помехозащищенность передачи информации наряду со снижением требований к отношению сигнал/шум в канале на 3...6 дБ. При этом число сигнальных точек увеличивается вдвое за счет добавления к информационным битам одного избыточного, образованного путем сверточного кодирования. Расширенный таким образом блок битов подвергается все той же КАМ. В процессе демодуляции производится декодирование принятого сигнала по алгоритму Витерби (см. гл. 9). Именно этот алгоритм за счет использования введенной избыточности и знания предыстории процесса приема позволяет по критерию максимального правдоподобия выбрать из сигнального пространства наиболее достоверную эталонную точку.

При треллис-модуляции выбор способов модуляции и кодирования передаваемых сигналов сводится к поиску такого заполнения сигнального пространства, при котором обеспечивается высокая скорость и высокая помехоустойчивость. Комбинирование различных ансамблей многопозиционных сигналов и помехоустойчивых кодов порождает множество вариантов сигнальных конструкций. Согласованные определенным образом варианты, обеспечивающие улучшение энергетической и частотной эффективности, и являются сигнально-кодовыми конструкциями. Задача поиска наилучшей СКК является одной из наиболее сложных задач теории связи.

В качестве помехоустойчивых кодов в системах СКК используют сверточные и каскадные коды, а многопозиционных сигналов — сигналы ФМ, АФМ и ЧМНФ. Сейчас распространены виды импульсной и цифровой модуляции:

- в спутниковой связи — QPSK, 16QAM, BPSK (*Binary Phase-Shift Keying* — бинарная фазовая манипуляция — скачкообразное переключение фазы синусоидального сигнала на $\pi$ при неизменной амплитуде, при этом фазе 0 ставится в соответствие логический нуль, а фазе $\pi$ — логическая единица);
- в радиорелейных линиях связи — BPSK, QPSK, 8PSK, 16QAM, 32QAM, 64QAM, 128QAM, 256QAM;
- в кабельных линиях связи — QPSK, 16QAM, 64QAM, 256QAM;
- в радиотелефонии систем сотовой связи — от 16QAM до 16384QAM.

И последнее. Поскольку амплитуда сигнала в сотовых системах подвергается наибольшему искажению, возникают так называемые *запрещенные точки*, т. е. точки, которые модем может не правильно распознать. Чтобы избежать этого, в очередную группу бит данных подмешивают специальные биты, по которым принимающий модем определяет, какие точки являются запрещенными.

## 2.6. Узкополосные сигналы

Современная радиоэлектроника связана с непрерывным совершенствованием способов и систем передачи информации. В радиосистемах передачи информации внедряются новые виды и сигналов, и методов их обработки. Рассмотренные модели аналоговых модулированных колебаний относятся к простейшим видам радиосигналов, имеющих ограниченный спектр, и зачастую предназначенных для облегчения их понимания. На практике часто имеют дело с радиосигналами, получаемыми при одновременной модуляции амплитуды и частоты (или фазы) несущего колебания по достаточно сложному закону. Для их представления требуются новые аналитические модели и формы представления.

### *Аналитическое представление узкополосных сигналов*

Рассмотрим особый класс сигналов с ограниченным спектром, которые возникают на выходе частотно-избирательных цепей (колебательных контуров, фильтров и пр.). Принято считать, что сигнал имеет *ограниченный спектр*, если после определенного номера все коэффициенты ряда Фурье равны нулю. Другими словами, на заданном отрезке времени сигнал представляется конечной суммой ряда Фурье. При этом говорят, что спектр сигнала ограничен частотой $F$, где $F$ — частота синусоиды при последнем ненулевом коэффициенте ряда Фурье. Одним из таких сигналов является узкополосный сигнал (процесс).

*Узкополосными* называются сигналы, спектральные составляющие которых группируются в относительно узкой по сравнению с некоторой центральной (обычно несущей) частотой $\omega_0$ полосе. Как правило, можно считать, что частота $\omega_0$ является *опорной частотой сигнала*. Однако в общем случае выбор ее достаточно произволен в пределах обозначенной полосы.

Упрощенно узкополосный сигнал можно представить в следующей форме

$$u(t) = U(t)\cos\psi(t) = U(t)\cos[(\omega_0 t + \varphi(t)]. \qquad (2.203)$$

В соотношении (2.203) $\omega_0$ — *опорная частота*; $U(t)$ — изменяющаяся во времени амплитуда, которую следует считать *амплитудной огибающей* (*огибающей мгновенных значений*) узкополосного колебания; $\psi(t)$ — *полная фаза* (иногда *фазовая функция*); $\varphi(t)$ — *начальная фаза*. В аналитическом представлении узкополосного сигнала (2.203) информация при амплитудной модуляции закладывается в огибающую $U(t)$, а при угловой — в колебательную составляющую $\cos\psi(t) = \omega_0 t + \varphi(t)$.

Создать оптимальную аналитическую (математическую) модель узкополосного сигнала можно следующим образом. Нетрудно заметить, что если $e_1(t)$ — низкочастотный сигнал, спектр которого сосредоточен в окрестности нулевой частоты, то косинусоидальное колебание $u_1(t) = e_1(t)\cos\omega_0 t$ при достаточно большом значении опорной частоты $\omega_0$ будет обладать всеми необходимыми признаками узкополосного сигнала, поскольку его спектр окажется перенесенным и сконцентрированным в малых окрестностях точек частоты $\pm\omega_0$. Узкополосным будет и аналогичный синусоидальный сигнал вида $u_2(t) = e_2(t)\sin\omega_0 t$, отличающийся от первого начальной фазой опорной гармоники на 90°.

Наиболее адекватную аналитическую модель узкополосного сигнала можно получить, составив следующую линейную комбинацию:

$$u(t) = A_u(t)\cos\omega_0 t - B_u(t)\sin\omega_0 t. \qquad (2.204)$$

Входящие в эту формулу функции времени $A_u(t)$ и $B_u(t)$ относятся к низкочастотным, поскольку их относительные изменения за период высокочастотных колебаний $T = 2\pi/\omega_0$ достаточно малы. В теории сигналов функцию $A_u(t)$ называют *синфазной амплитудой* узкополосного сигнала $u(t)$ при заданном значении опорной частоты $\omega_0$, а функцию $B_u(t)$ — его *квадратурной амплитудой*. Как видно из (2.204), синфазная и квадратурная амплитуды сдвинуты по опорной частоте друг относительно друга на 90°.

**Представление узкополосных сигналов комплексной формой.** В радиотехнике широко применяют комплексное описание сигналов, когда гармоническое колебание можно представить либо вещественной, либо мнимой частью комплексных функций. Кроме того, в различных системах передачи информации часто применяют узкополосные сигналы, спектр которых сосредоточен в окрестности некоторой частоты $\omega_0$. При анализе и первых и вторых электрических сигналов удобно пользоваться понятиями *комплексной огибающей*, *огибающей амплитуд* и *фазовой функции* сигнала.

Комплексная огибающая, объединяя в себе информацию об амплитуде и фазе исследуемого сигнала, является обобщением понятия *комплексной амплитуды*, широко используемого в теоретической радиотехнике. Введение понятия комплексной амплитуды позволяет упростить решение ряда задач, связанных со сложными сигналами.

Пусть имеется сигнал вида

$$u(t) = U_m e^{j(\omega_0 t + \varphi_0)} = \dot{U}_m e^{j\omega_0 t}, \qquad (2.205)$$

где $\dot{U}_m = U_m e^{j\varphi_0}$ — *комплексная амплитуда*; $\varphi_0$ — *начальная фаза* гармонического колебания.

Запишем функцию (2.205) в тригонометрической форме

$$u(t) = U_m e^{j(\omega_0 t + \varphi_0)} = U_m \cos(\omega_0 t + \varphi_0) + jU_m \sin(\omega_0 t + \varphi_0). \qquad (2.206)$$

Из (2.206) следует, что гармоническое колебание $u_1(t) = U_m \cos(\omega_0 t + \varphi_0)$ можно рассматривать как действительную часть комплексной функции $u(t)$.

Напомним, что условно это записывается так:

$$u_1(t) = U_m \cos(\omega_0 t + \varphi_0) = \mathrm{Re}(\dot{U}_m e^{j\omega_0 t}) = \mathrm{Re}(U_m e^{j\omega_0 t} e^{j\varphi_0}). \qquad (2.207)$$

Аналогично это же колебание можно записать в виде

$$u_2(t) = U_m \sin(\omega_0 t + \varphi_0) = \mathrm{Im}(\dot{U}_m e^{j\omega_0 t}) = \mathrm{Im}(U_m e^{j\omega_0 t} e^{j\varphi_0}). \qquad (2.208)$$

В последнем выражении, как известно, берется мнимая составляющая комплексной функции. Подобное представление позволяет использовать преимущества методов теории функций комплексной переменной с последующим возвратом к тригонометрической форме путем исключения мнимой части.

С физической точки зрения узкополосные сигналы относятся к *квазигармоническим колебаниям*. Следует попытаться так обобщить метод комплексных амплитуд, чтобы иметь возможность в рамках этого метода описывать узкополосные сигналы вида (2.204).

Введем в рассмотрение комплексную низкочастотную функцию

$$\tilde{U}_u(t) = A_u(t) + jB_u(t) \qquad (2.209)$$

и назовем ее *комплексной огибающей* узкополосного сигнала.

Нетрудно показать, что

$$u(t) = A_u(t)\cos\omega_0 t - B_u(t)\sin\omega_0 t = \mathrm{Re}[\tilde{U}_u(t)e^{j\omega_0 t}]. \qquad (2.210)$$

Комплексная огибающая, объединяя в себе информацию об амплитуде и фазе сигнала (за исключением несущей частоты $\omega_0$, которая предполагается известной), является обобщением понятия *комплексной амплитуды*, широко используемого в основах теории цепей. Это свойство комплексной огибающей, позволяющее при анализе узкополосных сигналов исключить из рассмотрения частоту $\omega_0$, имеет важное значение. Таким образом, применительно к узкополосному сигналу комплексная огибающая играет ту же роль, что и комплексная амплитуда по отношению к простому гармоническому колебанию. Однако в отличие от комплексной амплитуды гармонического колебания комплексная огибающая узкополосного сигнала в общем случае зависит от времени.

**Пример 2.9.** Узкополосный сигнал $u(t)$ содержит в своем составе два фиксированных значения частоты (рис. 2.61). При $t < 0$ он является первым гармоническим колебанием с частотой $\omega_0 = 2\pi/T_0$, а при $t > 0$ — вторым гармоническим колебанием с частотой $\omega_1 = 2\pi/T_1$; в момент времени $t = 0$ частота узкополосного сигнала изменяется скачком:

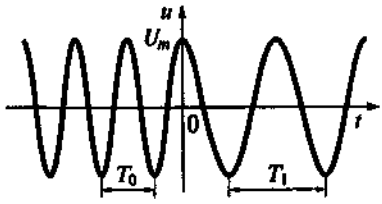



**Рис. 2.61.** Узкополосный сигнал

$$u(t) = \begin{cases} U_m \cos \omega_0 t, & t < 0; \\ U_m \cos \omega_1 t, & t > 0. \end{cases}$$

Определить комплексную огибающую этого сигнала.

Решение. Выбрав в качестве опорной круговой частоты $\omega_0$ и используя формулу (2.210), получим выражение для комплексной огибающей заданного сигнала

$$\widetilde{U}_u(t) = \begin{cases} U_m, & t < 0; \\ U_m e^{j(\omega_1 - \omega_0)t}, & t > 0. \end{cases}$$

### Физическая огибающая, полная фаза и мгновенная частота узкополосного сигнала

Комплексную огибающую (2.209) представим в экспоненциальном виде

$$\widetilde{U}_u(t) = U_u(t) e^{j\varphi_u(t)}, \qquad (2.211)$$

где $U_u(t)$ — вещественная положительная функция времени, называемая *физической огибающей* (часто, просто *огибающей*); $\varphi_u(t)$ — медленно меняющаяся во времени начальная фаза представляемого узкополосного сигнала.

Очень важно, что понятие физической огибающей узкополосного сигнала совпадает с понятием огибающей модулированных колебаний.

Физическая огибающая $U_u(t)$ и фаза $\varphi_u(t)$ связаны с синфазной и квадратурной амплитудами узкополосного сигнала следующими соотношениями:

$$A_u(t) = U_u(t)\cos[\varphi_u(t)]; \qquad (2.212)$$

$$B_u(t) = U_u(t)\sin[\varphi_u(t)]. \qquad (2.213)$$

Из (2.212) и (2.213) вытекает еще одна, обобщенная форма записи математической модели узкополосного сигнала, которая имеет практическое значение в теории модуляции

$$u(t) = U_u(t)\cos[\omega_0 t + \varphi_u(t)]. \qquad (2.214)$$

Согласно соотношению (2.214), узкополосный сигнал представляет собой достаточно сложное колебание, получающееся при одновременной модуляции несущего гармонического сигнала как по амплитуде, так и по фазовому углу.

**Пример 2.10.** Задан узкополосный сигнал, имеющий вид простого однотонального АМ-колебания

$$u(t) = U_m(1+M\cos\Omega t)\cos(\omega_0 t + \pi/4).$$

Определить комплексную огибающую $\tilde{U}_u(t)$, синфазную $A_u(t)$ и квадратурную $B_u(t)$ амплитуды этого сигнала.

Р е ш е н и е. Выберем в качестве опорной частоты узкополосного сигнала значение $\omega_0$. Тогда, согласно формуле (2.211), получим следующее выражение для комплексной огибающей узкополосного сигнала

$$\tilde{U}_u(t) = U_m(1+M\cos\Omega t)e^{j\pi/4}.$$

Поскольку $\cos(\pi/4) = \sin(\pi/4) = \sqrt{2}/2$, то, согласно (2.212) и (2.213), находим

$$A_u(t) = B_u(t) = (U_m\sqrt{2}/2)(1+M\cos\Omega t) = (U_m/\sqrt{2})(1+M\cos\Omega t).$$

По аналогии с радиосигналами с угловой модуляцией введем понятие «мгновенной (полной) фазы» узкополосного сигнала

$$\psi_u(t) = \omega_0 t + \varphi_u(t). \qquad (2.215)$$

Определим *мгновенную частоту* как производную по времени от полной фазы сигнала

$$\omega_u(t) = \omega_0 + \frac{d\varphi_u(t)}{dt}. \qquad (2.216)$$

**Основные свойства физической огибающей узкополосного сигнала.** Используя (2.212) и (2.213), выразим физическую огибающую $U_u(t)$ через синфазную и квадратурную амплитуды произвольного узкополосного сигнала:

$$U_u(t) = \sqrt{A_u^2(t) + B_u^2(t)}. \qquad (2.217)$$

Сравнив (2.209) и (2.217), нетрудно заметить, что физическая огибающая представляет собой модуль комплексной огибающей узкополосного сигнала.

Оценим влияние опорной частоты $\omega_0$ на эти две огибающие узкополосного сигнала. В общем случае комплексная огибающая узкополосного сигнала определяется неоднозначно. Если вместо опорной частоты $\omega_0$, входящей в соотношение (2.210), взять некоторую частоту $\omega_1 = \omega_0 + \Delta\omega$, то исходный сигнал $u(t)$ принимает вид

$$u(t) = \mathrm{Re}[\tilde{U}_u(t)e^{j\omega_0 t}] = \mathrm{Re}[\tilde{U}_u(t)e^{-j\Delta\omega t}e^{j\omega_1 t}]. \qquad (2.218)$$

Тогда новое значение комплексной огибающей

$$\tilde{U}'_u(t) = \tilde{U}_u(t)e^{-j\Delta\omega t}. \qquad (2.219)$$

Однако физическая огибающая узкополосного сигнала (напомним, это модуль комплексной огибающей) при изменении частоты *останется неизменной*, поскольку $|e^{-j\Delta\omega t}| = 1$.

Второе свойство физической огибающей заключается в том, что в любой момент времени для узкополосного сигнала $|u(t)| \le U_u(t)$. Справедливость этого утверждения непосредственно вытекает из формулы (2.214). Знак равенства здесь соответствует моментам времени, когда множитель $\cos[\omega_0 t + \varphi_u(t)] = 1$. По существу можно считать, что физическая огибающая действительно «огибает» амплитуды узкополосного сигнала и является его мгновенной амплитудой. Ценность понятия огибающей обусловлена тем, что в радиотехнике и системах связи широко используются амплитудные детекторы (демодуляторы), способные с высокой точностью воспроизводить огибающую узкополосного сигнала.

**Основные свойства мгновенной частоты узкополосного сигнала.** Если комплексную огибающую узкополосного сигнала представить вектором, который вращается на комплексной плоскости с некоторой постоянной угловой скоростью $\Omega$, т. е. аналитически сигнал описывается функцией $\tilde{U}_u(t) = U_u(t)e^{\pm j\Omega t}$, то, согласно формуле (2.216), мгновенная частота этого колебания постоянна во времени и поэтому $\omega_u = \omega_0 \pm \Omega$.

Можно показать, что в общем случае мгновенная частота узкополосного сигнала изменяется во времени по закону

$$\omega_u(t) = \omega_0 + \frac{d}{dt}(\operatorname{arctg}\frac{B_u}{A_u}). \qquad (2.220)$$

**Связь между спектрами узкополосного сигнала и его комплексной огибающей.** Пусть $S(\omega)$ — спектральная плотность узкополосного сигнала $u(t)$, комплексная огибающая $\tilde{U}_u(t)$ которого, в свою очередь, имеет спектральную плотность $Y_u(\omega)$. С помощью соотношения (2.210) определим связь между спектральными плотностями физического сигнала и его комплексной огибающей, записав прямое преобразование Фурье

$$S(\omega) = \int_{-\infty}^{\infty} \mathrm{Re}\left[\tilde{U}_u(t)e^{j\omega_0 t}\right]e^{-j\omega t}dt = 0{,}5\int_{-\infty}^{\infty}\tilde{U}_u(t)e^{-j(\omega-\omega_0)t}dt +$$
$$+ 0{,}5\int_{-\infty}^{\infty}\tilde{U}_u^*(t)e^{-j(\omega+\omega_0)t}dt = 0{,}5Y_u(\omega-\omega_0) + 0{,}5Y_u^*(-\omega-\omega_0), \qquad (2.221)$$

где $\tilde{U}_u^*(t)$ — комплексно-сопряженная огибающая; $Y_u^*(\omega)$ — комплексно-сопряженная спектральная плотность комплексной огибающей узкополосного сигнала $\tilde{U}_u(t)$.

Из последнего выражения следует, что спектральная плотность узкополосного сигнала $\mathbf{S}(\omega)$ может быть найдена путем переноса спектра комплексной огибающей $\mathbf{Y}_u(\omega)$ из окрестности $\omega = 0$ в окрестности опорных частот $\omega = \pm\,\omega_0$. При этом амплитуды всех спектральных составляющих сигнала уменьшаются вдвое. Отметим, что для определения спектра сигнала в области отрицательных частот применяется операция комплексного сопряжения.

Формула (2.221) позволяет по известной спектральной плотности узкополосного сигнала найти спектр его комплексной огибающей, которая, в свою очередь, в полной мере определяет его физическую огибающую и мгновенную частоту.

**Пример 2.11.** Узкополосный сигнал представляет собой радиоимпульс экспоненциальной формы, аналитически записываемый как

$$u(t) = U_m e^{-\alpha t}\sin\omega_0 t.$$

Определить комплексную огибающую $\tilde{U}_u(t)$, спектральную плотность заданного сигнала $\mathbf{S}(\omega)$ и спектральную плотность $\mathbf{Y}_u(\omega)$ его комплексной огибающей.

Р е ш е н и е. Пусть опорная частота $\omega_0$. Поскольку $\sin\omega_0 t = \cos(\omega_0 t - \pi/2)$, то начальная фаза $\varphi_u(t) = -\pi/2$. Используя соотношение (2.211) и формулу Эйлера, получим следующее выражение для комплексной огибающей сигнала

$$\tilde{U}_u(t) = (U_m e^{-\alpha t})e^{-j\pi/2} = U_m e^{-\alpha t}[\cos(\pi/2) - j\sin(\pi/2)] = -jU_m e^{-\alpha t}.$$

С помощью прямого преобразования Фурье находим спектральную плотность комплексной огибающей

$$\mathbf{Y}(\omega) = \int_{-\infty}^{\infty}\tilde{U}_u(t)e^{-j\omega t}\,dt = \int_{-\infty}^{\infty} -jU_m e^{-\alpha t}e^{-j\omega t}\,dt = -\frac{jU_m}{\alpha + j\omega}.$$

Аналогично вычисляем спектральную плотность узкополосного сигнала

$$\mathbf{S}(\omega) = \int_{-\infty}^{\infty}(U_m e^{-\alpha t}\sin\omega_0 t)e^{-j\omega t}\,dt = \frac{U_m\omega_0}{[\alpha + j(\omega - \omega_0)][\alpha + j(\omega + \omega_0)]}.$$

## *Аналитический сигнал и преобразования Гильберта*

Поскольку в радиотехнике и электросвязи в основном используются сигналы с ограниченным спектром (например, модулированные сигналы), для их адекватного математического представления во временно́й области учеными было введено важное понятие «аналитический сигнал» (1951–1983; США — Г.Б. Велкер, Е. Бедросян; СССР — С.И. Тетельбаум, Д.Е. Вакман; широкое использование аналитического сигнала в теории связи началось с 1962 г.), основанное на использовании преобразования Д. Гильберта. Аналитический сигнал — способ комплексного представления физического сигнала, который применяют при анализе сложных колебаний и систем их обработки. Он, кстати, позволил ввести в анализ сигналов понятия огибающей и мгновенной частоты сигнала.

**Аналитический сигнал.** С помощью известной формулы Эйлера представим гармоническое косинусоидальное колебание единичной амплитуды суммой двух комплексно-сопряженных функций

$$\cos\omega t = 0{,}5(\mathrm{e}^{j\omega t} + \mathrm{e}^{-j\omega t}).$$

Возможность такого представления гармонических колебаний позволяет и произвольный физический сигнал $u(t)$ с известной спектральной плотностью $S(\omega)$ записать (через обратное преобразование Фурье) в виде суммы двух составляющих, каждая из которых содержит либо только *положительные*, либо только *отрицательные частоты*

$$u(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S(\omega)\mathrm{e}^{j\omega t}d\omega = \frac{1}{2\pi}\int_{-\infty}^{0} S(\omega)\mathrm{e}^{j\omega t}d\omega + \frac{1}{2\pi}\int_{0}^{\infty} S(\omega)\mathrm{e}^{j\omega t}d\omega. \qquad (2.222)$$

В радиотехнике и теории сигналов функцию

$$z_u(t) = \frac{1}{\pi}\int_{0}^{\infty} S(\omega)\mathrm{e}^{j\omega t}d\omega \qquad (2.223)$$

из представления (2.222) стали называть *аналитическим сигналом*, отвечающим вещественному (физическому) сигналу $u(t)$.

Заметим, что аналитический сигнал $z_u(t)$, описываемый формулой (2.223), есть комплексный сигнал, сформированный из вещественного сигнала $u(t)$.

Теперь проведем некоторые преобразования с первым из интегралов в правой части формулы (2.222). Заменив в нем переменную $\Omega = -\omega$ и проделав несложные выкладки в виде обратного преобразования Фурье, приходим еще к одному интересному равенству:

$$\frac{1}{2\pi}\int_{-\infty}^{0} S(\omega)\mathrm{e}^{j\omega t}d\omega = -\frac{1}{2\pi}\int_{\infty}^{0} S(-\Omega)\mathrm{e}^{-j\Omega t}d\Omega = \frac{1}{2\pi}\int_{0}^{\infty} S(-\Omega)\mathrm{e}^{-j\Omega t}d\Omega = \frac{1}{2\cdot 2\pi}\int_{-\infty}^{\infty} S(-\Omega)\mathrm{e}^{-j\Omega t}d\Omega = 0{,}5z_u^*(t),$$

где $z_u^*(t)$ — сигнал, комплексно-сопряженный с аналитическим сигналом $z_u(t)$.

Нетрудно заметить, что формула (2.222) устанавливает следующую связь между физическим $u(t)$ и аналитическим $z_u(t)$ сигналами:

$$u(t) = 0{,}5[z_u(t) + z_u^*(t)]. \qquad (2.224)$$

**Преобразования Гильберта.** Еще более упростить анализ узкополосных сигналов позволяют известные в математике преобразования Гильберта. Представим произвольный сигнал $u(t)$ как произведение двух функций (2.203):

$$u(t) = U(t)\cos\psi(t) = U(t)\cos[(\omega_0 t + \varphi(t)],$$

т. е. выделим его амплитудную огибающую $U(t)$ и полную фазу $\psi(t)$. Очевидно, что способов сделать это бесконечно много, поскольку одной функции $u(t)$

необходимо поставить в соответствие набор из двух функций $U(t)$ и $\psi(t)$. Однако такое представление должно удовлетворять нескольким ограничениям по огибающей и фазе:

- абсолютное значение (часто, модуль) сигнала $u(t)$ в любой момент времени не превышает значений огибающей $U(t) \geq |u(t)|$;
- касательные, проведенные к кривым $U(t)$ и $u(t)$ в тех точках, где предыдущее неравенство превращается в равенство, совпадают, что означает равенство их производных;
- малым изменениям $u(t)$ соответствуют малые изменения $U(t)$;
- необходимо, чтобы для анализируемого гармонического сигнала *обязательно* выполнялось равенство комплексной (2.209) и физической огибающих $\tilde{U}_u(t) = |\tilde{U}_u(t)| = U$ и полная фаза $\psi(t) = \omega_0 t + \varphi(t)$;
- полная фаза $\psi(t)$ и мгновенная частота не должны зависеть от мощности сигнала, т. е. $\omega(t) = \omega_0$.

Разумно и очевидное требование, что полная фаза не должна меняться при умножении или делении сигнала на произвольный постоянный коэффициент. С учетом этих требований способ выделения амплитудной огибающей и полной фазы из произвольного сигнала оказывается единственным: эта операция производится с помощью *преобразования Гильберта.* Поясним это. Несложные вычисления по формуле (2.224) показывают, что

$$u(t) = \operatorname{Re} z_u(t). \tag{2.225}$$

Мнимую составляющую аналитического сигнала

$$\hat{u}(t) = \operatorname{Im} z_u(t) \tag{2.226}$$

называют *сопряженным по Гильберту сигналом* (*сопряженным сигналом*, иногда *квадратурным дополнением*) по отношению к физическому колебанию $u(t)$.

Физический сигнал $u(t)$ и сопряженный ему сигнал $\hat{u}(t)$ ортогональны:

$$\int_0^T u(t)\hat{u}(t)dt = 0, \tag{2.227}$$

где $T$ — период следования физического сигнала.

Поэтому действительную и мнимую части аналитического сигнала называют *квадратурными составляющими.*

Итак, аналитический сигнал можно представить через физический и сопряженный по Гильберту сигналы в виде суммы

$$z_u(t) = u(t) + j\hat{u}(t). \tag{2.228}$$

Очевидно, что аналитический сигнал на комплексной плоскости отображается вектором, модуль и фазовый угол которого изменяются во времени. Проекция аналитического сигнала на вещественную ось в любой момент времени равна исходному сигналу $u(t)$.

Согласно прямому преобразованию Гильберта, сопряженный сигнал связан с физическим следующим уравнением:

$$\hat{u}(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{u(\tau)}{t-\tau} d\tau. \tag{2.229}$$

Обратное преобразование Гильберта от сопряженного сигнала $\hat{u}(t)$ дает физический сигнал

$$u(t) = -\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\hat{u}(\tau)}{t-\tau} d\tau = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\hat{u}(\tau)}{\tau-t} d\tau. \tag{2.230}$$

Символическая их запись имеет вид:

$$\hat{u}(t) = \mathrm{H}[u(t)]; \quad u(t) = \mathrm{H}^{-1}[\hat{u}(t)]. \tag{2.231}$$

С помощью физического и сопряженного по Гильберту сигналов легко определить огибающую $U_u(t)$, полную фазу $\psi_u(t)$ и мгновенную частоту $\omega_u(t)$ физического сигнала $u(t)$

$$U_u(t) = \sqrt{u^2(t) + \hat{u}^2(t)}; \tag{2.232}$$

$$\psi_u(t) = \operatorname{arctg} \frac{\hat{u}(t)}{u(t)}; \tag{2.233}$$

$$\omega_u(t) = \frac{d\psi_u(t)}{dt}. \tag{2.234}$$

Дифференцирование формулы (2.232) после предварительного возведения в квадрат обеих ее частей

$$U_u(t) \frac{dU_u(t)}{dt} = u(t) \frac{du(t)}{dt} + \hat{u}(t) \frac{d\hat{u}(t)}{dt}$$

в точках соприкосновения, где $U_u(t) = u(t)$ и $\hat{u}(t) = 0$, дает равенство производных

$$\frac{dU_u(t)}{dt} = \frac{du(t)}{dt}, \tag{2.235}$$

т. е. в точках соприкосновения сигнал и его огибающая не только совпадают, но и имеют одинаковые скорости изменения. Все это оправдывает название *огибающей* для функции $U_u(t)$.

Поскольку аналитический сигнал является комплексной функцией, на комплексной плоскости он отображается вектором, вращающимся против часовой стрелки с опорной частотой $\omega_0$, при этом его модуль и фазовый угол изменяются во времени. Проекция аналитического сигнала на вещественную ось комплексной плоскости в любой момент времени равна физическому сигналу $u(t)$.

Смысл термина «аналитический сигнал» заключается в том, что при переходе к переменной $t = \tau + jx$ функция $z_u(t) = z_u(\tau + jx)$, определяемая в соответствии с (2.223) интегралом

$$z_u(t) = \frac{1}{\pi}\int_0^\infty \mathbf{S}(\omega)e^{j\omega\tau}e^{j\omega x}d\omega,$$

является аналитической функцией для всех $x > 0$.

Нетрудно заметить, что прямое преобразование Гильберта (2.229) представляет собой свертку сигнала $u(t)$ и функции $1/(\pi t)$. Это означает, что преобразование Гильберта может быть выполнено линейной системой с постоянными параметрами (такие системы описаны в гл. 4). Из этого следует, что можно легко определить комплексный коэффициент передачи преобразования Гильберта

$$\mathbf{K}(\omega) = \int_{-\infty}^{\infty}\frac{1}{\pi t}e^{-j\omega t}dt = \begin{cases} j, & \omega < 0; \\ 0, & \omega = 0; \\ -j, & \omega > 0. \end{cases} \qquad (2.236)$$

Известно, что модуль комплексного коэффициента передачи представляет амплитудно-частотную характеристику. Из (2.236) следует, что АЧХ преобразования Гильберта равна единице всюду, кроме нулевой частоты, т. е. преобразование Гильберта не меняет амплитудных соотношений в спектре сигнала, лишь удаляя из него постоянную составляющую. Фазы всех спектральных составляющих в области положительных частот уменьшаются на 90° (коэффициент $j$), в области отрицательных частот — увеличиваются на 90° (коэффициент $-j$). Можно показать (см. гл. 4), что результат преобразования Гильберта (2.229) есть реакция описанной линейной системы с импульсной характеристикой $1/(\pi t)$ при подаче на ее вход сигнала $u(t)$ (забегая вперед, заметим, что система, осуществляющая преобразование Гильберта, является *физически нереализуемой*, поскольку ее импульсная характеристика имеет бесконечную протяженность в обоих направлениях временно́й оси).

С формальной точки зрения способ получения сопряженного сигнала с помощью прямого преобразования Гильберта (2.229) можно представить себе следующим образом. Исходное физическое колебание $u(t)$ подается на вход некоторого устройства, которое осуществляет поворот фаз всех его спектральных составляющих на угол –90° в области положительных частот и на угол 90° в области отрицательных частот, не изменяя их при этом по амплитуде. По существу преобразователь Гильберта должен представлять собой идеальное фазосдвигающее устройство (*фазовращатель*), вносящее на всех частотах фазовый сдвиг, равный 90°. Устройство, обладающее подобными свойствами, в радиотехнике принято называть *квадратурным фильтром*.

**Пример 2.12.** Определить сигнал, сопряженный с гармоническим колебанием $u(t) = \cos\omega_0 t$. Записать выражение для аналитического сигнала.

Р е ш е н и е. Результаты можно получить непосредственно из формулы прямого преобразования Гильберта (2.229). Для этого, введя новую переменную $x = \tau - t$ и осуществив несложные преобразования, запишем

$$\hat{u}(t) = -\frac{1}{\pi}\cos\omega_0 t \int_{-\infty}^{\infty}\frac{\cos\omega_0 x}{x}dx + \frac{1}{\pi}\sin\omega_0 t \int_{-\infty}^{\infty}\frac{\sin\omega_0 x}{x}dx.$$

Из курса математики известно, что

$$\int_{-\infty}^{\infty}\frac{\cos\omega_0 x}{x}dx = 0; \qquad \int_{-\infty}^{\infty}\frac{\sin\omega_0 x}{x}dx = \pi.$$

Тогда после подстановок этих значений получаем, что гармоническому колебанию $u(t) = \cos\omega_0 t$ соответствует сопряженный по Гильберту сигнал

$$\hat{u}(t) = \sin\omega_0 t,$$

проходящий через нуль тогда, когда значение физического сигнала максимальное.

Из полученных вычислений легко заметить, что прямое преобразование Гильберта обеспечивает необходимый выбор мнимой части комплексного сигнала для гармонического колебания $u(t) = \cos\omega_0 t$. Операция сдвигает все гармонические составляющие сигнала по фазе на $\pm \pi/2$ и удаляет постоянную составляющую.

Если в соотношение (2.228) подставить физический и сопряженный сигналы, то для гармонического колебания $u(t) = \cos\omega_0 t$ аналитический сигнал:

$$z_a(t) = u(t) + j\,\hat{u}(t) = \cos\omega_0 t + j\sin\omega_0 t.$$

По аналогии с решенным примером нетрудно убедиться, что синусоидальному колебанию $u(t) = \sin\omega_0 t$ соответствует сопряженный сигнал вида

$$\hat{u}(t) = -\cos\omega_0 t.$$

Также можно записать и аналитический сигнал для данного колебания:

$$z_a(t) = u(t) + j\,\hat{u}(t) = \sin\omega_0 t - j\cos\omega_0 t.$$

Если исходный физический сигнал состоит из суммы гармонических колебаний (без постоянной составляющей)

$$u(t) = \sum_{n=1}^{\infty}(a_n \cos n\omega_1 t + b_n \sin n\omega_1 t), \qquad (2.237)$$

то сопряженный сигнал

$$\hat{u}(t) = \sum_{n=1}^{\infty}(a_n \sin n\omega_1 t - b_n \cos n\omega_1 t). \qquad (2.238)$$

Ряд (2.237) называется сопряженным ряду (2.238).

**Пример 2.13.** Задано простое гармоническое колебание единичной амплитуды $u(t) = \cos\omega_0 t$. Определить его огибающую, полную фазу и мгновенную частоту.

Р е ш е н и е. Огибающая исходного сигнала согласно (2.232)

$$U_u(t) = \sqrt{u^2(t) + \hat{u}^2(t)} = 1$$

равна его амплитуде и не зависит от времени.

Полную фазу заданного сигнала определим по (2.233)

$$\psi_u(t) = \operatorname{arctg} \frac{\hat{u}(t)}{u(t)} = \operatorname{arctg} \frac{\sin \omega_0 t}{\cos \omega_0 t} = \omega_0 t,$$

а мгновенную частоту — по (2.234):

$$\omega_u(t) = \frac{d\psi_u(t)}{dt} = \frac{d(\omega_0 t)}{dt} = \omega_0.$$

Из примера следует, что нахождение физической огибающей, полной фазы и мгновенной частоты гармонического сигнала с помощью преобразования Гильберта приводит к результатам, согласующимся с обычными представлениями о свойствах гармонических колебаний.

Ввод аналитического и сопряженного сигналов не позволяет получить каких-либо новых сведений о физическом сигнале $u(t)$. Однако они значительно расширяют систематические методы исследования узкополосных процессов.

**Спектральная плотность аналитического сигнала и комплексной огибающей.** Сравним амплитудные спектры аналитического сигнала и комплексной огибающей (рис. 2.62).

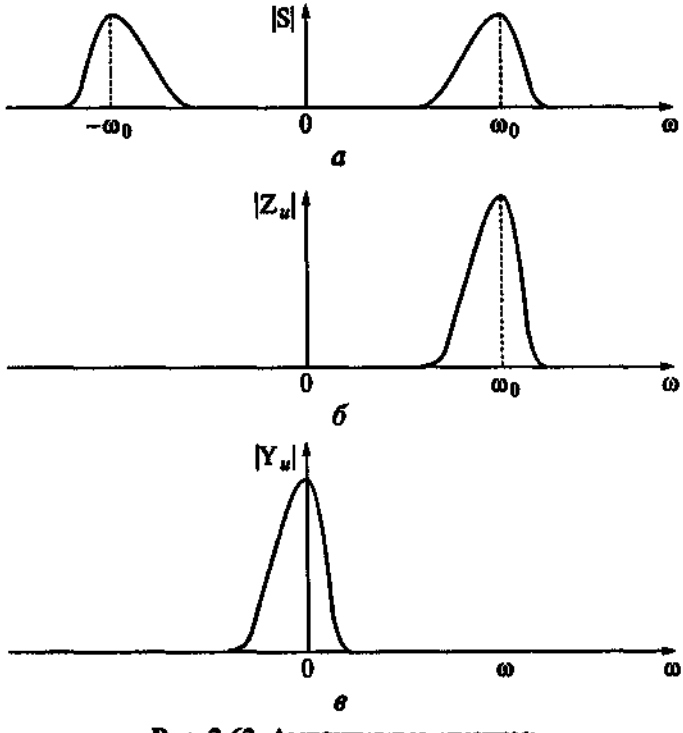



**Рис. 2.62.** Амплитудные спектры:

*а* — вещественного сигнала; *б* — аналитического сигнала; *в* — комплексной огибающей

Положим, что $\mathbf{S}(\omega)$ — спектральная плотность физического сигнала $u(t)$ (рис. 2.62, *а*). Введем функцию $\mathbf{Z}_u(\omega)$, связанную с аналитическим сигналом $z_u(t)$ прямым преобразованием Фурье

$$\mathbf{Z}_u(\omega) = \int_{-\infty}^{\infty} z_u(t) \mathrm{e}^{-j\omega t} dt \tag{2.239}$$

и являющуюся *спектральной плотностью аналитического сигнала.*

Определим параметры введенной спектральной плотности аналитического сигнала. Как отмечалось ранее, аналитический сигнал получают путем добавления к вещественному сигналу $u(t)$ сопряженной (мнимой) его части в виде его преобразования Гильберта:

$$z_u(t) = u(t) + j\hat{u}(t).$$

Обозначим через $\hat{\mathbf{S}}(\omega)$ спектральную плотность сопряженного сигнала $\hat{u}(t)$. Тогда в силу линейности прямого преобразования Фурье спектральная плотность аналитического сигнала запишется как сумма

$$\mathbf{Z}_u(\omega) = \mathbf{S}(\omega) + j\hat{\mathbf{S}}(\omega). \tag{2.240}$$

Учитывая, что преобразование Гильберта является линейным и его коэффициент передачи определяется формулой (2.236), находим

$$\mathbf{Z}_u(\omega) = \begin{cases} 2\mathbf{S}(\omega), & \omega \geq 0; \\ 0, & \omega < 0. \end{cases} \tag{2.241}$$

Равенство (2.241) выполняется тогда, когда спектральные плотности исходного и сопряженного сигналов связаны между собой следующим образом:

$$\hat{\mathbf{S}}(\omega) = \begin{cases} j\mathbf{S}(\omega), & \omega \geq 0; \\ -j\mathbf{S}(\omega), & \omega < 0. \end{cases} \tag{2.242}$$

Полученный результат (2.241) весьма интересен. В области положительных частот спектры вещественного сигнала и добавленной мнимой части (с учетом дополнительного фазового сдвига в 90°, вносимого множителем $j$) складываются, давая удвоенный результат. В области же отрицательных частот эти спектры оказываются противофазными и взаимно уничтожаются. Таким образом, *спектральная плотность аналитического сигнала* равна удвоенному значению спектральной плотности физического сигнала и находится в области *только положительных частот*, т. е. оказывается *односторонней* (рис. 2.62, *б*).

Найдем спектральную плотность комплексной огибающей $\mathbf{Y}_u(\omega)$ физического сигнала $u(t)$. Для этого используем связь между аналическим сигналом $z_u(t)$ и комплексной огибающей $\tilde{U}_u(t)$

$$z_u(t) = \tilde{U}_u(t) \mathrm{e}^{j\omega_0 t}.$$

Очевидно, что

$$\tilde{U}_u(t) = z_u(t)\,e^{-j\omega_0 t}.$$

Тогда

$$\mathbf{Y}_u(\omega) = \int_{-\infty}^{\infty} \tilde{U}_u(t)e^{-j\omega t}dt = \int_{-\infty}^{\infty} z_u(t)e^{-j\omega_0 t}e^{-j\omega t}dt = \int_{-\infty}^{\infty} z_u(t)e^{-j(\omega+\omega_0)t}dt = \mathbf{Z}_u(\omega+\omega_0). \quad (2.43)$$

Итак, спектральная плотность комплексной огибающей представляет сдвинутую на $\omega_0$ спектральную плотность аналитического сигнала. Спектр комплексной огибающей сигнала не обязательно симметричен относительно нулевой частоты. Следует помнить, что понятие комплексной огибающей имеет смысл только при указании выбранной центральной частоты $\omega_0$, относительно которой эта комплексная огибающая вычислена.

Представление узкополосных сигналов их комплексными огибающими для анализа прохождения сигналов через различные цепи называют *методом низкочастотного эквивалента.* При этом важно, что любую комплексную огибающую следует вычислять относительно одной и той же центральной частоты $\omega_0$.

**Пример 2.14.** Задан низкочастотный физический сигнал $u(t)$ с равномерной спектральной плотностью $S_0$ в полосе частот $-\omega_в \le \omega \le \omega_в$ (рис. 2.63, *а*). Определить физический и сопряженный с ним сигналы.

Р е ш е н и е. Согласно (2.223) аналитический сигнал

$$z_u(t) = \frac{S_0}{\pi}\int_0^{\omega_в} e^{j\omega t}d\omega = \frac{S_0}{j\pi t}(e^{j\omega_в t} - 1).$$

Выделив вещественную и мнимую составляющие аналитического сигнала, находим вещественный сигнал

$$u(t) = u(\omega t) = \frac{S_0\omega_в}{\pi}\frac{\sin\omega_в t}{\omega_в t} \quad (2.244)$$

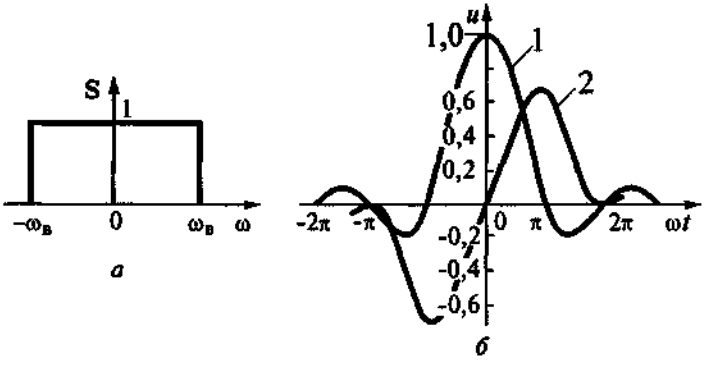



**Рис. 2.63.** К примеру 2.14:
*а* — спектр физического сигнала; *б* — физический 1 и сопряженный 2 сигналы

и сопряженный ему

$$\hat{u}(t) = \hat{u}(\omega t) = \frac{S_0\omega_в}{\pi}\frac{\sin^2(\omega_в t/2)}{\omega_в t/2}. \qquad (2.245)$$

Временные диаграммы физического (1) и сопряженного (2) сигналов приведены на рис. 2.63, *б* для случая $\frac{S_0\omega_в}{\pi} = 1$.

Результаты примера 2.14 отражают одно из основных свойств преобразования Гильберта: если в какой-то момент времени $t$ исходный сигнал $u(t)$ достигнет экстремума, то в окрестности этой точки сопряженный ему сигнал пройдет через нуль.

Одно из простейших и в тоже время очень важных свойств преобразований Гильберта — линейность:

$$H[a_1u_1(t) + a_2u_2(t)] = a_1H[u_1(t)] + a_2H[u_2(t)] \qquad (2.246)$$

при любых постоянных коэффициентах $a_1$ и $a_2$, в чем можно убедиться непосредственно решением.

**Преобразование Гильберта для узкополосного сигнала.** Пусть известна спектральная плотность $\mathbf{Y}_u(\omega)$ комплексной огибающей $\tilde{U}_u(t)$ узкополосного сигнала $u(t)$, имеющего опорную (центральную) частоту $\omega_0$. Согласно соотношению (2.242), спектр этого сигнала

$$\mathbf{S}(\omega) = 0{,}5\,\mathbf{Y}_u(\omega - \omega_0) + 0{,}5\,\mathbf{Y}_u^*(-\omega - \omega_0).$$

Нетрудно заметить, что первое слагаемое в правой части уравнения соответствует области частот $\omega > 0$, второе $-\omega < 0$. Тогда на основании формулы (2.242) спектральная плотность сопряженного сигнала

$$\hat{\mathbf{S}}(\omega) = 0{,}5\mathbf{Y}_u(\omega-\omega_0)e^{j\pi/2} + 0{,}5\mathbf{Y}_u^*(-\omega-\omega_0)e^{j\pi/2}. \qquad (2.247)$$

Как следует из последней формулы, спектральная плотность комплексной огибающей сопряженного сигнала

$$\mathbf{Y}_{\hat{u}}(\omega) = \mathbf{Y}_u(\omega)e^{-j\pi/2}. \qquad (2.248)$$

Следовательно, узкополосному сигналу соответствует также узкополосный сопряженный сигнал.

В последние годы методы, связанные с понятием аналитического сигнала и преобразованиями Гильберта, имеют широкое распространение в радиоэлектронике и теории связи.

## Контрольные вопросы

1. По каким признакам классифицируют радиотехнические сигналы?
2. Какие сигналы относятся к аналоговым, дискретным и цифровым?
3. В чем заключено основное отличие детерминированных и случайных сигналов?
4. Что представляют собой шумы и помехи?
5. Какие сигналы являются ортогональными и ортонормированными?
6. Чем обусловлено широкое применение гармонических сигналов в радиотехнике?
7. Для каких целей применяют спектральное представление сигналов?
8. Какие формы ряда Фурье применяются в теории сигналов?
9. Как зависит спектр периодической последовательности импульсов от периода?
10. Что такое спектральная плотность сигнала?
11. Почему для спектрального представления непериодических сигналов используется прямое преобразование Фурье? Как связаны между собой преобразования Фурье?
12. Как выглядит спектр прямоугольного импульса?
13. Как изменяется спектральная плотность импульса при изменении его длительности?
14. Как связаны спектры импульса и последовательности таких же импульсов?
15. Какими основными свойствами обладают преобразования Фурье?
16. Что такое дельта-функция и функция включения?
17. Какова связь между функцией Хевисайда и дельта-функцией?
18. Что такое АКФ детерминированного сигнала? Какими свойствами обладает АКФ?
19. Какой вид имеет АКФ прямоугольного импульса?
20. Что характеризует энергетический спектр импульсного сигнала?
21. Как связаны между собой АКФ и энергетический спектр сигнала?
22. Для каких целей применяют ВКФ двух детерминированных сигналов?
23. В чем заключается физический процесс модуляции несущего колебания?
24. Покажите спектры АМ-колебания при модуляции одним тоном и сложным сигналом.
25. В чем состоит принцип построения векторной диаграммы АМ-сигнала?
26. Для каких целей применяются балансная и однополосная модуляции?
27. Какова связь между спектральными плотностями радиоимпульса и его огибающей?
28. В чем заключаются различия и сходство однотональных ЧМ- и ФМ-сигналов?
29. Как определяются и как связаны между собой частота модуляции, девиация частоты и индекс модуляции в однотональных ЧМ- и ФМ-сигналах?
30. Чем отличаются спектры АМ- и ЧМ-сигналов при малых индексах модуляции?
31. Чему равна практическая ширина спектра сигналов с угловой модуляцией?
32. Для каких целей применяется полярная модуляция?
33. На каком физическом принципе основано сжатие ЛЧМ-импульса во времени?
34. Какие существуют виды импульсной модуляции?
35. В чем отличие спектров последовательности импульсов и АИМ-сигнала?
36. В чем заключается принцип импульсно-кодовой модуляции?
37. Как осуществляют цифровые фазовую и частотную модуляции?
38. Дайте пояснение огибающей, фазе и мгновенной частоте узкополосного сигнала.
39. Какими свойствами обладает физическая огибающая узкополосного сигнала?
40. С какой целью вводится понятие «аналитический сигнал»?
41. Дайте определение аналитическому и сопряженному сигналам.
42. Запишите прямое и обратное преобразования Гильберта.

# Глава 3. СЛУЧАЙНЫЕ И ШУМОПОДОБНЫЕ СИГНАЛЫ

В последние годы широкое развитие получила *статистическая радиотехника*, что объясняется необходимостью изучения явлений при передаче сообщений в условиях, когда детерминированное описание сигналов невозможно. Статистическая радиотехника рассматривает вопросы приема сигналов в шумах и базируется на аппарате теории вероятностей и случайных процессов. Случайное изменение параметров полезных сигналов обусловлено воздействием на них дестабилизирующих факторов, таких, как случайный характер модуляции, нестабильность частоты и т. д. Такой же вид имеют и шумы. Поэтому они объединены и терминологически — понятия «случайный сигнал» и «случайный процесс» используются как синонимы и часто обозначаются как «случайные процессы». При статистическом подходе к этой проблеме не обязательно определять точный результат одного исследования, а можно основываться на анализе множества опытов и найти закономерности, характеризующие случайный процесс в среднем.

Начало применения статистических методов к проблемам приема сигналов в присутствии шумов положил С. Райс. Значительный вклад в развитие статистической радиотехники внесли наши соотечественники В.А. Котельников, В.И. Тихонов, Ю.Б. Кобзарев, В.И. Бунимович; американские ученые Д. Миддлтон, Ф. Вудворд, А. Зигерт, Х. Хелстрем, Р. Фано, Р. Хемминг.

Важное значение в радиотехнике, наряду с узкополосными, имеют широкополосные сигналы. К ним относятся и *шумоподобные* (*pseudonoise*) *сигналы*, называемые так по причине близости их спектра к белому шуму.

## 3.1. Общие сведения о случайных сигналах

Как отмечалось в главе 2, отличительной чертой случайного сигнала является то, что его мгновенные значения заранее не предсказуемы. Случайный сигнал представляет случайную функцию времени, и его рассмотрение во временно́й области, как и детерминированного сигнала, является наиболее наглядным. Вместе с тем, если детерминированный сигнал однозначно описывается временно́й функцией, то такое представление случайного процесса невозможно.

Итак, в отличие от детерминированных сигналов, значения случайных сигналов в произвольные моменты времени не могут быть вычислены достоверно, а могут быть только предсказаны в определенном диапазоне значений с определенной вероятностью, меньшей единицы. Однако, изучая такой случайный сиг-

нал детально, можно заметить, что ряд его характеристик весьма точно описывается в вероятностном смысле. Важно и то, что чаще всего наблюдают относительно небольшие отклонения амплитудных значений случайного сигнала от некоторого среднего уровня; чем больше отклонения по абсолютному значению, тем реже их наблюдают. Уже в этом проявляется статистическая закономерность. Располагая сведениями о вероятностях флуктуации различного уровня, удается создать математическую модель случайного колебания, приемлемую для детального анализа случайного процесса.

Случайный процесс в каждый момент времени представляет случайную величину, которая может принимать различные значения из области возможных для нее. С общих вероятностных позиций под случайной величиной понимают результат фиксации такого явления (процесса), который является либо полностью непредсказуемым, либо предсказуем только «в среднем», с определенной *средней квадратической* (упрощенно — *среднеквадратической*) *погрешностью* (СКП; *experimental standard deviation*). В качестве универсальной координаты распределения случайных величин по независимой переменной часто используют переменную $t$ и трактуют ее для удобства как временну́ю координату. Таким образом, во временно́й области случайный процесс может быть описан только с использованием вероятностных характеристик (теория вероятности — раздел математики, в котором изучаются закономерности случайных событий, явлений и процессов независимо от их физической природы). Вероятностные законы возникают всегда, если физическая система, порождающая случайный процесс, представляет собой объединение очень большого числа более мелких подсистем (устройств, или цепей), формирующих некоторые индивидуальные сигналы, в большей или меньшей степени не зависимые друг от друга.

Важно то, что теория вероятности рассматривает случайные величины в «статике». Случайные сигналы, напротив, «динамические» явления, развивающиеся во времени или по любой другой переменной. Задачи описания и изучения случайных сигналов «в динамике», как отображений случайных явлений, рассматривает теория случайных процессов. Количественные характеристики случайных сигналов, позволяющие производить их оценку и сравнение, называют *статистическими*. Специфика рассмотрения случайных процессов в рамках теоретической радиотехники и теории связи заключается, прежде всего, в выборе их характеристик и очерченного круга задач анализа.

В общем случае в радиотехнике и теории связи существуют два основных класса сигналов, нуждающихся в вероятностном описании. Во-первых, практически случайными являются все реальные сигналы, несущие *информацию*, поэтому для описания закономерностей, присущих осмысленным сообщениям, прибегают к вероятностным моделям. Во-вторых, *шумы* — хаотически изменяющиеся во времени электромагнитные колебания, возникающие в разных физических системах из-за беспорядочного движения носителей заряда.

По структуре и характеру радиотехнические случайные процессы весьма разнообразны, поэтому потребность в изучении их свойств привела к развитию ряда специфических методов их анализа.

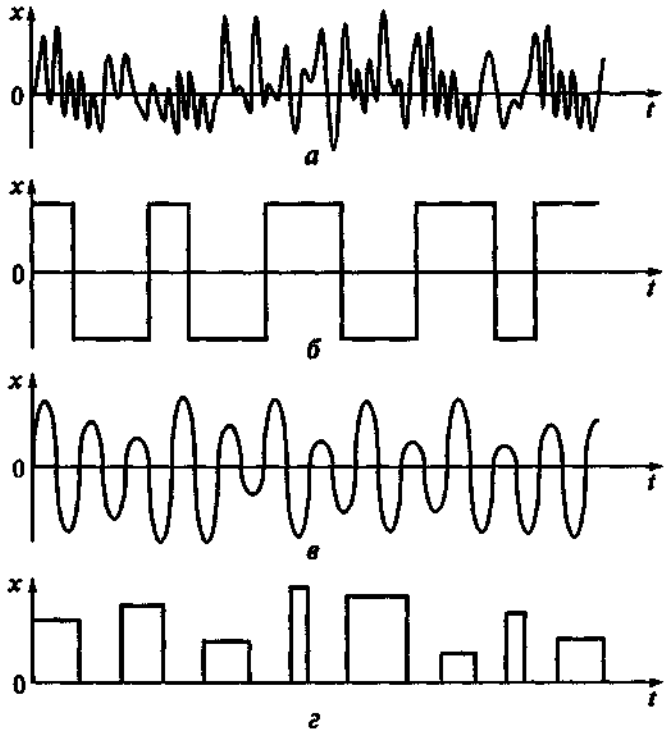



**Рис. 3.1.** Различные виды случайных процессов:
*а* — реализация непрерывного шума; *б* — «телеграфный» сигнал;
*в* — случайный процесс на входе детектора; *г* — импульсный случайный сигнал

Случайные процессы бывают непрерывными и дискретными (рис. 3.1). Примером непрерывного случайного процесса служит шум в радиотехнической цепи. Непрерывный шум (одна его реализация — временна́я диаграмма) может иметь любое значение из области возможных (рис. 3.1, *а*).

Дискретный случайный процесс принимает только определенные значения. Как сложилось на практике, дискретными будем также называть случайные процессы, получавшиеся при дискретизации непрерывных случайных процессов во времени. Такие случайные процессы представляют последовательность импульсов, амплитуды которых соответствуют мгновенным значениям исходного непрерывного случайного процесса. Примером такого процесса является колебание на выходе идеального ограничителя при подаче на его вход непрерывного случайного процесса — аналог «телеграфного сигнала» (рис. 3.1, *б*). Возможен случайный процесс смешанного типа, например случайный процесс на входе детектора приемника (рис. 3.1, *в*). Отдельно следует выделить импульсные случайные процессы (рис. 3.1, *г*).

Из всего возможного разнообразия случайных процессов выделим только те, которые наиболее часто встречаются в инженерной и исследовательской практике, и проведем их классификацию по основным признакам. Такая классификация позволяет определить место каждого вида случайного процесса сре-

ди возможного их разнообразия и ввести устоявшуюся в радиотехнике терминологию, которая будет использована в дальнейшем. Подчеркнем и такой редко упоминаемый момент — случайный вид полезных информационных сигналов можно рассматривать как стохастический (от греч. *stochastikos* — умеющий угадывать) процесс, течение которого зависит от случая и для которого определена вероятность того или иного его течения. Очевидно, что стохастические случайные процессы более предсказуемы, чем просто случайные. Например, можно заранее знать или предугадать несущую частоту или диапазон частот передаваемого сигнала.

### Вероятностное представление случайных величин и процессов

Известно, что при решении многих практических задач радиотехники и техники электросвязи случайный процесс, помимо вероятностного описания, может быть представлен рядом неслучайных числовых характеристик, постоянных или изменяющихся во времени.

В теории сигналов исследуемый случайный процесс представляют бесконечным множеством некоторых временны́х функций. Рассмотрим (для наглядности низкочастотный) случайный процесс, состоящий из множества случайных сигналов $x_1(t)$, $x_2(t)$, ..., $x_k(t)$, ... , называемых *реализациями случайного процесса* (рис. 3.2), и аналитически описываемый некоторой обобщающей его случайной функцией $X(t)$. Совокупность всех реализаций случайного процесса называют *ансамблем* (рис. 3.2, *а*). Ансамбль реализаций — математическая абстракция, модель рассматриваемого процесса, но конкретные реализации (см., например, $k$-ю реализацию на рис. 3.2, *б*), используемые на практике, представляют физические объекты или явления и входят в ансамбль как его неотъемлемая часть.

Типичными примерами случайных процессов в радиотехнике являются тепловые шумы в пассивных и активных элементах, действием которых сопровождается работа всех радиотехнических устройств. Реализацией случайного процесса является зафиксированный осциллографом, самописцем или другим регистрирующим устройством отрезок развития во времени случайного напряжения. Ансамблем реализаций случайного процесса является группа сигналов, наблюдаемых одновременно на выходах идентичных генераторов шумового напряжения. Отдельную реализацию случайного процесса трудно описать явной математической формулой. Однако поскольку конкретный вид реализации (например, принятый сигнал в системе связи) позволяет определить практически все параметры функции $x_k(t)$ (измерить амплитуду, оценить частоту изменений мгновенных значений и т. д.), то она является уже *детерминированным сигналом*.

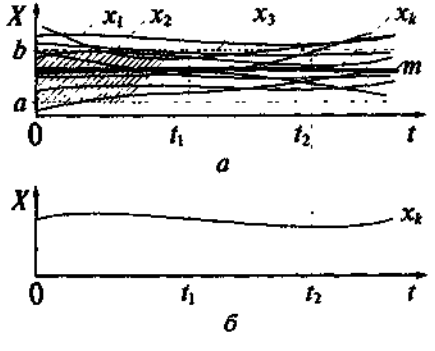



**Рис. 3.2.** Реализации случайного процесса:
*а* — ансамбль; *б* — $k$-я реализация

Отметим на рис. 3.2 некоторый момент времени $t = t_1$. Значения, которые могут принимать конкретные реализации $x_1(t)$, $x_2(t)$, ..., $x_k(t)$ всего ансамбля в заданный момент времени $t_1$, образуют совокупность случайных величин $x_1(t_1)$, $x_2(t_1)$, ..., $x_k(t_1)$, которую обозначим случайной величиной $X(t_1)$. Эту величину $X(t_1)$, отражающую совокупность всех возможных значений реализаций случайного процесса $X(t_1)$ в момент времени $t_1$, называют *сечением* случайного процесса. Случайная величина $X(t_1)$ может иметь любые заранее неизвестные значения в некотором интервале изменения амплитуд.

Аналитическое описание случайных процессов определяется основными статистическими характеристиками, к которым относятся одномерные, двумерные и многомерные функции распределения (интегральные и дифференциальные), числовые характеристики (среднее значение, дисперсия и др.), спектральные и корреляционные функции. Одномерные и двумерные функции распределения позволяют найти наиболее вероятные значения параметров случайных процессов, а многомерные функции распределения дают возможность получить практически полную информацию о случайном процессе, его характеристиках и основных параметрах.

Одной из важных одномерных характеристик случайной величины $X(t_1)$ является *интегральная функция распределения* (проще, *функция распределения*) или *функция распределения вероятности* (*cumulative distribution function*, *CDF*) $F(x)$. Численно эта функция определяется как вероятность того, что все значения случайной величины $X(t_1)$ не превышают некоторого заданного уровня $x$:

$$F(x) = P[X(t_1) < x], \tag{3.1}$$

где $P$ — символ, отражающий вероятность.

Основные свойства интегральной функции распределения вероятности:

- для случайной величины $X(t_1)$, имеющей любые вещественные значения, функцию распределения определяют на интервале $0 \le F(x) \le 1$ при $-\infty < x < \infty$;
- функция распределения $F(x)$ не уменьшается при возрастании аргумента $x$;
- для интегральной функции распределения $F(x)$ справедливо равенство

$$P(x_1 < X \le x_2) = P(x_2) - P(x_1).$$

Если случайная величина $X(t_1)$ является непрерывной во времени, то зачастую вместо функции распределения удобнее пользоваться ее производной

$$p(x, t_1) = \frac{dF(x)}{dx} = \frac{dF}{dx}, \tag{3.2}$$

получившей название *одномерной плотности распределения вероятности* (или, проще, *плотности вероятности — probability density function*, *PDF*).

Зададим некоторый интервал $a - b$ изменения мгновенного значения $x$ случайного процесса (см. рис. 3.2). Тогда из (3.2) следует, что плотность вероятности

$$p(x, t_1)dx = F(b) - F(a) = P[a < X(t_1) < b] \tag{3.3}$$

— есть вероятность попадания случайной величины $X(t_1)$ в заданный интервал.

Пусть параметр $a \to -\infty$, а $b$ принимает текущее значение переменной $x$. В этом случае интегральная функция распределения примет вид:

$$F(x) = P[-\infty < X(t_1) \le x] = \int_{-\infty}^{x} p(x, t_1)dx. \qquad (3.4)$$

Следует отметить, что одномерная плотность вероятности всегда неотрицательная величина и удовлетворяет условию нормировки

$$\int_{-\infty}^{\infty} p(x, t_1)dx = 1. \qquad (3.5)$$

Это соотношение можно трактовать следующим образом: площадь под кривой плотности вероятности $p(x, t_1)$ всегда равна единице.

На рис. 3.3 в качестве наглядного примера приведены графики одной из распространенной функции плотности вероятностей $p(x) = \exp(-|x|)$ и соответствующей интегральной функции распределения вероятностей $F(x)$.

Одномерная плотность вероятности (или функция распределения) и связанные с ней различные характеристики позволяют получить весьма важную информацию о свойствах случайного процесса. Для решения же многих практических задач радиотехники и техники связи таких сведений зачастую недостаточно, так как они дают вероятностное представление о случайном процессе $X(t)$ только в отдельные моменты времени и ничего не говорят о том, как он изменяется в широких интервалах времени. Поэтому для описания его временны́х характеристик необходимо использовать корреляционную функцию или привлечь для этого спектральные характеристики случайного процесса.

Достаточно исчерпывающей характеристикой случайного процесса служит $n$-мерная плотность вероятности $p(x_1,..., x_k; t_1,..., t_n)$, также $n$-мерная функция распределения, полученная для $k$ реализаций в $n$ различных фиксированных моментах времени $t_1, t_2, ..., t_n$. *Многомерные плотности вероятности* используют на практике редко, поскольку они сложны и требуют для определения обработки много экспериментальных данных. В прикладных задачах статистической радиотехники, наряду с одномерной, применяют *двумерную плотность вероятности* $p[x(t_1), x(t_2)]$. Для ее определения надо располагать двумя сечениями случайного процесса $X(t_1)$, $X(t_2)$, полученными в разные моменты времени $t_1$ и $t_2$ (см. рис. 3.2).

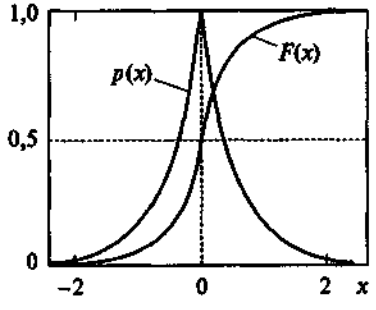



**Рис. 3.3.** Графики функций $p(x) = \exp(-|x|)$ и $F(x)$

**Статистическое усреднение. Числовые характеристики (моменты) случайных процессов.** При решении многих задач радиотехники нет особой необходимости в полной вероятностной характеристике случайных величин,

которую дают функция плотности распределения вероятности или интегральная функция распределения. При этом часто приходится также иметь дело с анализом случайных величин, плотности вероятности которых не отражаются аналитическими функциями либо вообще неизвестны. Вместе с тем задание одномерной плотности вероятности $p(x, t_1)$ позволяет произвести статистическое усреднение и самой величины $x$, и любой функции $f(x)$, а также основных числовых характеристик распределений. Под *статистическим усреднением* (*ensemble averaging*) подразумевают усреднение по ансамблю реализаций в каком-либо сечении процесса, т. е. в фиксированный момент времени. Результатами экспериментов над случайными величинами являются их средние значения.

Если $\varphi(x)$ — известная функция от $x$ (исхода случайного испытания), то по определению ее *среднее значение* (*average value*; часто *mean value*)

$$\overline{\varphi(x)} = \int_{-\infty}^{\infty} \varphi(x) p(x,t) dx. \tag{3.6}$$

Черта сверху над функцией в формуле (3.6) означает операцию *усреднения по множеству результатов* случайных испытаний (реализаций). Отметим, что наибольший вклад в среднее значение дают те величины на оси $x$ случайного процесса, где одновременно велики как усредняемая функция $\varphi(x)$, так и плотность вероятности $p(x)$.

Итак, в статистической радиотехнике, наряду с вероятностными зависимостями $F(x)$ и $p(x)$ для описания свойств случайных процессов, используется ряд неслучайных числовых характеристик, называемых *моментами распределения* (*моментами*) случайной величины. Моменты распределения случайных величин позволяют оценить случайные процессы устойчивыми и уже неслучайными интегральными оценками путем усреднения их по ансамблю реализаций.

Момент $n$-го порядка случайной величины $X(t)$ есть среднее значение $n$-й степени случайной переменной

$$m_n(t) = \overline{x^n(t)} = \int_{-\infty}^{\infty} x^n p(x,t) dx. \tag{3.7}$$

Наиболее простые и часто используемые в статистической радиотехнике и теории электросвязи характеристики закона распределения случайного процесса — математическое ожидание (момент первого порядка) и дисперсия (центральный момент второго порядка).

Запишем известные соотношения для вычисления основных числовых характеристик случайных процессов. В многих практических задачах радиотехники приходится вычислять математическое ожидание функции $f$ от случайной величины $x$, имеющей плотность вероятности $p_x(x)$, по следующей обобщенной формуле

$$m\{f(x)\} = \int_{-\infty}^{\infty} f(x) p_x(x,t) dx. \tag{3.8}$$

Поскольку принято $f(x) = x$, то первый момент, или *математическое ожидание* (*mathematical expectation*),

$$m_x(t) = \int_{-\infty}^{\infty} xp(x,t)dx \tag{3.9}$$

представляет собой среднее значение случайного процесса в текущий момент времени $t$, полученное усреднением по всему ансамблю. Математическое ожидание (линия $m$ на рис. 3.2, *а*) служит теоретической оценкой среднего значения случайной величины, получаемой в больших сериях испытаний.

Средний квадрат (второй момент) случайного процесса вычисляют по формуле

$$\overline{x^2(t)} = \int_{-\infty}^{\infty} x^2 p(x,t)dx. \tag{3.10}$$

При практическом исследовании реальных случайных процессов давно замечено, что чаще всего имеют место относительно небольшие отклонения мгновенных значений реализаций от среднего уровня; чем больше отклонения по абсолютному значению, тем реже они наблюдаются. Уже в этом проявляется некоторая статистическая закономерность случайных процессов. Располагая сведениями о вероятностях флуктуаций различного уровня, удается создать математическую модель случайного колебания, вполне приемлемую как в научном, так и в прикладном смысле. В большинстве радиоэлектронных систем и устройств аналогового принципа действия обычно имеют место случайные процессы с *нулевым математическим ожиданием*, когда реализации группируются по обе стороны от оси времени. Такие процессы называют *центрированными*.

*Дисперсия* (*variance*), или *второй центральный момент*, определяется как математическое ожидание квадрата отклонения случайного процесса от его математического ожидания

$$D_x(t) = \sigma_x^2(t) = \int_{-\infty}^{\infty} [x(t) - m_x(t)]^2 p(x,t)dx. \tag{3.11}$$

При нулевом математическом ожидании дисперсия характеризует среднюю мощность *флуктуаций* случайного процесса. Итак, дисперсия равна разности среднего квадратического значения и квадрата среднего значения случайного процесса. Дисперсия случайной величины неотрицательна. Если случайная величина равна константе, то ее дисперсия равна нулю. Дисперсию $D_x(t)$ случайной величины часто обозначают через квадрат *среднего квадратического отклонения* $\sigma_x^2 = D_x$. Удобство использования дисперсии связано с рядом причин. Прежде всего, отклонение в большую или меньшую сторону от среднего рáвно измеряют разброс значений наблюдаемой величины, так что усредняться должны не сами отклонения, а симметричная функция от них. Кроме квадрата можно взять абсолютную величину, четвертую степень, синус и т. п. Выбор квадра-

тичной функции упрощает модель, при этом среднее квадратичное отклонение естественно входит как параметр в закон нормального распределения, который распространен в радиотехнике. Существуют и иные оценки «параметра положения», кроме среднего, и «параметра масштаба», помимо дисперсии.

Среди двумерных функций распределения особое место занимает второй смешанный центральный момент — *функция корреляции* (корреляционная функция), которая характеризует статистическую связь между значениями одного и того же случайного процесса в два различных момента времени. Наличие корреляционной функции позволяет использовать преобразование Фурье, связывает ее со *спектральной плотностью мощности случайного процесса* (см. далее). Соотношение между корреляционной функцией и спектральной плотностью мощности лежит в основе корреляционного и спектрального анализа случайных процессов.

*Функция корреляции*

$$R_x(t_1,t_2)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[x(t_1)-m_x(t_1)\right]\left[x(t_2)-m_x(t_2)\right]p(x_1,x_2;t_1,t_2)dx_1dx_2 \qquad (3.12)$$

представляет собой меру связи между сечениями случайного процесса, взятыми в моменты времени $t_1$ и $t_2$.

Когда $t_1 = t_2$, т. е. при совмещении сечений случайного процесса, функция корреляции численно равна дисперсии

$$R_x(t,t)=R_x(0)=\sigma_x^2(0). \qquad (3.13)$$

На рис. 3.4 приведены графики некоррелированного и коррелированного случайных процессов, которые характеризуются практически одинаковыми математическими ожиданиями $m$.

Из представленных графиков следует, что хотя пространство состояний некоррелированного (рис. 3.4, *а*) и коррелированного (рис. 3.4, *б*) процессов практически одно и то же, динамика развития их реализаций существенно различается. Отдельные реализации коррелированных процессов в произвольный момент времени могут быть такими же случайными, как и некоррелированных, а в пределе, во всех сечениях оба процесса могут иметь один и тот же закон распределения и одинаковые математические ожидания. Однако динамика развития по временно́й координате $t$ единичной реализации коррелированного процесса по сравнению с некоррелированным является более плавной, а следовательно, в коррелированном процессе имеется определенная связь между последовательными значениями случайных величин.

Случайные процессы, изучаемые в радиотехнике, часто обладают следующим специфическим свойством: их функция корреляции стремится к нулю с увеличением временно́го сдвига $\tau$ (кстати, это является достаточным условием эргодичности процесса). Чем быстрее убывает функция $R(\tau)$, тем слабее оказывается статистическая связь между мгновенными значениями случайного сигнала в два несовпадающих момента времени.

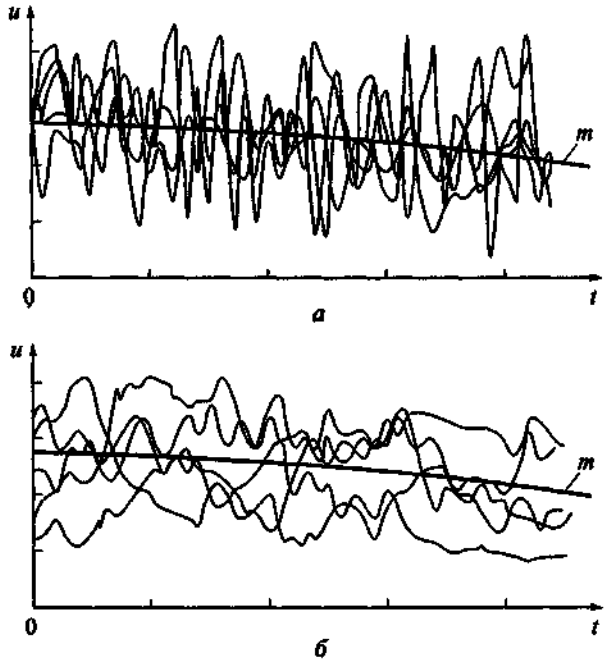



**Рис. 3.4.** Виды случайных процессов:
*а* — некоррелированный; *б* — коррелированный

Наиболее простой и удобной числовой характеристикой быстротечности случайного процесса («скорости изменения» реализаций) является *интервал корреляции*, который определяется как

$$\tau_к = \frac{1}{R(0)} \int_0^\infty |R(\tau)| d\tau. \qquad (3.14)$$

Если известна информация о поведении реализации случайного процесса «в прошлом», то возможен вероятностный его прогноз на интервал $\tau_к$.

## 3.2. Законы распределения случайных процессов

Для задания случайной величины нужно знать ее возможные значения и вероятности, с которыми принимаются эти значения. Соответствие между ними называется *законом распределения* случайной величины. Закон распределения дает практически полную информацию о свойствах и статистических параметрах случайного процесса и позволяет ответить на поставленные вопросы о его поведении во времени. Он может иметь вид таблицы, формулы или графика.

В радиотехнике весьма часто встречается гауссовский процесс — случайный процесс с нормальным распределением мгновенных значений. При анализе случайных процессов, кроме гауссовского (нормального), применяют и



А.М. Ляпунов

равномерный закон. Некоторые виды распределений получаются при преобразовании нормального закона. Так можно получить экспоненциальный (Лапласа второго рода), похожий по форме на нормальный, но с более разбросанными «хвостами» (т. е. вероятность бóльших отклонений для него значительнее, чем для нормального), и рэлеевский законы распределения, и другие, в том числе несимметричные законы. Этим законам чаще подчиняются случайные процессы, имеющие место в большинстве практических случаев.

**Нормальный закон распределения плотности вероятности**

Центральная предельная теорема теории вероятности определяет условия, при которых реальный случайный процесс приближается к нормальному. Одна из трактовок *центральной предельной теоремы Ляпунова* (Александр Михайлович Ляпунов — выдающийся русский математик, один из основоположников теории случайных процессов; 1857–1918) гласит: «Если случайная величина $X$ представляет собой сумму очень большого числа взаимно независимых случайных величин, влияние каждой из которых на всю сумму ничтожно мало, то $X$ имеет распределение, близкое к нормальному». Указанное обстоятельство в большой степени объясняет то важное место, которое занимают гауссовские процессы в практике исследований, поскольку для большинства случайных величин выполняются условия теоремы Ляпунова.

Плотность вероятности гауссовского (нормального) случайного процесса имеет вид симметричного колокола, быстро убывающего по мере отклонения от центра. Одномерный нормальный закон распределения плотности вероятности

$$p(x)=\frac{1}{\sqrt{2\pi}\sigma_x}\mathrm{e}^{-\frac{(x-m_x)^2}{2\sigma_x^2}}. \qquad (3.15)$$

На рис. 3.5 показаны графики плотности вероятности нормального закона распределения случайного процесса для трех значений СКО $\sigma_x$ (заметим, что при $\sigma_x = 1$ плотность вероятности $p = 0{,}4$). Функция плотности вероятности $p(x - m_x)$ данного закона является симметричной относительно нуля (или среднего значения — математического ожидания). С увеличением $\sigma_x$ максимум функции уменьшается, а кривая плотности вероятности становится более пологой относительно оси абсцисс.

Интегральную функцию распределения процесса вычислим как вероятность пребывания величины $x$ в некотором интервале случайных величин $-\Delta, \Delta$ (здесь $\Delta = x - m_x$ — новая переменная). Подставляя соотношение (3.15) в (3.4) при $m_x = 0$, получим

$$F(x)=P(-\Delta\le x\le\Delta)=\int_{-\infty}^{x}p(x,t_1)dx=\frac{1}{\sqrt{2\pi}\sigma_x}\int_{-\Delta}^{\Delta}\mathrm{e}^{-\frac{x^2}{2\sigma_x^2}}dx=\frac{2}{\sqrt{2\pi}\sigma_x}\int_{0}^{\Delta}\mathrm{e}^{-\frac{x^2}{2\sigma_x^2}}dx. \qquad (3.16)$$

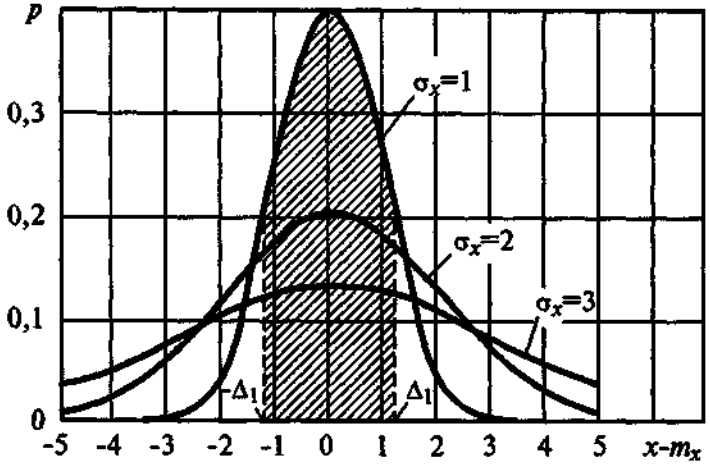



**Рис. 3.5.** Графики плотности вероятности нормального закона распределения

Отметим геометрическую интерпретацию закона распределения (3.16). На графике плотности вероятности (рис. 3.5) для конкретного СКО $\sigma_x$ и интервала значений $(-\Delta_1, \Delta_1)$ вероятность численно равна площади $S$ заштрихованной фигуры, ограниченной функцией $p(x)$, отрезком оси $\Delta$ от $-\Delta_1$ до $\Delta_1$ и ординатами $p(-\Delta_1)$, $p(\Delta_1)$. Чем шире интервал значений $x$ $(-\Delta_1, \Delta_1)$, тем больше площадь $S$, т. е. больше вероятность попадания случайных величин $\Delta$ в этот интервал. Для интервала $-\infty$, $\infty$ вероятность $P(-\infty < \Delta < \infty) = 1$. Чтобы удобнее вести анализ процесса и расчеты числовых характеристик, свяжем $\Delta$ с средним квадратическим отклонением $\sigma_x$, введя новую переменную $y = x/\sigma_x$. Тогда функция распределения (3.16) будет:

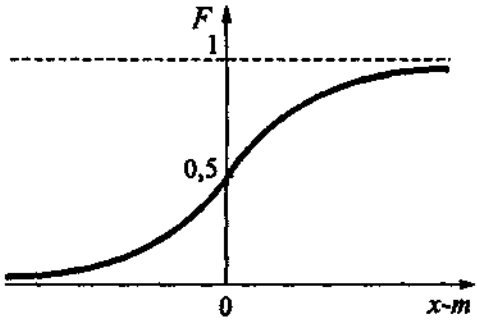



**Рис. 3.6.** Интегральная функция нормального закона распределения

$$F(x) = \frac{2}{\sqrt{2\pi}} \int_0^{\Delta/\sigma_x} e^{-\frac{y^2}{2}} dy = 2\Phi\left(\frac{\Delta}{\sigma_x}\right) = 2\Phi(z), \qquad (3.17)$$

где $z = \Delta/\sigma_x$.

Выражение

$$\Phi(z) = \Phi\left(\frac{x - m_x}{\sigma_x}\right) = \frac{1}{\sqrt{2\pi}} \int_0^{z} e^{-\frac{y^2}{2}} dy \qquad (3.18)$$

представляет собой *функцию Лапласа*, или *интеграл вероятности*. Эта вероятностная функция хорошо изучена и табулирована (табл. 3.1).

Таблица 3.1. **Значения интеграла вероятности Φ(z)**

| $z$ | $\Phi(z)$ | $z$ | $\Phi(z)$ | $z$ | $\Phi(z)$ | $z$ | $\Phi(z)$ |
|---|---|---|---|---|---|---|---|
| 0,00 | 0,000 | 0,70 | 0,516 | 1,40 | 0,839 | 2,25 | 0,976 |
| 0,10 | 0,080 | 0,80 | 0,576 | 1,50 | 0,866 | 2,50 | 0,988 |
| 0,20 | 0,159 | 0,90 | 0,632 | 1,60 | 0,890 | 2,75 | 0,994 |
| 0,30 | 0,236 | 1,00 | 0,683 | 1,70 | 0,911 | 3,00 | 0,997 |
| 0,40 | 0,311 | 1,10 | 0,729 | 1,80 | 0,928 | 3,30 | 0,999 |
| 0,50 | 0,383 | 1,20 | 0,770 | 1,90 | 0,943 | 3,50 | 0,999 |
| 0,60 | 0,452 | 1,30 | 0,806 | 2,00 | 0,955 | 4,00 | 0,999 |

Нормальный закон распределения, представленный в зависимости от относительного аргумента $z = \Delta/\sigma_x$, называют *нормированным* (иногда употребляют термин «стандартный») *нормальным законом* и задают более простым соотношением

$$p_н(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}. \qquad (3.19)$$

График функции нормированного нормального закона при $z = \Delta$ совпадает с графиком нормального закона (3.15) для среднеквадратического отклонения $\sigma_x = 1$ (см. рис. 3.5). Нормирование нормального закона распределения случайных величин приводит к переносу начала координат в центр распределения и выражению абсциссы в долях среднеквадратического отклонения. Значения дифференциальной функции нормированного нормального закона распределения случайных величин сведены в таблицы, которые можно найти в литературе по теории вероятности и теории случайных процессов.

Интегральная функция распределения нормального закона распределения случайных величин имеет вид монотонной нечетной симметричной кривой, принимающей значения от нуля до единицы при изменении аргумента $x$ от $-\infty$ до $\infty$ (см. рис. 3.6). Заметим, нормальный случайный процесс занимает исключительное положение среди случайных процессов. Наибольшее число теоретических результатов в статистической радиотехнике получено именно применительно к нормальным процессам. Фактически любая многомерная плотность вероятности гауссова случайного процесса определяется двумя характеристиками — математическим ожиданием и функцией корреляции. При негауссовом случайном процессе на входе отыскание закона распределения на выходе цепи является сложной задачей, не имеющей прямого однозначного решения.

### Равномерный закон распределения плотности вероятности

Наиболее удобным для анализа законом распределения плотности вероятности является равномерное распределение (*uniform distribution*) случайных величин (рис. 3.7), у которого плотность вероятности постоянна для данного ин-

тервала и равна нулю за его пределами. Закон наиболее характерен для случайных сигналов при преобразовании исследуемых непрерывных физических величин в дискретные и цифровые значения, например преобразования аналоговых сигналов в дискретные и затем цифровые в АЦП (погрешности квантования).

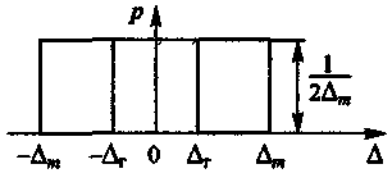



**Рис. 3.7.** График равномерного закона распределения

Все возможные значения, описываемые равномерным законом распределения, расположены в интервале $(-\Delta_m, \Delta_m)$, где $\Delta_m$ — максимальное значение. Аналитически плотность вероятности равномерного закона распределения

$$p(\Delta)=\begin{cases}\dfrac{1}{2\Delta_m}, & -\Delta_m \le \Delta \le \Delta_m;\\ 0, & \Delta < -\Delta_m \text{ и } \Delta > \Delta_m.\end{cases} \tag{3.20}$$

Вероятность того, что любое случайное значение $\Delta$ находится в некотором симметричном интервале значений $(-\Delta_1, \Delta_1)$, определяется с помощью (3.4) при подстановке в него значения плотности вероятности $p(\Delta) = 1/(2\Delta_m)$

$$P(-\Delta_1 \le \Delta \le \Delta_1) = \int_{-\Delta_1}^{\Delta_1} p(\Delta)d\Delta = \int_{-\Delta_1}^{\Delta_1} \frac{1}{2\Delta_m}d\Delta = \frac{1}{2\Delta_m}\int_{-\Delta_1}^{\Delta_1} d\Delta = \frac{\Delta_1}{\Delta_m}. \tag{3.21}$$

На графике равномерного закона распределения (см. рис. 3.7) площадь заштрихованного прямоугольника с основанием $2\Delta_1$ и высотой $1/(2\Delta_m)$ равна вероятности, вычисленной по (3.21). Для равномерного закона распределения, симметричного относительно центра $\Delta = 0$, расчет СКО $\sigma_x$ выполняют с помощью известной в теории вероятности формулы для дисперсии случайной величины

$$\sigma_x = \sqrt{D} = \sqrt{\int_{-\infty}^{\infty} \Delta^2 p(\Delta)d\Delta} = \sqrt{\frac{1}{2\Delta_m}\int_{-\Delta_m}^{\Delta_m} \Delta^2 d\Delta} = \sqrt{\frac{\Delta_m^2}{3}} = \frac{\Delta_m}{\sqrt{3}}. \tag{3.22}$$

Тогда собственно дисперсия $D = \Delta_m^2/3$.

### Закон распределения Пуассона

В ряде задач телекоммуникаций сталкиваются с распределением случайных целочисленных дискретных величин, представляющим последовательность точек, расположенных случайным образом, в частности, на оси времени. Такие точки могут соответствовать различным событиям, например моментам времени поступления заявок на обслуживание, моментам времени наступления отказов в какой-либо системе и др. Точечный случайный поток может встречаться и в задаче распределения вызовов на телефонной станции в течение суток. Общее число вызовов в течение суток — величина случайная.

Для каждого временно́го интервала $T$ путем наблюдений можно установить среднее число вызовов $\lambda T$ (математическое ожидание). Коэффициент пропорциональности $\lambda$ характеризует интенсивность телефонных вызовов (среднее число вызовов в единицу времени).



С. Пуассон

Вероятность появления $n$ вызовов на любой телефонной станции $P_T(n)$ на некотором интервале времени $(0, T)$ на практике чаще всего определяется *распределением Пуассона* (Симеон Пуассон — *Simeon Poisson*, 1781–1840; французский математик)

$$P_T(n) = \frac{\lambda^n}{n!} e^{-\lambda}. \tag{3.23}$$

Такое распределение задают только одним параметром (*параметром* закона Пуассона) $\lambda$ — некоторой положительной величиной, значение которого равно среднему значению случайных величин распределения и одновременно задает значение дисперсии распределения. Для распределения Пуассона

$$D_{\text{П}} = \lambda, \quad \sigma_{\text{П}} = \sqrt{\lambda}. \tag{3.24}$$

Сумма всех плотностей вероятностей для распределения Пуассона

$$\sum_{i=1}^{\infty} p_i = e^{-\lambda} + \lambda e^{-\lambda} + \frac{\lambda^2 e^{-\lambda}}{2!} + \ldots + \frac{\lambda^n e^{-\lambda}}{n!} + \ldots =$$
$$= e^{-\lambda}\left(1 + \lambda + \frac{\lambda^2}{2!} + \ldots + \frac{\lambda^n}{n!} + \ldots\right) = e^{-\lambda} \cdot e^{\lambda} = 1 \tag{3.25}$$

(здесь учтено, что в скобках в правой части записано разложение в ряд Тейлора функции $e^x$ при значении $x = \lambda$).

Графически закон распределения дискретной случайной величины удобно представлять в виде многоугольника распределения — ломаной линии, соединяющей точки плоскости с координатами $(n, p_i)$. На рис 3.8 приведены многоугольники (*полигоны*) распределения случайной величины $X$, имеющей распределение Пуассона с параметром $\lambda$ = 0,5; 1; 2; 3,5; 5. Распределение Пуассона называют *распределением вероятностей количества редких событий*, для оценки которых оно применяется наиболее часто, безотносительно к их физической сущности.

Условия справедливости распределения Пуассона применительно к случайным событиям достаточно просты.

1. Случайная величина может принимать только нулевые и целые положительные значения (как, например, число отказов аппаратуры или телефонных звонков за определенный интервал времени).

2. Вероятность отдельного события на бесконечно малом интервале времени счета событий пропорциональна величине этого интервала и при его стремле-

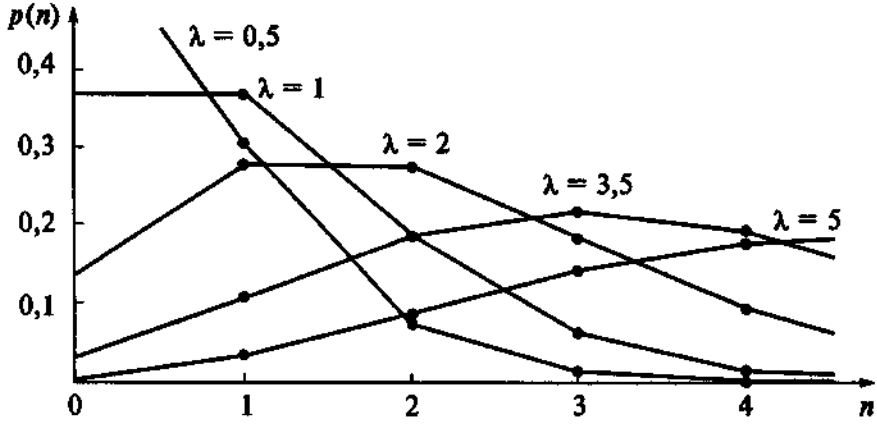



**Рис. 3.8.** Распределение Пуассона для ряда параметров λ

нии к нулю также стремится к нулю, как бесконечно малая первого порядка, а вероятности событий более высокой кратности (2 и более событий) стремятся к нулю, как бесконечно малые более высоких порядков.

3. События в неперекрывающихся интервалах времени их счета статистически независимы.

**Пример 3.1.** Найти плотность вероятности, математическое ожидание и дисперсию случайного процесса $U(t)$ вида «телеграфного сигнала», реализация которого $u(t)$ показана на рис. 3.9.

В теории связи таким сигналом называют случайный процесс, реализации $u(t)$ которого принимают значения уровней $U = +1$ и $U = -1$, причем «опрокидывания» уровня происходят в случайные моменты времени и число $n$ «опрокидываний» уровня, происходящих за некоторый период $T$, является случайной величиной с дискретным распределением вероятности, описываемым законом Пуассона.

Для данного случайного процесса в соотношении (3.23) параметр λ — коэффициент, определяющий среднюю частоту возникновения изменения полярности данного сигнала в единицу времени; $P_T(n)$ — вероятность того, что за период $T$ произойдет $n$ изменений полярности; при этом $P(1) = P(-1) = 0{,}5$.

Р е ш е н и е. Скачки уровня сигнала происходят в случайные моменты времени, поэтому аналитически записать отдельную реализацию данного случайного процесса оказывается весьма затруднительно.

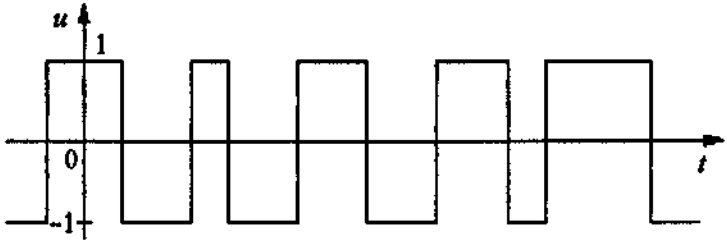



**Рис. 3.9.** Реализация случайного телеграфного сигнала

Конкретную реализацию случайного процесса удобнее задать бесконечным множеством случайных величин ± 1 — моментов изменения уровня, а характеристики случайного процесса определять статистическими свойствами этих величин.

1. Плотность вероятности

$$P(n) = P(1) + P(-1) = 1.$$

2. Математическое ожидание

$$m_u = 0,$$

что очевидно из графика реализации.

3. Дисперсия

$$D(t) = \int_0^T [u(t) - m(t)]^2 P(n)dt = \int_0^T 1^2 P(n)dt = T.$$

4. При вычислении корреляционной функции телеграфного сигнала каждое отдельное произведение $u_k(t)u_k(t+\tau)$ равно либо $U^2 = 1$, либо $-U^2 = -1$ в зависимости от совпадения или несовпадения знаков $u_k(t)$ и $u_k(t+\tau)$, причем вероятность значения корреляционной функции $U^2 = 1$ равна сумме «четных» вероятностей $P(0) + P(2) + P(4) + \ldots$, а вероятность корреляционной функции $-U^2 = -1$ определяют суммой «нечетных» вероятностей $P(1) + P(3) + \ldots$ Следовательно,

$$R_u(\tau) = \mathrm{M}\{u_k(t)u_k(t+\tau)\} = U^2 \sum_{n=0}^{\infty} (-1)^n P(n) =$$

$$= U^2 e^{-\alpha|\tau|} \sum_{n=0}^{\infty} (-1)^n (\lambda|\tau|)^n \frac{1}{n!} = U^2 e^{-2\lambda|\tau|} = e^{-2\lambda|\tau|}.$$

Параметр $\lambda$ полностью определяет корреляционные и спектральные свойства телеграфного сигнала, приведенные на рис. 3.10, *а*. При $\lambda \to 0$ характеристики сигнала приближаются к характеристикам постоянной составляющей, при $\lambda \to \infty$ — к характеристикам белого шума.

Можно показать, что чем больше $\lambda$, тем меньше время корреляции. При $\lambda \to 0$ процесс вырождается в детерминированный (стремится к постоянной составляющей). При $\lambda \to \infty$ процесс вырождается в белый шум с некоррелированными отсчетами на соседних точках времени.

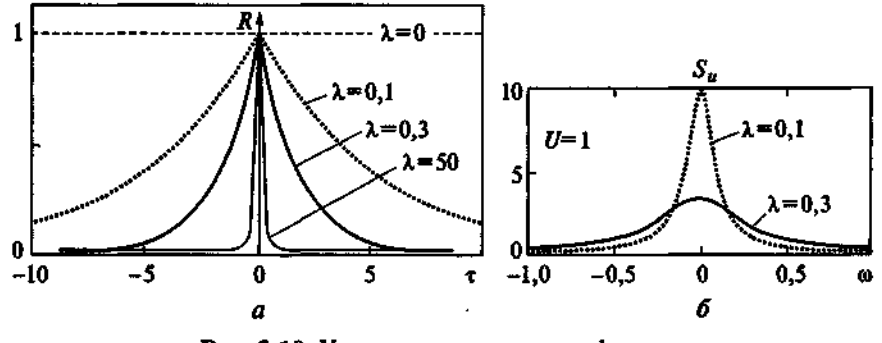



**Рис. 3.10.** Характеристики телеграфного сигнала:
*а* — функция корреляции; *б* — спектральная плотность

5. Спектральная плотность телеграфного сигнала (рис. 3.10, *б*)

$$S_u(\omega) = \int_{-\infty}^{\infty} R_u(\tau) e^{-j\omega\tau} d\tau = \frac{\lambda U^2}{\lambda^2 + \omega^2} = \frac{\lambda}{\lambda^2 + \omega^2}.$$

### *Характеристическая функция*

Как отмечалось ранее, при негауссовском случайном процессе на входе отыскание закона распределения на выходе инерционной, а особенно, нелинейной цепи является чрезвычайно сложной задачей, зачастую не имеющей прямого решения. Существуют лишь приближенные методы решения, связанные с большими вычислительными проблемами.

Один из таких методов основан на использовании характеристических функций случайного процесса и известных соотношений между характеристической функцией и моментами распределения процесса. Так, в частности, в радиотехнике и теории передачи информации часто требуется определить вероятностные характеристики информационных случайных сигналов, преобразованных нелинейными или параметрическими цепями и устройствами (например, в процессе нелинейного усиления, при модуляции, детектировании и т. д.). В этих случаях весьма эффективно применение теории характеристических функций. Пусть имеется случайный процесс $X(t)$, описываемый *одномерной плотностью вероятности* $p(x)$. Конкретные составляющие $x_1(t)$, $x_2(t)$, ..., $x_k(t)$, ..., $x_n(t)$ этой функции, отражающие весь ансамбль реализаций случайного процесса, являются отдельными преобразуемыми стохастическими колебаниями, составляющими групповой сигнал. Такие сигналы на практике имеют место в различных системах радиосвязи с многостанционным доступом.

В теории вероятности для таких случайных процессов большую роль играет статистическое среднее вида

$$G(v) = \overline{e^{jvx}} = \int_{-\infty}^{\infty} p(x) e^{jvx} dx, \qquad (3.26)$$

представляющее собой по существу обратное преобразование Фурье (без привычного коэффициента $1/(2\pi)$) от плотности распределения вероятности $p(x)$ и называемое *характеристической функцией* случайной величины $X$. В этом случае функция $G(v)$ является прямым преобразованием Фурье от плотности вероятности (с коэффициентом $1/(2\pi)$), поэтому

$$p(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} G(v) e^{-jvx} dv. \qquad (3.27)$$

Итак, плотность вероятности $p(x)$ случайной величины может рассматриваться как частотный спектр характеристической функции данной случайной величины, а последняя, в свою очередь, как координатное представление ее закона распределения.

Из математики известно, что можно провести асимптотическое разложение характеристических функций для произвольного распределения. В качестве примера определим характеристическую функцию гауссовского случайного процесса с заданным математическим ожиданием $m$ и СКО σ:

$$G(v)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{(x-m)^2}{2\sigma^2}}e^{jvx}dx=e^{jmv-\frac{\sigma^2v^2}{2}}. \qquad (3.28)$$

Для центрированного гауссовского процесса ($m = 0$) имеем

$$G(v)=e^{-\frac{\sigma^2v^2}{2}}. \qquad (3.29)$$

Для случайной величины, равномерно распределенной на отрезке $0 \le x \le a$,

$$G(v)=\frac{e^{jav}-1}{jav}. \qquad (3.30)$$

Через операцию определения характеристической функции упрощается вычисление моментов случайной величины любого порядка. Такое удобство использования характеристической функции при анализе случайных процессов и вычислении их параметров следует из свойств преобразования Фурье. В частности, характеристическая функция суммы независимых случайных величин (суммы случайных процессов) равна произведению их характеристических функций. Если известна характеристическая функция, то легко найти числовые характеристики случайного процесса. Действительно, поскольку из формулы (3.26) получаем после дифференцирования по степени $n$

$$\frac{d^nG}{dv^n}=j^n\int_{-\infty}^{\infty}x^np(x)e^{jvx}dx,$$

то, полагая здесь $v = 0$ и сравнивая результат с моментом $n$-го порядка случайной величины (3.7), находим

$$m_n=j^{-n}G^{(n)}(0). \qquad (3.31)$$

Начальные моменты случайного процесса можно определить через производные характеристической функции при $v = 0$. Раскладывая характеристическую функцию случайного процесса в ряд Тейлора, запишем

$$G(v)=\sum_{n=0}^{\infty}\left[\frac{d^nG(v)}{dv^n}\right]_{v=0}\frac{v^n}{n!}. \qquad (3.32)$$

С учетом (3.31) получим

$$G(v)=\sum_{n=0}^{\infty}\frac{m_n}{n!}(jv)^n. \qquad (3.33)$$

Следовательно, характеристическая функция случайного процесса определяется его моментами распределения. Учитывая связь характеристической функции и плотности вероятности распределения, можно сказать, что моменты распределения вероятности определяют и плотность вероятности случайного процесса. Чем большее число начальных моментов распределения случайного процесса известно, тем точнее может быть определена характеристическая функция и плотность вероятности случайного процесса.

С помощью характеристической функции удобно также находить плотность вероятности случайной величины, подвергнутой функциональному преобразованию, например нелинейному усилению, модуляции, детектированию и т. д. В частности, при $u = f(x)$ получаем

$$G_u(v) = \overline{e^{jvu}} = \overline{e^{jvf(x)}}.$$

Теперь, если осуществимо прямое преобразование Фурье вида (3.27), то поставленная задача вычисления спектра выходного сигнала любого устройства преобразования или усиления будет решена.

Еще одно очень важное дополнение к теории применения характеристических функций при практическом анализе сигналов. Из курса теории вероятностей известно, что математическое ожидание и дисперсия случайного процесса достаточно просто определяются через производные логарифма характеристической функции.

**Пример 3.2.** Пусть задан случайный сигнал $u = U_0\cos x$, у которого амплитуда $U_0 = \text{const}$, а фаза $x$ — значение случайной величины, равномерно распределенной на отрезке $-\pi \le x \le \pi$.

С помощью характеристической функции вычислить плотность вероятности распределения амплитуд случайного сигнала.

Р е ш е н и е. Поскольку плотность вероятности распределения фазы случайного сигнала определяется как $p_x(x) = 1/(2\pi)$, то из формулы (3.26) находим характеристическую функцию

$$G(v) = \int_{-\infty}^{\infty} p(x)e^{jvu}dx = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{jvU_0\cos x}dx = J_0(vU_0),$$

где $J_0$ — функция Бесселя первого рода нулевого порядка.

Применяя известный в математике табличный интеграл и воспользовавшись последней формулой, получаем для функции плотности распределения амплитуд случайного сигнала

$$p_u(u) = \frac{1}{2\pi}\int_{-\infty}^{\infty} J_0(vU_0)e^{jvu}dv = \begin{cases} \dfrac{1}{\pi\sqrt{U_0^2 - u^2}}, & |u| \le U_0; \\ 0, & |u| > U_0. \end{cases}$$

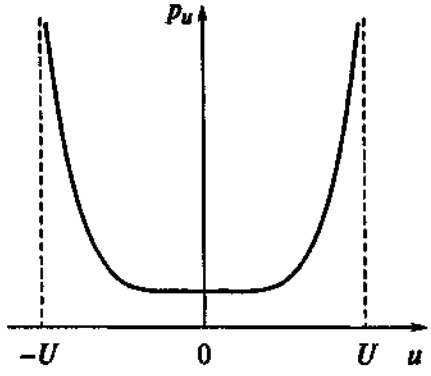



**Рис. 3.11. График плотности вероятности распределения амплитуд сигнала со случайной начальной фазой**

Такой вид графика распределения плотности вероятности случайного процесса со случайной начальной (рис. 3.11) фазой связан с тем, что если выполнить большую серию испытаний, каждый раз случайным образом выбирая значения фазы $x$ случайного сигнала из заданной ее области, то сигнальная функция $U_0\cos x$ чаще будет принимать мгновенные значения, близкие к амплитудам $\pm U_0$, нежели близкие к нулевому уровню напряжения амплитуд.

## 3.3. Стационарные случайные процессы

Вероятностные и корреляционные характеристики случайных процессов определяются с помощью одного или нескольких моментов времени (сечений). Однако существует класс случайных процессов, у которых зависимость характеристик от времени отсутствует и при определенных условиях ряд вероятностных характеристик может быть определен путем усреднения по всему ансамблю реализаций. В других случаях для данных целей может быть осуществлено *усреднение по времени* (*time averaging*) с использованием одной, $k$-й реализации $x_k(t)$ случайного процесса $X(t)$. Наличие и отсутствие зависимости вероятностных характеристик от времени или от номера реализации определяет такие фундаментальные свойства процесса, как стационарность и эргодичность.

Особое место среди случайных процессов занимает стационарный случайный процесс, с которым наиболее часто приходится сталкиваться при решении разнообразных практических задач радиотехники.

*Стационарными* называют случайные процессы, статистические характеристики которых не изменяются во времени. Примером стационарного случайного процесса является внутренние шумы приемников, тепловой шум (имеется в виду тепловой ток) транзистора, стабилитрона, диода и других полупроводниковых и электронных приборов. В радиотехнических приложениях теории случайных процессов условие стационарности обычно ограничивается требованием независимости от времени только одномерной и двумерной плотностей вероятности. Выполнение этого условия позволяет считать, что среднее значение (пер-

вый момент), средний квадрат и дисперсия случайного процесса не зависят от времени, а корреляционная функция зависит не от самих моментов времени $t_1$ и $t_2$, а только от интервала между ними $\tau = t_2 - t_1$, т. е. от одного аргумента. Случайные процессы, удовлетворяющие условиям стационарности на ограниченных интервалах, также обычно относят к числу стационарных в широком смысле и называют *квазистационарными*.

С учетом предложенных ограничений при записи статистических параметров стационарного случайного процесса можно опускать обозначения фиксированных моментов времени. В этом случае математическое ожидание и дисперсия не зависят от времени, т. е. формулы (3.9) и (3.11) примут вид:

$$m_x = \overline{x(t)} = \int_{-\infty}^{\infty} xp(x)dx; \tag{3.34}$$

$$\overline{x^2} = \int_{-\infty}^{\infty} x^2 p(x)dx; \tag{3.35}$$

$$D_x = \sigma_x^2 = \int_{-\infty}^{\infty} (x - m_x)^2 p(x)dx. \tag{3.36}$$

Нетрудно показать, что функция корреляции случайного стационарного процесса зависит только от разности $\tau = t_2 - t_1$, и поэтому

$$R_x(t_1, t_2) = R_x(\tau). \tag{3.37}$$

Из определения стационарности случайного процесса следует, что его функция корреляции является четной относительно $\tau = 0$:

$$R_x(\tau) = R_x(-\tau). \tag{3.38}$$

В реальности даже для технических систем зачастую трудно ожидать стационарности случайных процессов. Вероятностные характеристики нестационарных процессов зависят от времени и для их описания требуется сложный математический аппарат. Поэтому разработка методов исследования нестационарных процессов остается актуальной проблемой. Часто ограничиваются применением методов, предполагающих стационарность, но на коротких временны́х интервалах (например, кратковременный спектральный анализ), меняя с течением времени их параметры. Более радикальное решение — переход к вейвлетным (*wavelete*) методам. В них упрощающее предположение — гипотеза о подобии свойств стационарного случайного процесса при больших и малых масштабах (возможность масштабирования).

*Стационарность* — не единственное полезное свойство случайных процессов, позволяющее подробно их исследовать. Еще одним свойством такого рода является *эргодичность* (*ergodicity*; от греч. *ergon* — работа).

Условие эргодичности случайного процесса включает в себя и условие его стационарности. Свойством эргодичности обладает значительная часть стационарных случайных процессов. Понятие «эргодичность» заимствовано из статистической физики, где изучают свойства физических систем (например, газов), состоящих из большого числа одинаковых частиц. В таких системах математические ожидания характеристик можно вычислить не только усреднением по всему множеству составляющих систему частиц, но и по траектории движения одной частицы. Эргодичность проявляется в том, что со временем процесс становится однородным, т. е. любая реализация этого процесса окажется в какой-то момент сколь угодно близко к произвольному заданному состоянию системы.

Стационарный случайный процесс является *эргодическим*, если усреднение по ансамблю реализаций можно заменить усреднением по времени одной реализации в пределах бесконечно длинного интервала времени $T_x$. Приведем простой пример: если у вас есть кубик (игральная кость) с числами на гранях от 1 до 6, то при 600 выбрасываниях число 1 выпадет примерно 100 раз. Можно поступить и по-другому: взять одинаковых 600 кубиков и бросить их все одновременно один раз. При этом около 100 кубиков также покажут грань с 1. Эргодические процессы просто описываются аналитически и, что не менее важно, их вероятностные характеристики можно легко получить экспериментально.

*Математическое ожидание* эргодического процесса вычисляется путем усреднения по бесконечному интервалу времени значений заданной реализации. Обозначая усреднение по времени угловыми скобками, запишем математическое ожидание эргодического случайного процесса

$$m_x = \langle x(t) \rangle = \lim_{T_x \to \infty} \frac{1}{T_x} \int_0^{T_x} x(t)dx. \tag{3.39}$$

Следует помнить, что математическое ожидание эргодического случайного процесса равно *постоянной составляющей* любой его реализации.

Средний квадрат

$$\langle x^2(t) \rangle = \lim_{T_x \to \infty} \frac{1}{T_x} \int_0^{T_x} x^2(t)dx \tag{3.40}$$

является средней мощностью всего случайного эргодического процесса.

*Дисперсия*

$$\sigma_x^2 = \langle [x(t) - m_x]^2 \rangle = \lim_{T_x \to \infty} \frac{1}{T_x} \int_0^{T_x} [x(t) - m_x]^2 dt = \lim_{T_x \to \infty} \frac{1}{T_x} \int_0^{T_x} x^2(t)dt - m_x^2 \tag{3.41}$$

определяет *мощность флуктуационной составляющей* эргодического процесса.

Как правило, при экспериментальном исследовании случайных процессов удается наблюдать одну реализацию сигнала, а не всего ансамбля. Если изучаемый процесс является эргодическим, то его реализация на достаточно большом интервале времени является «типичным» представителем всего ансамбля.

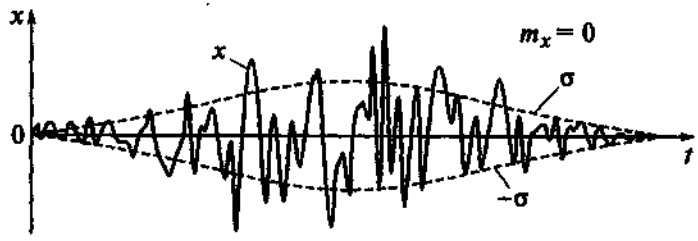



**Рис. 3.12.** Флуктуационная составляющая $x(t)$ со средним квадратическим отклонением $\pm\sigma$

На рис. 3.12 приведен пример реального случайного процесса $X(t)$ в виде одной из реализаций флуктуационной составляющей $x(t)$; там же показано СКО $\pm\sigma$ от математического ожидания $m_x$ (для упрощения графика выбрано $m_x = 0$).

В электрических цепях широко используют переходные (разделительные) *RC*-цепи, не пропускающие постоянной составляющей. Поэтому для реальных стационарных эргодических процессов математическое ожидание $m_x = 0$.

*Функция корреляции* в этом случае имеет более простой вид

$$R_x(\tau) = \lim_{T_x \to \infty} \frac{1}{T_x} \int_0^{T_x} x(t)x(t-\tau)dt. \qquad (3.42)$$

Выражение (3.42) внешне совпадает с определением (2.102) автокорреляционной функции детерминированного периодического сигнала. Непосредственно из (3.42) вытекает четность функции $R_x(\tau)$ относительно сдвига $\tau$.

Итак, функция корреляции, по существу, является одной из основных характеристик данного случайного процесса. Чем медленнее, плавней изменяется во времени $x(t)$, тем больше интервал времени $\tau$, в пределах которого наблюдается статистическая связь между мгновенными значениями случайной функции.

Важно заметить, что достаточным условием эргодичности случайного процесса, стационарного в широком смысле, является стремление к нулю его корреляционной функции с ростом временно́го сдвига $\tau$

$$\lim_{\tau \to \infty} R_x(\tau) = 0. \qquad (3.43)$$

Согласно приведенным формулам по единственной реализации можно определить математическое ожидание, дисперсию и корреляционную функцию эргодического случайного процесса. На практике интегрирование выполняется не в бесконечных пределах, а на конечном интервале, длина которого должна быть тем больше, чем выше требования к точности результатов исследования.

Далее изучение стационарного случайного процесса будем проводить с учетом свойства его эргодичности, одним из признаков которого является равенство среднего значения по множеству реализаций (3.34) среднему значению по времени одной реализации (3.39), т. е.

$$\overline{x(t)} = \langle x(t) \rangle.$$

В общем случае результаты усреднения случайных процессов по совокупности и по времени неодинаковы. Предел выборочного среднего по совокупности представляет вероятностную характеристику, выражающую зависимость вероятностных свойств процесса от текущего времени. Предел выборочного среднего по времени представляет вероятностную характеристику, выражающую зависимость вероятностных свойств процесса от номера реализации.

При экспериментальном исследовании случайных процессов используют временны́е характеристики (3.39) – (3.42), поскольку исследователю доступно наблюдение одной реализации, а не все их множество из ансамбля. Интегрирование выполняется не в бесконечных пределах, а на конечном интервале $T$, длительность которого должна быть тем больше, чем выше требование к точности результатов измерения. Кстати, в радиотехнике возможны и другие виды случайных процессов.

*Стационарный неэргодический* случайный процесс такой, у которого эквивалентны временны́е сечения — вероятностные характеристики не зависят от текущего времени, но не эквивалентны реализации — вероятностные характеристики зависят от номера реализации (наблюдаются в радиолокации).

*Нестационарный эргодический* случайный процесс — процесс, у которого эквивалентны реализации — вероятностные характеристики не зависят от номера реализации, но не эквивалентны временны́е сечения — вероятностные характеристики зависят от текущего времени (например, процессы на Солнце).

*Нестационарный неэргодический* — случайный процесс, у которого не эквивалентны ни временны́е сечения, ни реализации. В частности, сигналы связи во времени меняют свои мгновенные значения, причем эти изменения могут быть предсказаны лишь с некоторой (меньше 1) вероятностью. Значит, сигналы связи являются неэргодическими и нестационарными случайными процессами и методы их описания достаточно сложны.

Итак, обобщенная классификация на основе стационарности и эргодичности дает четыре класса случайных процессов: *стационарные эргодические, стационарные неэргодические, нестационарные эргодические, нестационарные неэргодические*. Учет этих свойств случайных процессов играет большую роль. В практической радиотехнике приходится сталкиваться со всеми перечисленными видами случайных процессов. Поэтому принято *моделировать* реальные сигналы стационарным и эргодическим случайным процессом, полученным в результате двойного усреднения — вначале по множеству реализаций ансамбля определяют числовые характеристики для достаточно большого числа моментов времени, а затем эти характеристики усредняют по времени. Полученная таким образом модель отражает некий «*среднестатистический*» сигнал, параметры которого и используются при практических расчетах. Ясно, что в расчетах неизбежно возникают ошибки, которые преодолевают некоторым завышением требований к рассчитываемым устройствам с помощью компьютерных и натурных экспериментов и т. д. Постоянно проводятся работы по накоплению статистических материалов для совершенствования моделей сигналов.

**Пример 3.3.** Случайный процесс $U(t)$ состоит из гармонических реализаций $u(t) = U_m\cos(\omega_0 t+\varphi)$, где амплитуда $U_m$ и частота $\omega_0$ — постоянные параметры, а начальная фаза реализации $\varphi$ — случайная величина, которая с одинаковой вероятностью принимает значение в интервале $-\pi$, $\pi$ (рис. 3.13). Найти числовые характеристики процесса и определить, является ли он стационарным?

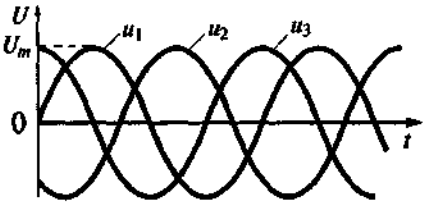



**Рис. 3.13.** Ансамбль гармонических колебаний со случайной фазой

Р е ш е н и е. Заданное распределение начальных фаз означает, что плотность вероятности случайной фазы любого колебания $p(\varphi) = 1/(2\pi)$. Тогда, согласно (3.34), математическое ожидание для амплитуд гармонических напряжений

$$m_x = \frac{U_m}{2\pi}\int_{-\pi}^{\pi}\cos(\omega_0 t + \varphi)d\varphi = 0.$$

По формуле (3.36) находим дисперсию

$$D(x) = \sigma_x^2 = \int_{-\pi}^{\pi}U_m^2\cos^2(\omega_0 t + \varphi)p(\varphi)d\varphi = \frac{U_m^2}{2\pi}\int_{-\pi}^{\pi}\cos^2(\omega_0 t + \varphi)d\varphi = \frac{U_m^2}{2}.$$

Тот факт, что реализации случайного процесса являются периодическими функциями, позволяет упростить вычисления, заменив усреднение по бесконечному промежутку времени усреднением по одному периоду, равному в этом случае $T = 2\pi/\omega_0$. тогда функцию корреляции получаем усреднением по времени произведения двух напряжений

$$u(t_1)u(t_2) = U_m\cos(\omega_0 t_1 + \varphi)U_m\cos(\omega_0 t_2 + \varphi) = 0{,}5U_m^2\{\cos\omega_0(t_2 - t_1) + \cos[\omega_0(t_2 + t_1) + 2\varphi]\}.$$

В правой части этого выражения первое слагаемое в фигурных скобках является детерминированным колебанием, поскольку в нем отсутствует случайная фаза. Второе слагаемое при статистическом усреднении по фазе с помощью одномерной плотности вероятности обращается в нуль. Поэтому функция корреляции

$$R_x(t_1, t_2) = 0{,}5U_m^2\cos\omega_0(t_2 - t_1) = 0{,}5U_m^2\cos\omega_0\tau,$$

где $\tau = t_2 - t_1$.

Все искомые числовые характеристики не зависят от времени, и заданный случайный процесс — *стационарным*. Следует подчеркнуть, что *любой* случайный процесс, реализации которого являются гармоническими функциями, идентичными по форме и различающимися лишь равномерно распределенной в пределах периода «начальной фазой», будет не только стационарным, но и *эргодическим*.

**Пример 3.4.** Случайный процесс $U(t)$ состоит из реализаций $u(t) = U_m\cos(\omega_0 t + \varphi)$, причем $\omega_0$ и $\varphi$ — постоянные параметры, амплитуда $U_m$ — случайная величина с произвольным законом распределения и равновероятная в интервале от 0 до $U_{max}$ (рис. 3.14). Определить, является ли этот процесс стационарным.

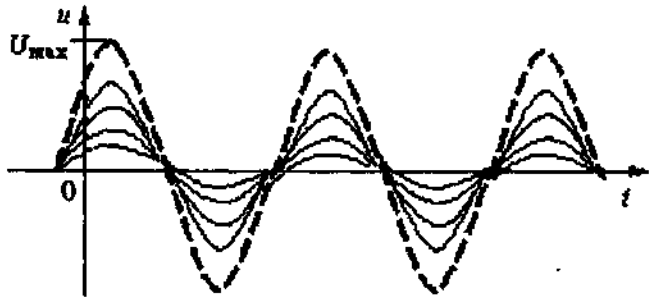



**Рис. 3.14.** Ансамбль гармонических колебаний со случайной амплитудой

Р е ш е н и е. Математическое ожидание $\bar{u} = \bar{U}_m \cos(\omega_0 t + \varphi)$ не зависит от времени лишь при $\bar{U}_m = 0$. Поэтому случайный процесс нестационарный.

## 3.4. Спектральный и корреляционный анализ случайных процессов

Для оценки частотных и временны́х свойств случайных процессов наиболее рационально использовать спектральные и корреляционные характеристики.

При теоретическом исследовании случайного процесса одномерная плотность вероятности и связанные с ней числовые характеристики позволяют получить важную информацию о его свойствах. Между тем для решения многих практических задач этих сведений часто недостаточно, поскольку они дают вероятностное представление о случайном процессе только в отдельные моменты и ничего не говорят о том, как он изменяется во времени.

Чтобы описать его временны́е характеристики, приходится использовать корреляционную функцию или привлечь для этого спектральные характеристики исследуемого случайного процесса. Теорию случайных процессов, основанную на использовании моментных функций не выше второго порядка, называют *корреляционной теорией*. Функция корреляции характеризует скорость протекания случайного процесса и определяет статистическую связь между отдельными значениями одного и того же или двух различных случайных процессов в разные моменты времени.

В главе 2 была изложена спектральная теория детерминированных сигналов. Это определяет целесообразность обращения к методам спектрального анализа и при исследовании случайных процессов. Однако особенности случайных процессов (в частности, вероятностный характер реализаций) делают невозможным простой перенос методов анализа детерминированных сигналов на случайные процессы. Вместе с тем часто удается получить ряд важных спектральных характеристик случайных процессов, используя преобразование Фурье функций, полученных путем усреднения реализаций на конечном интервале времени $T_x$, т. е. анализируются стационарные эргодические процессы (в общем случае реализации случайного процесса не являются абсолютно интегрируемыми на всей оси

времени, поэтому к их спектральным плотностям следует относиться как к *обобщенным функциям*). Вместе с тем спектральный анализ случайных процессов несет особенности, которые требуют отдельного рассмотрения.

Рассмотрим конкретную реализацию $x_k(t)$ стационарного эргодического центрированного случайного процесса на интервале $T_x = t_2 - t_1$ (см. рис. 3.2, *б*). Поскольку параметры реализации можно определить по ее виду или аналитическому представлению, то она является детерминированной функцией, и к ней можно применить преобразования Фурье. Естественно, при этом различные реализации будут иметь разные спектры. Интерес же представляют *статистически усредненные* характеристики случайного процесса. С помощью прямого преобразования Фурье вычислим спектральную плотность реализации $x_k(t)$

$$\mathbf{S}_x(\omega) = \int_{-T_x/2}^{T_x/2} x_k(t) \mathrm{e}^{-j\omega t} dt. \tag{3.44}$$

Используя *равенство Парсеваля* (2.71), определим энергию реализации

$$Э_x = \int_{-\infty}^{\infty} x_k^2(t) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |\mathbf{S}_x(\omega)|^2 d\omega. \tag{3.45}$$

Поделив энергию на период $T_x$, найдем среднюю мощность реализации

$$P_x = \frac{Э_x}{T_x} = \overline{x_{kT}^2(t)} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{|\mathbf{S}_x(\omega)|^2}{T_x} d\omega. \tag{3.46}$$

При увеличении интервала времени $T_x$ энергия отрезка реализации сигнала неограниченно возрастает, а ее средняя мощность стремится к некоторому пределу. Совершив предельный переход по интервалу $T_x \to \infty$, получим мощность $k$-й реализации случайного стационарного эргодического процесса

$$P_{xT} = \left\langle \overline{x_k^2(t)} \right\rangle = \frac{1}{2\pi} \int_{-\infty}^{\infty} \lim_{T_x \to \infty} \frac{|\mathbf{S}_x(\omega)|^2}{T_x} d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathbf{W}_{xk}(\omega) d\omega. \tag{3.47}$$

Здесь

$$\mathbf{W}_{xk}(\omega) = \lim_{T_x \to \infty} \frac{|\mathbf{S}_x(\omega)|^2}{T_x} \tag{3.48}$$

— *спектральная плотность мощности*, или *спектр мощности* (*power spectral density, PSD*), $k$-й реализации случайного процесса.

Если осуществить в отношении (3.47) статистическое усреднение по времени по всему ансамблю реализаций и операцию предельного перехода для пе-

риода $T_x \to \infty$, то можно определить *спектральную плотность мощности*, или *спектр мощности случайного процесса*,

$$\mathbf{W}(\omega) = \lim_{T_x \to \infty} \frac{|\mathbf{S}_x(\omega)|^2}{T_x}. \qquad (3.49)$$

Поскольку анализируется центрированный стационарный эргодический случайный процесс, то средняя мощность любой его реализации равна дисперсии. Поэтому

$$D_x = \sigma_x^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathbf{W}(\omega) d\omega. \qquad (3.50)$$

Из определения спектральной плотности мощности случайного процесса (3.49) очевидно, что $\mathbf{W}(\omega)$ — *неотрицательная вещественная четная функция частоты*. Она не содержит информации о фазах спектральных составляющих и не позволяет восстановить отдельные реализации случайного процесса. И еще, как следует из (3.48) и (3.49), для нахождения спектральной плотности мощности всего (даже и неэргодического) процесса $\mathbf{W}(\omega)$ необходимо усреднение по множеству реализаций. Однако, если ограничиться рассмотрением эргодических процессов, можно считать, что найденная по одной реализации (т. е. путем усреднения по времени) функция $\mathbf{W}_{xk}(\omega)$ характеризует весь случайный процесс в целом.

### Теорема Винера–Хинчина

Установим связь между корреляционной функцией и спектром мощности стационарного эргодического случайного процесса. Предварительно для этого введем модуль произведения

$$|\mathbf{S}_x(\omega)|^2 = \mathbf{S}_x(\omega)\mathbf{S}_x^*(\omega), \qquad (3.51)$$

в котором

$$\mathbf{S}_x^*(\omega) = \int_{-T_x/2}^{T_x/2} x_k(t) e^{j\omega t} dt \qquad (3.52)$$

— комплексно-сопряженная спектральная функция конкретной реализации $x_k(t)$.

Как отмечалось ранее, корреляционная функция детерминированного сигнала и его энергетический спектр связаны преобразованиями Фурье. Воспользуемся формулой (3.51) и применим это свойство к отрезку конкретной реализации $x_k(t)$ случайного процесса длительностью $T_x$:

$$\int_{-T_x/2}^{T_x/2} x_k(t) x_k(t-\tau) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |\mathbf{S}_x(\omega)|^2 e^{j\omega\tau} d\omega. \qquad (3.53)$$

Разделим обе части этого равенства на $T_x$ и устремим этот интервал к бесконечности

$$\lim_{T_x\to\infty}\frac{1}{T_x}\int_{-T_x/2}^{T_x/2}x_k(t)x_k(t-\tau)dt=\frac{1}{2\pi}\int_{-\infty}^{\infty}\lim_{T_x\to\infty}\frac{|\mathbf{S}_x(\omega)|^2}{T_x}e^{j\omega\tau}d\omega. \qquad (3.54)$$

Пусть анализируется эргодический процесс. Тогда в левой части равенства (3.54) представлена корреляционная функция, полученная усреднением по времени; в правой части под интегралом содержится спектральная плотность мощности всего случайного процесса, отражаемая формулой (3.51). В результате получим, что корреляционная функция случайного процесса

$$R(\tau)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\mathbf{W}(\omega)e^{j\omega\tau}d\omega. \qquad (3.55)$$

Это соотношение является фундаментальным в теории случайных сигналов и в некоторой мере может рассматриваться как обратное преобразование Фурье корреляционной функции. Существует и прямое преобразование Фурье

$$\mathbf{W}(\omega)=\int_{-\infty}^{\infty}R(\tau)e^{-j\omega\tau}d\tau. \qquad (3.56)$$

Заметим, что отсутствие в (3.55) и (3.56) и далее индексов у функций $\mathbf{W}_x(\omega)$ и $R_x(\tau)$ показывает, что результаты справедливы по отношению ко всем случайным процессам. Соотношения (3.55) и (3.56) отражают содержание *теоремы Винера–Хинчина* (доказана в 1934 г. Н. Винером и независимо известным советским математиком А.Я. Хинчиным; работы А.Я. Хинчина сыграли большую роль в корреляционной теории случайных процессов), утверждающей, что спектр мощности и функция корреляции случайного процесса связаны между собой интегральными преобразованиями Фурье. Причем, чем шире этот спектр, тем хаотичнее реализации случайного процесса. Для теоремы Винера–Хинчина справедливы известные свойства преобразования Фурье для детерминированных сигналов. Например, *чем шире энергетический спектр случайного процесса, тем меньше интервал корреляции, и, соответственно, чем больше интервал корреляции, тем у́же спектр процесса.*

Как уже отмечалось, при осуществлении передачи сигнала по каналу связи его точная форма часто заранее неизвестна. Поэтому реальный сигнал, поступающий в радиоканал от микрофона, передающей телевизионной камеры и других видов датчиков, можно в некотором приближении рассматривать как типичную реализацию из стационарного эргодического ансамбля. Если плотность вероятности такого случайного процесса известна (чаще всего ее считают близкой к нормальной), то единственная информация о всей совокупности возможных сигналов заключена в спектре мощности или в функции корреляции.

На практике часто применяемая аналитическая модель случайного процесса оказывается такой, что воспользоваться непосредственно определением дисперсии (3.50) для расчета спектра мощности не представляется возможным. Однако, если при этом удается вычислить корреляционную функцию случайного стационарного процесса, то получить спектральную информацию позволяет теорема Винера–Хинчина.

Физический смысл понятия спектра мощности легко выяснить, если принять в соотношении (3.55) $\tau = 0$:

$$R(0) = \sigma^2 = \frac{1}{2\pi}\int_{-\infty}^{\infty} W(\omega)d\omega. \qquad (3.57)$$

Итак, дисперсия $\sigma^2$ (индекс $x$ в обозначении СКО опущен), равная средней мощности флуктуаций стационарного эргодического случайного процесса, пропорциональна площади под кривой спектра мощности $W(\omega)$.

Размерность функции корреляции случайного процесса [$В^2$], а спектра мощности [($В^2 \cdot с$)/рад].

**Односторонний спектр мощности.** Поскольку $R(\tau)$ — четная функция аргумента $\tau$, то соответствующий спектр мощности $W(\omega)$ представляет собой также четную функцию частоты $\omega$. Отсюда следует, что пару преобразований Фурье (3.55) и (3.56) теоремы Винера–Хинчина можно записать, используя лишь интегралы в полубесконечных пределах, т. е.

$$R(\tau) = \frac{1}{\pi}\int_{0}^{\infty} W(\omega)\cos\omega\tau d\omega; \qquad (3.58)$$

$$W(\omega) = 2\int_{0}^{\infty} R(\tau)\cos\omega\tau d\tau. \qquad (3.59)$$

Для последующего анализа удобно ввести принятый ранее в радиолокации *односторонний спектр мощности* $F(\omega)$ случайного процесса, определив его следующим образом:

$$F(\omega) = \begin{cases} 0, & \omega < 0; \\ W(\omega)/\pi & \omega \geq 0. \end{cases} \qquad (3.60)$$

Функция одностороннего спектра мощности $F(\omega)$ позволяет вычислить дисперсию стационарного случайного процесса путем интегрирования по положительным (физическим) частотам

$$\sigma^2 = R(0) = \int_{0}^{\infty} F(\omega)d\omega. \qquad (3.61)$$

### Сравнение детерминированных и случайных сигналов

На основании проведенного в рамках спектральной теории анализа двух видов сигналов — детерминированного и стационарного эргодического случайного — можно провести их сравнение. Во-первых, к обоим сигналам подходят и спектральный и временно́й анализ. Во-вторых, базовыми соотношениями для детерминированного сигнала, устанавливающими связь между спектральной и временно́й характеристиками, являются прямое и обратное преобразования Фурье, определяемые (2.54), (2.55). Для случайного сигнала, относящегося к классу стационарных эргодических процессов, такими фундаментальными соотношениями, устанавливающими связь энергетического спектра и корреляционной функции, являются формулы Винера–Хинчина (3.56) и (3.55).

При внешнем сходстве формул преобразований Фурье (2.54), (2.55) и теоремы Винера–Хинчина (3.56), (3.55), надо иметь в виду их принципиальные отличия. Так, обратное преобразование Фурье (2.55) позволяет по спектральной плотности найти единственный, соответствующий ей импульсный сигнал. Определить же временну́ю форму случайного сигнала по спектру мощности по формуле Винера–Хинчина (3.56) невозможно. Это связано с тем, что в спектре мощности случайного процесса неизвестны начальные фазы спектральных составляющих. Одинаковый спектр мощности могут иметь совершенно разные детерминированный и случайный сигналы. Так, например, нормальный случайный процесс может иметь разнообразную спектральную плотность мощности, а белый шум — произвольную функцию распределения. Единственная «точка соприкосновения» вероятностных и корреляционных характеристик случайного процесса — это возможность расчета его дисперсии как исходя из одномерной плотности вероятности (3.11), так и на основе корреляционной функции (3.55).

Отмеченное демонстрируется на рис. 3.15: показаны временны́е реализации $x(t)$ двух случайных процессов (рис. 3.15, *а*, *б*), соответствующие им корреляционные функции $R(\tau)$ (рис. 3.15, *в*, *г*) и спектры мощностей $W(\omega)$ (рис. 3.15, *д*, *е*).

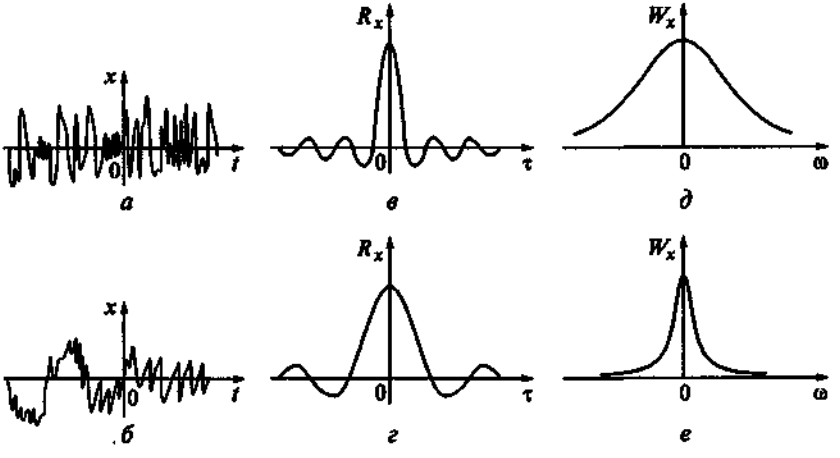



**Рис. 3.15.** Случайные процессы и взаимосвязь их характеристик:
*а*, *б* — реализации; *в*, *г* — корреляционные функции; *д*, *е* — спектры мощностей

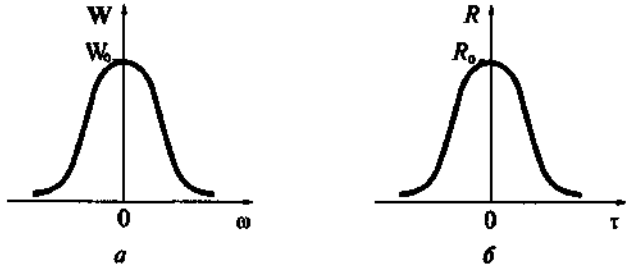



**Рис. 3.16.** Гауссовский характер параметров случайного процесса:
*а* — спектр мощности; *б* — функция корреляции

Следует отметить и такое различие между энергетическим спектром $W_и(\omega)$ детерминированного импульсного сигнала $u(t)$ и спектром мощности $W(\omega)$ стационарного случайного процесса $X(t)$. Если функция $W_и(\omega)$ характеризует распределение энергии импульсного сигнала по оси частот [$В^2 \cdot с$)/Гц], то функция $W(\omega)$ — распределение мощности случайного сигнала по оси частот [$В^2 \cdot с$)/рад]. Этот факт отражается и в разных физических размерностях энергетического спектра детерминированного сигнала $W_и(\omega)$ и спектра мощности случайного процесса $W(\omega)$.

**Пример 3.5.** Определить функцию корреляции стационарного случайного процесса, имеющего спектральную плотность мощности, распределенную по нормальному закону, причем при $\omega = 0$ — $W = W_0$ (рис. 3.16, *а*).

Р е ш е н и е. Поскольку спектр мощности описывается четной функцией, то в соответствии с (3.58) запишем

$$R(\tau) = \frac{W_0}{\pi} \int_{-\infty}^{\infty} e^{-\alpha^2\omega^2} \cos\omega\tau d\omega = \frac{W_0}{2\alpha\pi} e^{-\frac{\tau^2}{4\alpha^2}}.$$

Итак, при нормальном распределении спектра мощности случайного процесса функция корреляции имеет аналогичный характер распределения (рис. 3.16, *б*) и при $\tau = 0$

$$R(0) = \frac{W_0}{2\alpha\pi} e^{-\frac{\tau^2}{4\alpha^2}}.$$

**Пример 3.6.** Вычислить спектр мощности случайного стационарного процесса, функция корреляции которого равномерна на заданном интервале (рис. 3.17, *а*):

$$R(\tau) = \begin{cases} \sigma^2, & -\tau_x/2 \le \tau \le \tau_x/2; \\ 0, & \tau < -\tau_x/2, \quad \tau > \tau_x/2. \end{cases}$$

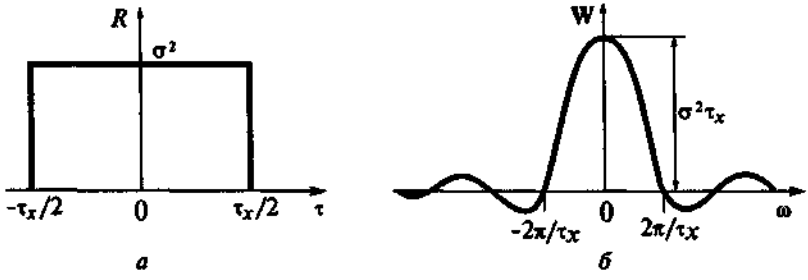



**Рис. 3.17.** Характеристики случайного процесса:
*а* — равномерная функция корреляции; *б* — спектр мощности

Р е ш е н и е. Используя формулу (3.59), получим

$$W(\omega) = 2\sigma^2 \int_0^{\infty} \cos\omega\tau\, d\tau = \sigma^2 \tau_x \frac{\sin(\omega\tau_x/2)}{\omega\tau_x/2}.$$

Спектр мощности показан на рис. 3.17, *б*, из которого следует, что его распределение совпадает со спектральной плотностью прямоугольного импульса.

### Белый шум

Для анализа радиотехнических цепей и сигналов большой практический интерес представляет специфический случайный процесс, теоретически имеющий равномерный спектр мощности во всей полосе частот:

$$W_x(\omega) = W_0 = \text{const}, \quad -\infty < \omega < \infty. \tag{3.62}$$

Как уже отмечалось, такой случайный процесс получил название «*белый шум*» по аналогии с применяемым в оптике понятием «белый свет», который содержит все цвета спектра с составляющими, имеющими практически одинаковую (равномерную) мощность. Белый шум представляет собой не реальный случайный процесс, а математическую абстракцию и идеализацию шумов. Он физически не реализуем и практически не может существовать в природе. Это объясняется, прежде всего, бесконечностью его дисперсии (т. е. бесконечностью средней мощности). Однако в тех случаях, когда полоса пропускания исследуемой системы существенно у́же эффективной ширины спектра шума, который на нее воздействует, можно для упрощения анализа приближенно заменить реальный случайный процесс белым шумом, который служит весьма удобной и полезной моделью. В технических расчетах вводят односторонний спектр мощности $N(f)$, представляющий среднюю мощность белого шума, приходящуюся на интервал частот в 1 Гц

$$N(f) = N_0 = \begin{cases} 0, & f < 0; \\ 2W(2\pi f) = 2W_0, & f \geq 0. \end{cases} \tag{3.63}$$

Нетрудно убедиться, что дисперсия белого шума в [$В^2$/Гц]

$$\sigma^2 = \int_0^\infty N_0 df. \tag{3.64}$$

Используя (3.55) и (3.62), определим функцию корреляции белого шума

$$R(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathbf{W}(\omega) e^{j\omega\tau} d\omega = \frac{W_0}{2\pi}\int_{-\infty}^{\infty} e^{j\omega\tau} d\omega = W_0\delta(\tau). \tag{3.65}$$

Таким образом, мгновенные значения напряжения (или мощности) белого шума в любые два сколь угодно близкие моменты времени *не коррелированны* (как бы ни был мал интервал $\tau$, сигнал за это время может измениться на любое значение). Поэтому теоретически интервал корреляции белого шума $\tau_к = 0$, а функция корреляции обращается в бесконечность. Временны́е значения белого шума не коррелированны, и поэтому его в радиотехнике и теории связи часто называют *дельта-коррелированным шумом*.

Поясним понятие «белый шум». По физическому смыслу спектральная плотность — мощность процесса, которая приходится на 1 Гц полосы частот. Тогда, если по равенству Парсеваля определить дисперсию белого шума, то получим значение $\sigma_x^2 = \infty$, что представляется абсурдным с физической точки зрения. Связано это с тем, что полученный результат является следствием абстрактного представления о бесконечной протяженности спектра мощности белого шума по оси частот. Как уже отмечалось, на практике ни один случайный процесс в действительности не может быть белым шумом. Тем не менее, многие помехи в радиотехнике, в технике связи и в других отраслях радиоэлектроники рассматривают как белый шум, если выполняется следующее соотношение между шириной спектров полезных сигналов $\Delta F_с$ и шумов $\Delta F_ш$

$$\Delta F_с / \Delta F_ш << 1,$$

и спектральная плотность шумов слабо изменяется в интервале спектра полезного сигнала.

На практике наблюдать белый шум можно только после того, как он пройдет через реальную радиотехническую систему, имеющую конечную ширину полосы рабочих частот. Тогда при обработке сигнала в приемнике так называемым *оптимальным фильтром* ширина спектра белого шума будет ограничиваться. Вследствие этого шум на выходе оптимального фильтра считают уже коррелированным. Следовательно, пока ширина полосы шума существенно больше ширины полосы обрабатывающей системы, можно утверждать, что шум имеет бесконечную ширину полосы. При выделении из белого шума с помощью фильтра относительно узкой полосы частот случайный процесс будет относиться к узкополосным случайным процессам. Принцип оптимальной фильтрации сигнала в шумах, будет представлен в главе 9.

## 3.5. Узкополосные случайные процессы

В радиотехнике и теории связи чрезвычайно важен особый класс случайных процессов, односторонний спектр мощности $F(\omega)$ (3.60) которых имеет *резко выраженный максимум* вблизи некоторой центральной частоты $\omega_0$, отличной от нуля. Как правило, функция $F(\omega)$ симметрична относительно частоты $\omega_0$, которая выбрана таким образом, что не содержит слагаемого, линейно зависящего от времени $t$. Такие сигналы относят к *узкополосным случайным процессам*. Их рассмотрение в учебнике ограничено стационарными и эргодическими случайными гауссовыми (нормальными и центрированными) процессами, наиболее часто встречающимися на практике. К тому же именно для гауссовых процессов удается получить ряд важных результатов, не выходя за рамки их корреляционной теории. Здесь необходимо также подчеркнуть, что нормальное распределение характеризует физическое колебание $x(t)$, т. е. мгновенное значение колебания в любой момент времени $t$.

Итак, *узкополосным* называется случайный процесс $X(t)$, односторонний спектр мощности $F(\omega)$ которого сосредоточен в полосе частот $\Delta\omega$ (ее часто называют *эффективной шириной спектра*), значение которой значительно меньше средней частоты $\omega_0$, т. е. $\Delta\omega << \omega_0$ (рис. 3.18). При этом реализации узкополосного случайного процесса представляют собой квазигармонические колебания, случайно модулированные по амплитуде и фазовому углу. Узкополосными случайными процессами являются модулированные сигналы, принятые и обрабатываемые в приемниках различных радиотехнических систем передачи информации. В качестве несущего колебания при этом служит гармоническое или узкополосное случайное колебание, а модулирующим сигналом может быть как детерминированный, так и случайный сигнал.

Для полного описания свойств узкополосного процесса необходимо знать законы распределения, а также корреляционные функции и все его параметры.

**Функция корреляции узкополосного случайного процесса.** Рассмотрим узкополосный стационарный случайный процесс $X(t)$ с односторонним спектром мощности $F(\omega)$. Узкополосный характер спектра $F(\omega)$ говорит о том, что корреляционная функция $R_x(\tau)$ должна также иметь вид узкополосного сигнала. Действительно, согласно теореме Винера–Хинчина (3.58), функция корреляции и односторонний спектр мощности случайного процесса связаны преобразованием Фурье. Для удобства мысленно сместим спектр случайного процесса $F(\omega)$ из окрестности частоты $\omega_0$ в окрестность нулевой частоты, выполнив для этого замену переменной $\omega = \omega_0 + \Omega$.

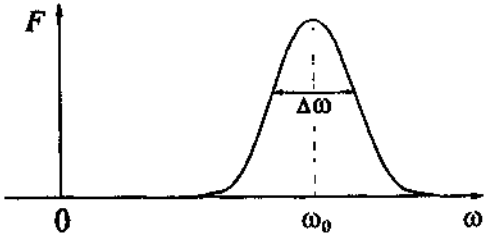



**Рис. 3.18.** Спектр мощности узкополосного случайного процесса

Тогда формула (3.58) при частоте $\omega_0 = \omega - \Omega$ примет вид

$$R(\tau) = \int_{-\omega_0}^{\infty} F(\omega_0 + \Omega)\cos[(\omega_0 + \Omega)\tau]d\Omega. \qquad (3.66)$$

Поскольку ранее было условлено, что спектр мощности $F(\omega)$ узкополосного процесса $X(t)$ ничтожно мал на частотах, близких к нулю, то в выражении (3.66) можно практически без заметной погрешности заменить нижний предел интегрирования $-\omega_0$ на $-\infty$. Тогда функцию корреляции можно записать в виде:

$$R(\tau) = a(\tau)\cos\omega_0\tau - b(\tau)\sin\omega_0\tau, \qquad (3.67)$$

где

$$a(\tau) = \int_{-\infty}^{\infty} F(\omega_0 + \Omega)\cos\Omega\tau d\Omega, \qquad (3.68)$$

$$b(\tau) = \int_{-\infty}^{\infty} F(\omega_0 + \Omega)\sin\Omega\tau d\Omega \qquad (3.69)$$

— медленно меняющиеся функции аргумента $\tau$, причем они представляют амплитуды соответственно косинусной и синусной составляющих реализации $x(t)$.

Если узкополосный случайный процесс обладает спектром мощности $F(\omega)$, *симметричным* относительно центральной частоты $\omega_0$, то функция корреляции имеет очень простой аналитический вид. Очевидно, что при этом в (3.67) «синусная» составляющая $b(\tau) = 0$ и функция корреляции

$$R(\tau) = a(\tau)\cos\omega_0\tau. \qquad (3.70)$$

В последнем соотношении параметр $a(\tau)$ играет роль огибающей функции корреляции, которая изменяется медленно во времени по сравнению с «быстрым» сомножителем $\cos\omega_0\tau$. На практике часто вводят нормированную огибающую $\rho(\tau)$ функции корреляции, определив ее через дисперсию (квадрат среднего квадратического отклонения) узкополосного случайного процесса $\sigma_x^2$ с помощью простого равенства $a(\tau) = \sigma_x^2\rho(\tau)$. Тогда функция корреляции типичного узкополосного случайного процесса (рис. 3.19) запишется в следующем виде

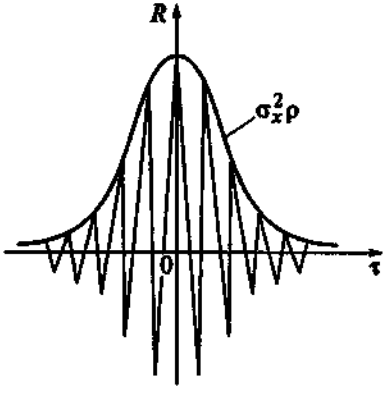



**Рис. 3.19.** Функция корреляции узкополосного процесса

$$R(\tau) = \sigma_x^2\rho(\tau)\cos\omega_0\tau. \qquad (3.71)$$

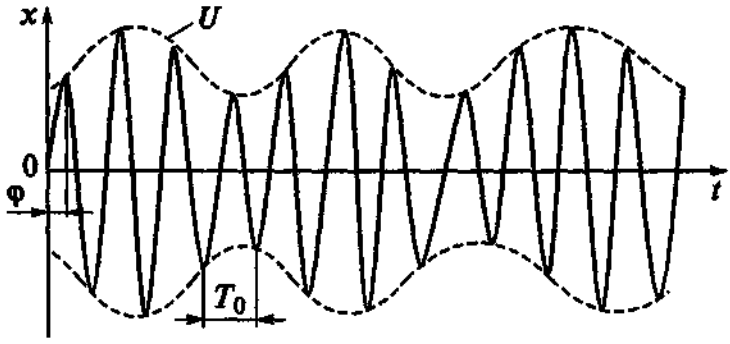



**Рис. 3.20.** Реализация узкополосного случайного процесса

**Огибающая и начальная фаза узкополосного случайного процесса.** Очевидно, что узкополосный спектр ($\Delta\omega << \omega_0$) и характерный, осциллирующий вид корреляционной функции (3.71), означают, что любые отдельные реализации узкополосного случайного процесса представляют собой *квазигармонические колебания* с периодом $T_0 = 2\pi/\omega_0$ (рис. 3.20)

$$x(t) = U(t)\cos[\omega_0 t + \varphi(t)], \tag{3.72}$$

у которых и огибающая $U(t)$ и начальная фаза $\varphi(t)$ являются *случайными* функциями, медленно (по сравнению с $\omega_0$ или $\cos\omega_0 t$) изменяющимися во времени. Заметим, что параметры любой реализации $U(t)$, $\varphi(t)$ и $\omega(t) = d\varphi(t)/dt$ случайного процесса могут подчиняться и не гауссовскому закону

Представим реализацию (3.72) как сумму синфазной и квадратурной амплитуд, точнее квадратурных колебаний или реализаций (2.210):

$$x(t) = A(t)\cos\omega_0 t - B(t)\sin\omega_0 t. \tag{3.73}$$

Очевидно, что колебания $A(t)$ и $B(t)$ являются медленно меняющимися функциями и тем они низкочастотнее, чем меньше эффективная ширина спектра $\Delta\omega$ процесса по сравнению с центральной частотой $\omega_0$.

Введем в рассмотрение узкополосный случайный процесс $Y(t)$, сопряженный с исходным процессом $X(t)$ (кстати, в этом смысле нет различия между случайным и детерминированным процессами). Его одиночной реализацией является реализация (2.229), сопряженная по Гильберту физическому колебанию $x(t)$

$$y(t) = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{x(\tau)}{t-\tau}d\tau. \tag{3.74}$$

Низкочастотность синфазной $A(t)$ и квадратурной $B(t)$ амплитуд позволяет сравнительно просто записать формулу для реализации сопряженного случайного процесса $Y(t)$, вынеся медленно меняющиеся множители по сравнению с частотой $\omega_0$ за знак преобразования Гильберта

$$y(t) = B(t)\cos\omega_0 t + A(t)\sin\omega_0 t. \tag{3.75}$$

Из (3.75) по аналогии с соотношениями (2.232) и (2.233) получаем формулы для мгновенных значений огибающей физической реализации $x(t)$

$$U(t)=\sqrt{x^2(t)+y^2(t)}=\sqrt{A^2(t)+B^2(t)} \tag{3.76}$$

и начальной фазы исходного случайного процесса $X(t)$

$$\varphi(t)=-\operatorname{arctg}\frac{B(t)}{A(t)}. \tag{3.77}$$

**Статистические свойства сопряженного узкополосного случайного процесса.** Чтобы изучить основные свойства огибающей и начальной фазы узкополосного случайного процесса, следует определить есть ли связь между статистическими характеристиками анализируемого $X(t)$ и сопряженного ему $Y(t)$ процессов? Для этого сначала определим математическое ожидание сопряженного случайного процесса, применив усреднение по времени к формуле (3.74):

$$\overline{y(t)}=\overline{\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{x(\tau)}{t-\tau}d\tau}=\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\overline{x(\tau)}}{t-\tau}d\tau=0. \tag{3.78}$$

Как и следовало ожидать, результат усреднения равен нулю, поскольку исходный процесс $X(t)$ является центрированным. Итак, если $\overline{x(t)}=0$, то и $\overline{y(t)}=0$. Далее, вследствие того что процесс $X(t)$ имеет нормальное распределение, а преобразование Гильберта есть линейное интегральное преобразование, то *нормальным* будет и сопряженный процесс $Y(t)$. Таким образом, реализации $x(t)$ и $y(t)$ в совпадающие моменты времени не коррелированны. Поскольку они, кроме того, имеют нормальное распределение, то из некоррелированности и следует их статистическая независимость.

Из свойств преобразования Гильберта (2.236) следует, что спектры конкретных реализаций процессов $x(t)$ и $y(t)$ связаны определенным образом. Как известно, если $S_x(\omega)$ — спектральная плотность исходной конкретной реализации $x(t)$, то спектральная плотность сопряженной реализации $y(t)$:

$$S_y(\omega)=\begin{cases} jS_x(\omega), & \omega<0;\\ 0, & \omega=0;\\ -jS_x(\omega), & \omega>0.\end{cases} \tag{3.79}$$

Иногда для удобства формулу (3.79) записывают в более компактном виде

$$S_y(\omega)=-jS_x(\omega)\operatorname{sign}(\omega). \tag{3.80}$$

Символом sign (сигнум-функция) обозначается функция

$$\operatorname{sign}t=\begin{cases}1, & \text{при } t>0;\\ -1, & \text{при } t<0.\end{cases} \tag{3.81}$$

Модули спектральных плотностей $S_x(\omega)$ и $S_y(\omega)$ совпадают, поэтому спектры мощности процессов $X(t)$ и $Y(t)$ одинаковы: $F_x(\omega) = F_y(\omega)$. Поскольку функции корреляции связаны со спектрами мощности обратным преобразованием Фурье, то они тоже должны быть одинаковыми, и согласно формуле (3.58) для одностороннего спектра (вместо $W(\omega)/\pi$ используем $F(\omega)$)

$$R_x(\tau) = R_y(\tau) = \int_0^\infty F_x(\omega)\cos\omega\tau d\omega. \tag{3.82}$$

Очевидно, что сопряженный процесс $Y(t)$ является также стационарным.

Определим функцию взаимной корреляции $R_{xy}(\tau)$ реализаций сопряженных узкополосных процессов $X(t)$ и $Y(t)$. Для этого заменим функцию $y(t)$ сопряженной функцией $x(t)$ по (3.74), введем новую переменную временно́го сдвига реализаций $\mu$, и затем, проведя несложные преобразования, получим

$$R_{xy}(\tau) = \overline{x(t)y(t+\tau)} = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\overline{x(t)x(\mu)}}{(t+\tau)-\mu}d\mu = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{R_x(\mu-t)}{\tau-(\mu-t)}d\mu = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{R_x(\nu)}{\tau-\nu}d\nu, \tag{3.83}$$

которая соответствует преобразованию Гильберта от функции корреляции исходного процесса $X(t)$. В формуле (3.83) переменная $\nu = \mu - t$.

Аналогично найдем функцию взаимной корреляции процесса $R_{yx}(\tau)$

$$R_{yx}(\tau) = -R_{xy}(\tau).$$

Таким образом, используя (2.231) и (3.58), получим

$$R_{xy}(\tau) = \mathrm{H}[R_x(\tau)] = \int_0^\infty F_x(\omega)\sin\omega\tau d\omega. \tag{3.84}$$

Нетрудно установить, что функция $R_{xy}(\tau)$ нечетна и обращается в нуль при отсутствии сдвига ($\tau = 0$). Поэтому процессы $X(t)$ и $Y(t)$ в совпадающие моменты времени статистически независимы (не коррелированны). Формулу (3.84) представляют в более удобной форме, выполнив замену переменной $\omega = \omega_0 + \Omega$:

$$R_{xy}(\tau) = \int_{-\omega_0}^\infty F(\omega_0+\Omega)\sin(\omega_0+\Omega)\tau d\Omega = a(\tau)\sin\omega_0\tau + b(\tau)\cos\omega_0\tau, \tag{3.85}$$

в котором функции $a(\tau)$ и $b(\tau)$ определяются в соответствии с (3.68) и (3.69).

**Вероятностные и корреляционные характеристики огибающей $U(t)$ и начальной фазы $\varphi(t)$ узкополосного процесса.** Для определения вероятностных характеристик перейдем от реализаций $x(t)$, $y(t)$ к низкочастотным синфазной $A(t)$ и квадратурной $B(t)$ функциям, которые найдем из (3.73) и (3.75). После несложных преобразований, получим

$$\begin{aligned} A(t) &= x(t)\cos\omega_0 t + y(t)\sin\omega_0 t, \\ B(t) &= -x(t)\sin\omega_0 t + y(t)\cos\omega_0 t. \end{aligned} \tag{3.86}$$

Вычислим функцию корреляции реализации $A(t)$, используя для этого первую формулу из системы (3.86). Выполнив элементарные преобразования по усреднению произведения, находим

$$R_A(\tau) = \overline{[x(t)\cos\omega_0 t + y(t)\sin\omega_0 t][x(t+\tau)\cos\omega_0(t+\tau) + y(t+\tau)\sin\omega_0(t+\tau)]} = \\ = R_x(\tau)\cos\omega_0\tau + R_{xy}(\tau)\sin\omega_0\tau. \qquad (3.87)$$

Подставив в это соотношение выражения функций $R_x(\tau)$ и $R_{xy}(\tau)$ из (3.67) и (3.85), приходим к очень простому результату

$$R_A(\tau) = a(\tau). \qquad (3.88)$$

Аналогично можно показать, что функции корреляции реализации $B(t)$ и взаимокорреляционная функция реализаций $A(t)$ и $B(t)$ соответственно равны

$$R_B(\tau) = a(\tau), \qquad (3.89)$$

$$R_{AB}(\tau) = b(\tau). \qquad (3.90)$$

Положив в (3.89) и (3.90) $\tau = 0$, имеем

$$\sigma_A^2 = \sigma_B^2 = \int_{-\infty}^{\infty} F_x(\omega)\, d\omega = \sigma_x^2. \qquad (3.91)$$

Таким образом, дисперсии синфазной $A(t)$ и квадратурной $B(t)$ низкочастотных колебаний оказываются равными дисперсии исходного узкополосного процесса $X(t)$.

**Двумерная (совместная) плотность вероятности огибающей и начальной фазы узкополосного случайного процесса.** Переход от реализации узкополосного случайного процесса $x(t)$ к ее синфазной $A(t)$ и квадратурной $B(t)$ составляющим существенно упрощает анализ, когда требуется вычислить двумерную (совместную) плотность вероятности $p(U, \varphi)$ огибающей $U(t)$ и начальной фазы $\varphi(t)$ узкополосного случайного процесса. Эта вероятностная характеристика, в свою очередь, дает возможность сравнительно легко найти одномерную плотность вероятности огибающей

$$p(U) = \int_0^{2\pi} p(U, \varphi)\, d\varphi \qquad (3.92)$$

и одномерную плотность вероятности начальной фазы

$$p(\varphi) = \int_0^{\infty} p(U, \varphi) dU. \qquad (3.93)$$

Мгновенное значение комплексного случайного процесса $X(t)$ можно отразить графически с помощью низкочастотных синфазной $A(t)$ и квадратурной $B(t)$

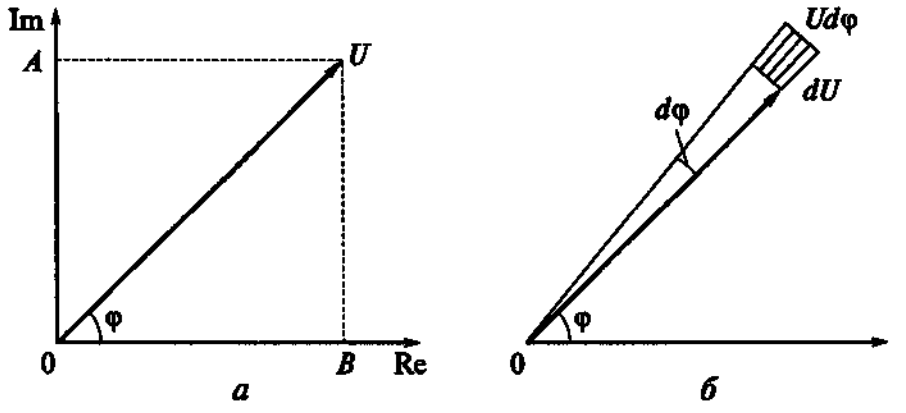



**Рис. 3.21.** К определению двумерной плотности вероятностей комплексной огибающей: *а* — квадратурных составляющих; *б* — модуля и аргумента

колебаний (рис. 3.21). Это позволяет сделать предположение о симметричности спектра мощности случайного процесса $F(\omega)$ и обращении в нуль амплитуды $b(\tau)$ синусной составляющей реализации $x(t)$.

Мгновенные значения амплитуд реализаций $A(t)$ и $B(t)$ образуют на комплексной плоскости двумерный гауссов вектор огибающей $U(t)$ (рис. 3.21, *а*), обе составляющие которого независимы и имеют одинаковые дисперсии $\sigma^2$. Поэтому совместная (двумерная) плотность вероятности квадратурных составляющих реализации случайного процесса $x(t)$

$$p(A, B) = p(A)p(B) = \frac{1}{2\pi\sigma_x^2}\exp\left[-\frac{A^2 + B^2}{2\sigma_x^2}\right].$$

Так как в новых переменных $A^2 + B^2 = U^2$, то искомая двумерная плотность вероятности

$$p(A, B) = \frac{1}{2\pi\sigma_x^2}\exp\left[-\frac{U^2}{2\sigma_x^2}\right]. \qquad (3.94)$$

Для определения статистических свойств огибающей и фазы следует в выражении (3.94) выполнить функциональное преобразование, переводящее случайный вектор огибающей $U(t)$ из декартовой $[A, B]$ в полярную $[U, \varphi]$ систему координат (см. рис. 3.21, *б*), и определить совместную плотность вероятности огибающей и начальной фазы $p(U, \varphi)$.

Связь между двумя представленными координатными системами выражается следующими формулами:

$$A = U\cos\varphi; \qquad B = U\sin\varphi. \qquad (3.95)$$

При смене системы координат вероятность попадания в бесконечно малую область в окрестностях каждой точки комплексной плоскости должна остаться

неизменной (площадь заштрихованного элемента на рис. 3.21, *б*). Площадь этой бесконечно малой области в декартовых координатах выражается как $dAdB$, а в полярных — как $UdUd\varphi$ (см. рис. 3.21).

Таким образом, нетрудно установить, что

$$p(U,\varphi)=Up(U)p(\varphi)=\frac{U}{2\pi\sigma_x^2}\exp\left[-\frac{A^2+B^2}{2\pi\sigma_x^2}\right]. \qquad (3.96)$$

Как уже отмечалось, в новых переменных $A^2 + B^2 = U^2$, поэтому двумерная плотность вероятности огибающей и начальной фазы узкополосного процесса (3.96) запишется в виде

$$p(U,\varphi)=\frac{U}{2\pi\sigma_x^2}\exp\left[-\frac{U^2}{2\sigma_x^2}\right]. \qquad (3.97)$$

**Одномерная плотность вероятности огибающей узкополосного случайного процесса.** Воспользовавшись формулами (3.92) и (3.97), найдем плотность вероятности огибающей

$$p(U)=\int_0^{2\pi}p(U,\varphi)d\varphi=\frac{U}{2\pi\sigma_x^2}\int_0^{2\pi}\exp\left[-\frac{U^2}{2\sigma_x^2}\right]d\varphi. \qquad (3.98)$$

Поскольку $p(U, \varphi)$ не зависит от начальной фазы $\varphi = \varphi(t)$ и $\int_0^{2\pi}d\varphi=2\pi$, то формула (3.98) для плотности вероятности огибающей $U = U(t)$ упрощается

$$p(U)=\frac{1}{2\pi}\int_0^{2\pi}\frac{U}{\sigma_x^2}\exp\left(-\frac{U^2}{2\sigma_x^2}\right)d\varphi=\frac{2\pi}{2\pi}\frac{U}{\sigma_x^2}\exp\left(-\frac{U^2}{2\sigma_x^2}\right)=\frac{U}{\sigma_x^2}\exp\left(-\frac{U^2}{2\sigma_x^2}\right). \qquad (3.99)$$

Плотность вероятности мгновенных значений огибающей узкополосного случайного процесса, описываемая законом (3.99), известна в математике под названием *закона* (*распределения*) *Рэлея*. Приведенный на рис. 3.22 график плотности вероятности огибающей наглядно показывает, что наиболее вероятны некоторые средние (порядка $\sigma_x$) значения огибающей. В то же время маловероятно, чтобы огибающая принимала значения как близкие к нулю, так и значительно превосходящие среднеквадратичный уровень $\sigma_x$ процесса. Максимальное значение функции $p(z)$ получается при $U = \sigma_x$. Это означает, что $U = \sigma_x$ является наивероятнейшим значением огибающей узкополосного случайного процесса.

Среднее же значение (математическое ожидание) огибающей узкополосного случайного процесса находим известным в математике способом с помощью плотности вероятности (3.99)

$$m_U=\overline{U}=\int_0^{\infty}Up(U)dU=\int_0^{\infty}U\frac{U}{\sigma_x^2}\exp\left[-\frac{U^2}{2\sigma_x^2}\right]dU=\sqrt{\frac{\pi}{2}}\,\sigma_x=1{,}253\sigma_x. \qquad (3.100)$$

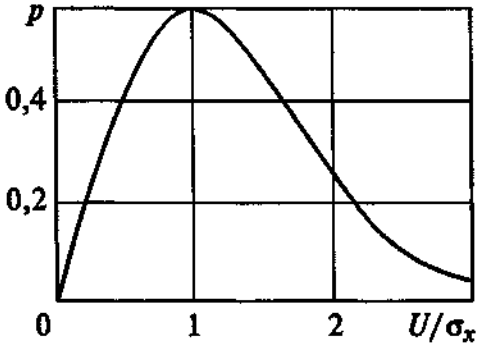



**Рис. 3.22.** График плотности вероятности огибающей, распределенной по закону Рэлея

Аналогично определяем ее дисперсию

$$\sigma_U^2 = \overline{U^2} = \int_0^\infty U^2 p(U)dU = \int_0^\infty \frac{U^3}{\sigma_x^2}\exp\left(\frac{-U^3}{2\sigma_x^2}\right)dU = \left(2-\sqrt{\frac{\pi}{2}}\right)\sigma_x^2 = 0{,}429\sigma_x^2. \quad (3.101)$$

Если случайный процесс центрированный ($\overline{U}=0$), то из (3.101) получим

$$\sigma_U^2 = 2\sigma_x^2. \quad (3.102)$$

Этот результат означает, что дисперсия огибающей равна удвоенной дисперсии узкополосного процесса. Это аналогично соотношению между квадратом амплитуды $U_0^2$ и средней мощностью гармонического колебания $u(t) = U_0\cos\omega_0 t$, равной $\overline{u^2(t)} = 0{,}5U_0^2$ (см. формулу (2.122) для мощности несущей АМ-сигнала).

Располагая одномерной плотностью вероятности огибающей, можно решить ряд задач теории узкополосных случайных процессов, в частности, найти вероятность превышения огибающей заданного уровня. Такой анализ проводят при исследовании процессов детектирования сигналов с шумами в приемниках.

**Одномерная плотность вероятности начальной фазы узкополосного случайного процесса.** Обратившись к (3.89) и (3.93), найдем плотность вероятности начальной фазы, вычислив интеграл по $U$

$$p(\varphi) = \int_0^\infty p(U,\varphi)\,dU = \int_0^\infty \frac{U}{2\pi\sigma_x^2}\exp\left[-\frac{U^2}{2\sigma_x^2}\right]dU = \frac{1}{2\pi}. \quad (3.103)$$

Значит, начальная фаза узкополосного случайного процесса распределена равномерно на отрезке [0, 2л]. Физически равномерное распределение означает отсутствие какого-либо преимущественного значения начальной фазы у отдельно взятых реализаций. Итак, мгновенное значение реализации узкополосного случайного процесса $u(t)$ и начальная фаза $\varphi(t)$ подчиняются нормальному закону, а его огибающая (амплитуда) $U(t)$ — закону Рэлея. Такой шум, образованный из белого шума, один из основных видов помех в радиоэлектронных цепях.

**Огибающая суммы гармонического сигнала и узкополосного гауссова шума.** В радиотехнике часто требуется оценить статистические свойства случайного колебания $X(t)$, имеющего место на выходе частотно-избирательного устройства, например резонансного усилителя. Пусть наряду с узкополосным нормальным шумом с центральной частотой $\omega_0$, равной резонансной частоте усилителя, на его выходе имеется детерминированный сигнал — гармоническое колебание $u(t) = U_m\cos\omega_0 t$ с амплитудой $U_m$. Наиболее просто в этом случае находится одномерная плотность вероятности огибающей суммарного колебания. Считая, что шум $n(t) = A(t)\cos\omega_0 t - B(t)\sin\omega_0 t$, запишем выражение реализации суммарного случайного процесса $X(t)$

$$x(t) = u(t) + n(t) = U_m\cos\omega_0 t + A(t)\cos\omega_0 t - B(t)\sin\omega_0 t = [U_m + A(t)]\cos\omega_0 t - B(t)\sin\omega_0 t.$$

В силу узкополосности процесса $x(t)$ любую реализацию можно выразить через медленно меняющиеся огибающую $U(t)$ и начальную фазу $\varphi(t)$ процесса

$$x(t) = U(t)[\cos\omega_0 t + \varphi(t)].$$

Из двух последних формул нетрудно установить, что между парами $\{A, B\}$ и $\{U, \varphi\}$ имеется такая связь:

$$A(t) = U(t)\cos\varphi(t) - U_m;$$
$$B(t) = U(t)\sin\varphi(t). \tag{3.104}$$

Учитывая двумерную плотность вероятности (3.94) и соотношения для квадратурных составляющих (3.103), получим в новых переменных при переходе от декартовой системы координат к полярной

$$p(U,\varphi) = \frac{U}{2\pi\sigma_x^2}\exp\left[-\frac{A^2(t)+B^2(t)}{2\sigma_x^2}\right] = \frac{U}{2\pi\sigma_x^2}\exp\left[-\frac{U^2+U_m^2-2UU_m\cos\varphi}{2\sigma_x^2}\right]. \tag{3.105}$$

Чтобы получить одномерную плотность вероятности огибающей (3.92)

$$p(U) = \int_0^{2\pi} p(U,\varphi)\,d\varphi,$$

следует проинтегрировать правую часть (3.105) по угловой координате $\varphi$.

Для упрощения выкладок применим известную в математике формулу

$$p(U) = \frac{1}{2\pi}\int_0^{2\pi} e^{x\cos\varphi}\,d\varphi = I_0(x),$$

где $I_0$ — модифицированная функция Бесселя первого рода нулевого порядка, в результате чего получим

$$p(U) = \frac{U}{\sigma_x^2}\exp\left[-\frac{U^2+U_m^2}{2\sigma_x^2}\right] I_0\left(\frac{UU_m}{\sigma_x^2}\right). \tag{3.106}$$

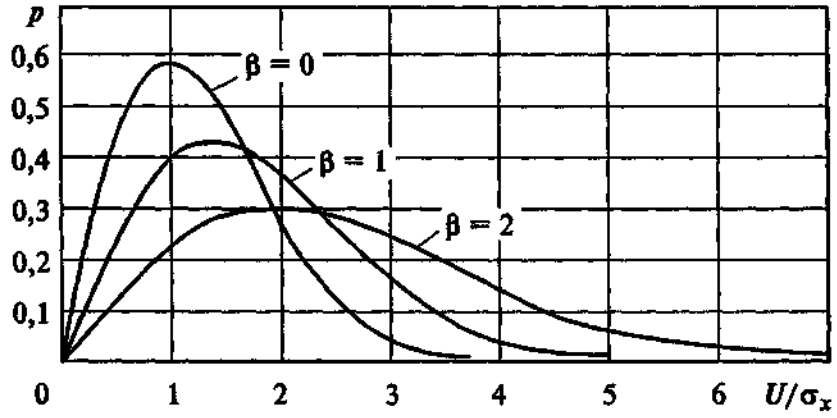



**Рис. 3.23.** Графики плотности вероятности случайной величины, распределенной по закону Райса

Формула (3.106) выражает закон, получивший в статистической радиотехнике название *закона Райса.*

На рис. 3.23 представлены графики плотности вероятности распределения случайной величины (в данном случае огибающей случайного процесса), распределенной по закону Райса при различных отношениях $\beta = U_m/\sigma_x$. Отметим, что при $U_m = 0$, т. е. в отсутствие детерминированного сигнала, закон Райса (3.107) переходит в закон Рэлея (3.99).

Отметим, что если амплитуда детерминированного (в данном случае гармонического) сигнала значительно превышает среднеквадратический уровень шума, т. е. $U_m/\sigma_x >> 1$, то при $U \approx U_m$ можно воспользоваться известным в математике асимптотическим представлением модифицированных функций Бесселя с большим аргументом.

Теперь нетрудно показать, что при этом соотношение (3.103) примет следующий вид

$$p(U) = \frac{U}{\sqrt{2\pi}\,\sigma_x} \exp\left[-\frac{(U-U_m)^2}{2\sigma_x^2}\right], \qquad (3.107)$$

т. е. огибающая результирующего сигнала распределена в этом случае приближенно нормально с дисперсией $\sigma_x^2$ и математическим ожиданием, равным амплитуде детерминированного сигнала $U_m$.

В практических случаях считают, что уже при соотношении $U_m/\sigma_x = 3$ огибающая суммарного колебания начинает подчиняться нормальному закону. Полезно вспомнить, что огибающая чистого шума, распределенная по закону Рэлея (3.101), имеет дисперсию, равную $\sigma_U^2 = 0{,}429\sigma_x^2$. Таким образом, аддитивное (суммарное) наложение в радиоэлектронной цепи на шум достаточно большого гармонического сигнала приводит более чем к двукратному росту дисперсии суммарной огибающей.

## 3.6. Шумоподобные сигналы

Необычно важными достижениями ученых и инженеров XX в. явились создание уже упоминавшихся в главе 1 широкополосных, или шумоподобных сигналов, и разработка на их основе новых систем радиолокации и радиосвязи. Широкое использование ШПС для создания радиосистем массового применения началось в последние 10–25 лет. Однако к использованию этих идей для повышения помехоустойчивости приема приступили еще в 40-х гг. XX в.

В обычных видах модуляции информационный поток изменяет амплитуду, частоту или фазу несущего гармонического колебания. При этом ширина спектра излучаемого в эфир радиосигнала соизмерима с шириной спектра модулирующего сигнала. Однако теперь переносчиком (несущим колебанием) информации стал не только гармонический, но и сложный широкополосный сигнал. Такой сигнал может быть сформирован различным образом: несущая может быть модулирована вспомогательной кодовой последовательностью или вспомогательным аналоговым сигналом по амплитуде, частоте или фазе.

Шумоподобные сигналы в отечественной литературе часто называют *широкополосными сигналами, сигналами без несущей*. Их корреляционные функции и спектры плотности мощности близки к аналогичным характеристикам квазибелого шума. Особое свойство ШПС заключается в том, что они относятся к классу сложных сигналов, база которых $B_c = 2F_cT_c >> 1$, где $F$ — полоса частот, а $T_c$ — длительность информационного (передаваемого) сигнала. В частности, в системах связи стандарта CDMA, использующих ШПС, поступающие на вход приемника помехи, подавляются в $B = W/F$ раз (обычно в ШПС $B$ = 100...10 000). Данное свойство ШПС является уникальным и позволяет многократно и одновременно использовать один и тот же частотный канал для связи разных абонентов на ограниченной территории.

В традиционных системах связи для исключения возможности возникновения помех между зонами, в которых используют один и тот же частотный канал, необходимо обеспечить значительный территориальный разнос. Таким образом, в системах CDMA достигается весьма высокая эффективность использования частотного спектра. Кроме того, ШПС позволяют путем специальной обработки принимаемых сигналов эффективно бороться с их замираниями в многолучевых каналах, разделяя отдельные лучи и осуществляя их когерентное сложение.

В 1953 г. в США был опубликован фундаментальный труд Ф.М. Вудворта «Теория вероятностей и теория информации с применениями в радиолокации», в котором автором был предложен метод создания сложных (шумоподобных) сигналов. Основополагающие работы, связанные с ШПС, были направлены на разработку методов радиоэлектронной борьбы с мощными активными радиотехническими помехами, мешающими приему радиолокационных сигналов, отраженных от целей. В последующие годы были разработаны и нашли широкое применение, в основном в системах радиосвязи, новые способы формирования

### *Широкополосные методы модуляции*

Первые идеи построения ШПС были связаны с частотной модуляцией несущей частоты вспомогательным сигналом, структура которого должна была быть известна на приеме. В 1942 г. американской актрисой Х. Ламар и ее мужем композитором Г. Атсейлу было сделано пионерское изобретение — способ формирования ШПС методом скачкообразного изменения частоты несущего колебания за время передачи одного информационного символа (*FH-SS — Frequency Hopping Spread Spectrum*). В отечественной литературе такие сигналы называют сигналами с *частотно-временной матрицей* (ЧВМ). Другим методом формирования ШПС является метод скачкообразного изменения относительного временно́го положения коротких кодовых импульсов за время передачи одного информационного символа (*TH-SS — Time Hopping Spread Spectrum*).

Один из широко применяемых сегодня способов формирования ШПС, который называется *методом прямого расширения спектра* путем непосредственной фазовой модуляции несущей кодовой последовательностью (*DS-SS — Direct Sequence Spread Spectrum*), был изобретен американскими инженерами Дж. Грином и М. Никольсоном в 1957 г. Ими был предложен метод построения бинарной кодовой последовательности.

Во всех отмеченных случаях возможно создание в общей полосе частот больших ансамблей сигналов, которые отличаются либо законом чередования фазы несущего колебания для сигналов *DS-SS*, либо законом изменения несущей частоты для сигналов *FH-SS*, либо определенной временно́й расстановкой коротких импульсов для сигналов *TH-SS*. Таким образом, ШПС могут использоваться в качестве переносчиков информации подобно тому, как используются гармонические колебания в обычных системах связи. При этом аналогом АМ является передача одного из ШПС, принадлежащего к определенному ансамблю сигналов, с заданной амплитудой, аналогом ФМн является манипуляция фазы ШПС, а аналогом обычной ЧМн является передача одного из двух возможных сигналов ШПС. В 1958 г. была создана первая система коротковолновой связи «Рейк» для работы в многолучевом канале, в которой ШПС применялись для разделения отдельных лучей и устранения замираний, вызванных их интерференцией. В 1963 г. была создана наземная система связи *RACEP* (*Random Access and Correlation for Extended Perfomance*), в которой для передачи сообщений применялась ФИМ. Система занимала полосу частот 4 МГц и работала в диапазоне 140 МГц. Она позволяла осуществлять передачу сигналов телефонии и цифровой информации и давала возможность организации на одной территории сети связи с емкостью 700 абонентов.

С 1963 г. на основе ШПС начинают создаваться спутниковые системы связи со свободным доступом к общему каналу связи, тропосферные радиорелейные системы связи с разделением отдельных лучей. Исследования эффективности применения ШПС в сравнении с другими методами модуляции в системах связи различных назначений начались с 1965 г. В 90-х гг. системы с ШПС начинают внедряться в системы сотовой подвижной связи.

Подобные системы будут применяться в сотовых системах подвижной связи, широкое внедрение которых начнется в XXI веке. Исследования вопросов эффективности использования ШПС в таких системах и разработка методов их частотного планирования были начаты российским ученым Л.Е. Варакиным.

**Расширенный спектр (*Spread Spectrum*).** Предположим, что создан узкополосный сигнал, промодулированный потоком данных с заданной скоростью. Пусть есть повторяющаяся, псевдослучайная цифровая последовательность (код) со значительно большей скоростью. Если менять фазу узкополосного сигнала в соответствии с псевдослучайной последовательностью, то получим ШПС с широким спектром, содержащим передаваемую информацию. Если рассмотреть, что происходит с точки зрения частоты — получится, что информационный сигнал «расплылся» (*spread*) по спектру ШПС. Теперь осталось передать этот широкополосный. Принятый и демодулированный сигнал перемножим с копией ШПС, который использовали для модуляции (здесь необходима высокая степень синхронизации приемника и передатчика) и получим узкополосную составляющую с высокой энергией на единицу частоты — переданный поток данных. Поскольку помехи и сигналы от других передатчиков не совпадают с используемым ШПС, то после перемножения они еще больше расползутся по спектру и их энергия на единицу частоты уменьшится. Таким образом, используя разные псевдослучайные последовательности (коды), можно организовать несколько независимых каналов передачи данных в одной и той же полосе частот. Приведенное описание технологии создания ШПС сильно упрощено.

Эффект первый. При перемещении точки приема эти условия периодически изменяются, так как сложение волн, приходящих с разных направлений, создает пространственную интерференционную картину. Эффект особенно сильно сказывается на мобильных пользователях и типичен для узкополосных систем. Для систем с *Spread Spectrum* действие эффекта ослабляется, поскольку на разных частотах в пределах его широкого спектра создаются разные интерференционные картины, что выравнивает результирующий сигнал.

Эффект второй. Время задержки сигналов при прохождении больших расстояний может меняться из-за изменений характеристик среды распространения, причем это сказывается по-разному на сигналах, приходящих разными путями, что при сложении вызывает временные флуктуации уровня (громкости) сигнала, называемые в радиосвязи замираниями. Несинфазное изменение уровня сигнала на разных частотах спектра *Spread Spectrum* сигнала приводит к сильному ослаблению действия этого эффекта на подобные системы.

Отметим три особенности шумоподобного сигнала:

- ширина спектра намного превышает базу передаваемого сигнала ($B_c >> 1$);
- при формировании используется фактор псевдослучайности, благодаря чему по своей структуре он становится близким к случайному сигналу;
- распознать его при приеме можно, только зная «ключ» кодирования при передаче, что обеспечивает скрытность связи. Это реализуется «размазыванием» передаваемого сигнала по спектру и требованием, чтобы приемник «знал» как его «собирать».

В настоящее время в теоретической радиотехнике, теории информации и прикладной математике усиленно разрабатывают методы синтеза сигналов с заданными (оптимальными) автокорреляционными и спектральными свойствами. Если рассматривать последовательности из $n$ импульсов прямоугольной формы, которые в соответствии с номером позиции $M$ могут принимать значения $\pm 1$, то простым перебором можно найти последовательности, для которых

$$\frac{B(0)}{E_1} = n; \quad E = nE_1, \tag{3.108}$$

где $B(0)$ — АКФ; $E$ — энергия всего сигнала; $E_1$ — энергия одного элемента.

В основе структуры шумоподобного сигнала и его анализа лежит понятие «секвента» (*Sequential*). Сигнал вида секвента имеет одинаковое число переходов через нуль. Обычно под секвентными сигналами понимается последовательность разнополярных импульсов одинаковой формы и единичной амплитуды. Структура секвентных сигналов хорошо приспособлена для современных цифровых систем связи, особенно это касается мобильных систем. Поэтому при анализе сигнала в секвентном базисе получают разложение сигнала на те разнополярные компоненты, которые имеют место в действительности.

Шумоподобный сигнал формируют по-разному: несущая может быть модулирована вспомогательной кодовой последовательностью или вспомогательным аналоговым сигналом по фазе, частоте или амплитуде. Наиболее распространенным примером технической реализации шумоподобных сигналов служат сформированные определенным образом псевдослучайные последовательности прямоугольных радиоимпульсов длительностью $\tau$, в частности, при манипуляции несущего колебания двоичными цифровыми кодами. База таких сигналов определяется числом модулирующих импульсов в исходной последовательности. При этом наиболее успешно развиваются цифровые методы передачи сигналов, созданные на основе дискретных ортогональных последовательностей в виде функций Баркера, линейных рекуррентных $M$-последовательностей (последовательности Голда, Лежандра и пр.), Радемахера, Уолша и др.

### Сигналы (коды) Баркера

В 50–60-е гг. XX в. были найдены целые классы широкополосных сигналов с совершенными корреляционными свойствами. Среди них большую известность получили *сигналы* (*коды*, *последовательности*) *Баркера*. Эти сигналы обладают уникальным свойством — независимо от номера позиции $M = n$ в этих последовательностях значения их АКФ при всех $n \neq 0$ не превышают единицы (т. е. абсолютное значение «остатков» — боковых лепестков — не превышает $1/n$ основного). В то же время энергия всех этих сигналов численно равна $n$. Модели сигналов Баркера в настоящее время удается реализовать лишь при числе позиций $M = 2, 3, 4, 5, 7, 11$ и $13$ (табл. 3.5). Случаи $M = 2$ и $M = 3$ являются тривиальными. Исследования показали, что пока не найдены модели сигналов Баркера с нечетным числом позиций, большим 13. Поэтому для $M > 13$ при-

Таблица 3.5. **Модели сигналов Баркера**

| $M$ | Модель сигнала | АКФ |
|---|---|---|
| 2 | 1, –1 | 2; –1 |
| 3 | 1, 1, –1 | 3; 0; –1 |
| 4 | 1, 1, –1, 1 | 4; 1; 0; –1 |
| 5 | 1, 1, 1, –1, 1 | 5; 0; 1; 0; 1 |
| 7 | 1, 1, 1, –1, –1, 1, –1 | 7; 0; –1; 0; –1; 0; –1 |
| 11 | 1, 1, 1, –1, –1, –1, 1, –1, –1, 1, –1 | 11; 0; –1; 0; –1; 0; –1; 0; –1; 0; –1 |
| 13 | 1, 1, 1, 1, 1, –1, –1, 1, 1, –1, 1, –1, 1 | 13; 0; 1; 0; 1; 0; 1; 0; 1; 0; 1; 0; 1; 0; 1 |

ходится довольствоваться последовательностями, имеющими «остатки» АКФ большее, чем $1/M$. При этом до сих пор неизвестно, можно ли построить сигнал Баркера с четным числом позиций $M > 4$.

На рис. 3.24 приведен вид 13-позиционного сигнала Баркера в виде биполярных импульсов, его фазовое кодирование и АКФ.

Прием последовательности Баркера выполняет *согласованный цифровой фильтр* (см. гл. 8). Импульсы последовательности Баркера с числом знаков $M = n = 13$ поступают сначала на фильтр $СФ_{ПИ}$, согласованный с единичным прямоугольным импульсом, а затем в линию задержки, имеющую отводы через промежутки времени $\Delta = \tau$, далее на фазоинверсные (–) и фазосохраняющие (+) каскады с одинаковыми коэффициентами передачи, суммирующее устройство и решающее устройство (РУ) (рис. 3.25). Фазоинверсные и фазосохраняющие каскады включены в порядке, соответствующем обратному порядку чередования импульсов последовательности (рис. 3.24, *а*): число этих каскадов равно числу ее элементов, т. е. 13. Первый каскад включен до линии задержки (ЛЗ), последний — на ее конце.

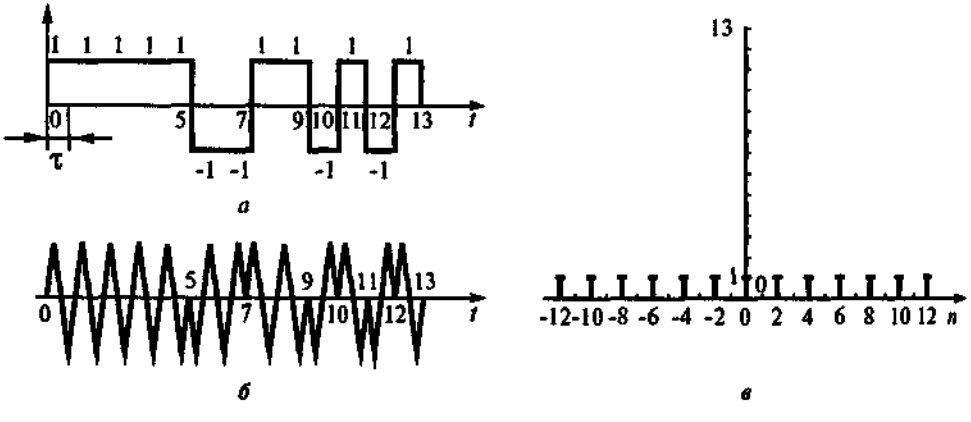



**Рис. 3.24.** Сигнал Баркера при $M = 13$:
*а* — амплитудное кодирование; *б* — фазовое кодирование; *в* — АКФ

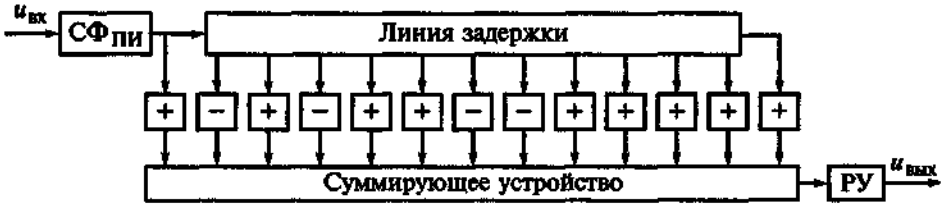



**Рис. 3.25.** Согласованный фильтр для последовательности Баркера при $M = 13$

При приеме последовательность Баркера движется по ЛЗ, и в момент, когда импульсы последовательности совпадут по знаку с «весами», включенными между отводами ЛЗ и суммирующим устройством, все импульсы сложатся синфазно, на выходе РУ появится наибольший импульс — согласованный фильтр зафиксирует адрес первого канала. При всех других сдвигах суммирование производится не в фазе (с разными знаками), и на выходе РУ появляются уровни, не превышающие по модулю $1/n = 1/13$ от максимального значения. Поскольку функции взаимной корреляции между последовательностями также имеют наибольшие значения, не превышающие $1/n$, то последовательность адреса чужого канала не может вызвать ложного срабатывания решающего устройства первого канала.

### Линейные *M*-последовательности

Несколько худшие автокорреляционные функции по сравнению с сигналами Баркера имеют линейные рекуррентные *M*-последовательности. Отметим, что *последовательность* — это половина среднего числа переходов через нуль за единицу времени. Для линейной *M*-последовательности отношение значения главного максимума к максимальному значению бокового лепестка автокорреляционной функции приближенно растет как $\sqrt{n}$, где $n = 2^{\mu} - 1$ — число импульсов в последовательности; $\mu$ — целое положительное число.

В определенной степени линейные рекуррентные последовательности обладают свойством хаотичности, которое заключается в следующем. Если из периода следования такой последовательности, содержащей $n$ членов, выбрать возможные отрезки по $\mu$ членов в каждом, то:

- среди этих отрезков не будет совпадающих;
- среди них найдутся любые комбинации из +1 и −1, состоящие из $\mu$ членов (кроме запрещенной комбинации, включающей только +1).

Эти свойства сходны со свойствами случайных разнополярных последовательностей, поэтому *M*-последовательности часто называют псевдослучайными (*pseudorandom*) или шумоподобными (*pseudonoise*) последовательностями. К тому же автокорреляционная функция *M*-последовательности имеет форму, сходную с АКФ квазибелого шума с ограниченным спектром. При передаче псевдослучайных последовательностей по высокочастотному каналу обычно применяется фазовая или относительная фазовая модуляция.

### Сигналы на основе функций Радемахера

Аналитическая интерпретация этих функций встречает определенные сложности, однако идея их построения сравнительно проста. Функции Радемахера образуются из синусоидальных функций с помощью соотношения

$$r_k(\theta) = \text{sign}[\sin(2^k\pi\theta)], \quad 0 \le \theta \le 1, \tag{3.109}$$

где аргумент $\theta = t/T$ — безразмерное время; $T$ — период функции; $k = 0, 1, 2, \ldots$ — порядок функции.

Функции Радемахера, принимающие значения $\pm 1$, можно трактовать как функции «прямоугольного синуса», и они имеют форму меандра. На рис. 3.26 показаны временны́е диаграммы первых четырех функций Радемахера.

Нетрудно заметить, что функции Радемахера $r(\theta)$

$$\int_0^1 r_i(\theta) r_k(\theta) d\theta = \begin{cases} 1, & i = k; \\ 0, & i \ne k, \end{cases} \tag{3.110}$$

ортонормированны на интервале $0 \le \theta \le 1$.

### Сигналы на основе функций Уолша

Дальнейшим развитием систем функций типа последовательностей Радемахера, имеющих форму «прямоугольной волны», стала ортонормированная система *функций* (последовательностей) *Уолша* (*Walsh*). Широко используемые в технике связи функции Уолша — одни из немногих известных в математике с 1922 г. ортонормированных функций, которые на отрезке своего существования (периоде следования $-T/2$, $T/2$) принимают лишь два значения: +1 или –1.

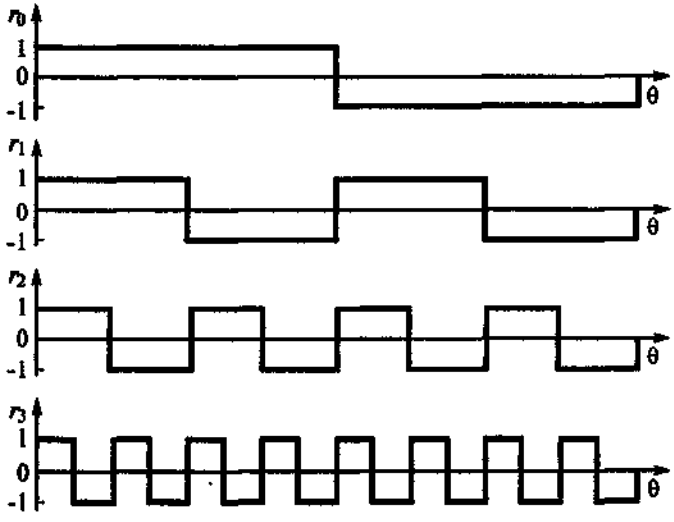



**Рис. 3.26.** Временны́е диаграммы первых четырех функций Радемахера

Введем безразмерное время, нормированное к периоду следования $\theta = t/T$, и обозначим $k$-ю ($k$ — порядок) функцию Уолша принятым в математике символом $\mathrm{wal}(k, \theta)$. Идею построения последовательностей Уолша можно уяснить из рис. 3.27, где показаны первые восемь этих функций. Ортонормированность данных функций следует из принципа их построения и может быть проверена непосредственно путем вычисления интеграла. Система $\mathrm{wal}(k, \theta)$ позволяет продемонстрировать наиболее важные свойства функций Уолша, которые определяются следующим разностным (рекуррентным) уравнением

$$\mathrm{wal}(2k+p, \theta) = (-1)^{[k/2]+p}[\mathrm{wal}(k, 2\theta + 0{,}5) + (-1)^{k+p}\mathrm{wal}(k, 2\theta - 0{,}5)], \qquad (3.111)$$

где обозначение $[k/2]$ представляет собой наибольшее целое число, меньшее или равное $k/2$; параметр $p$ может принимать только два значения: 1 или 0.

Функции Уолша нормированы к единице при любом порядке $k$

$$\|\mathrm{wal}(k, \theta)\|^2 = \int_{-1/2}^{1/2} \mathrm{wal}^2(k, \theta) d\theta = 1. \qquad (3.112)$$

При перемножении двух функций Уолша получается также функция Уолша (свойство мультипликативности):

$$\mathrm{wal}(k, \theta)\mathrm{wal}(n, \theta) = \mathrm{wal}(m, \theta), \qquad (3.113)$$

где $m$ — порядок функции является суммой порядков $k$ и $n$ по mod 2 (по модулю два обозначается $m = k \oplus n$).

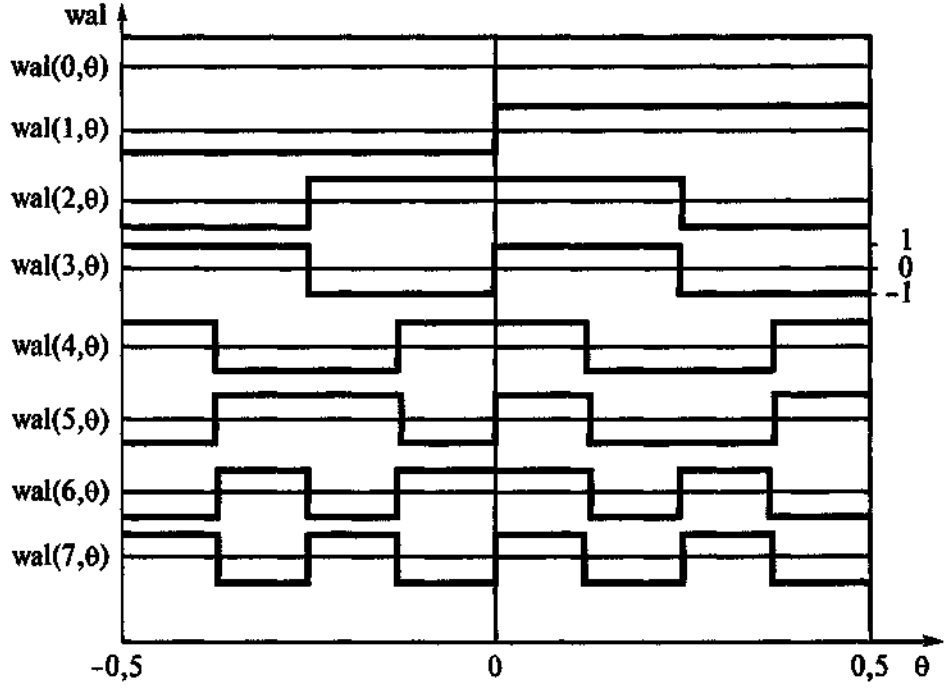



**Рис. 3.27.** Восемь первых функций Уолша

Модульное сложение двух цифр по модулю $h$ отличается от привычного тем, что полученная сумма приводит к числу, большему $h-1$. Из обычной суммы при модульном сложении вычитается число $h$. При сложении двух цифр по модулю 10, например, получим $8 \oplus 7 = 5 \pmod{10}$; при сложении цифр по модулю 2 находим $1 \oplus 0 = 1 \pmod 2$, но $1 \oplus 1 = 0 \pmod 2$ и т. д.

По существу, чтобы проделать такое сложение для функций Уолша, следует представить числа $k$ и $n$ в двоичной форме и сложить их без переносов в старшие разряды согласно правилу:

$$1 \oplus 0 = 1; \quad 0 \oplus 1 = 1; \quad 0 \oplus 0 = 0; \quad 1 \oplus 1 = 0.$$

Умножение функции Уолша самой на себя дает функцию нулевого порядка, так как результатом являются произведения вида (+1)(+1) и (–1)(–1):

$$\text{wal}(k, \theta)\text{wal}(k, \theta) = \text{wal}(0, \theta). \tag{3.114}$$

Также нетрудно показать, что умножение функции любого порядка wal($k$, θ) на функцию нулевого порядка wal(0, θ) оставляет первую без изменения:

$$\text{wal}(k, \theta)\text{wal}(0, \theta) = \text{wal}(k, \theta). \tag{3.115}$$

Из рис. 3.27 можно заметить, что четным относительно середины интервала определения (θ = 0) функциям wal($k$, θ) соответствуют четные номера $k$, а нечетным функциям — нечетные номера. Такое взаимно однозначное соответствие между четностью функций wal($k$, θ) и четностью их номера $k$ аналогично свойствам гармонических функций $\cos\left(k\frac{2\pi}{T}t\right)$ и $\sin\left(k\frac{2\pi}{T}t\right)$. Поэтому в ряде случаев используются обозначения cal($k$, θ) для четных функций и sal($k$, θ) для нечетных функций Уолша. Эти функции связаны следующими соотношениями:

$$\text{cal}(k, \theta) = \text{wal}(2k, \theta); \quad \text{sal}(k, \theta) = \text{wal}(2k-1, \theta). \tag{3.116}$$

Способ нумерации функций Уолша в системе называется *упорядочением*. Разработаны и существуют несколько принципов упорядочения функций Уолша. Функции Уолша, показанные на рис. 3.27, упорядочены по Уолшу. В ряде практических задач удобнее пользоваться другими способами упорядочения. Наиболее часто применяются системы функций Уолша, упорядоченные по Адамару (*Hadamard*) [had ($h$, θ)] и по Пэли (*Peley*) [pal ($p$, θ)]. Упорядочение по Уолшу характерно тем, что номер $k$ функции wal ($k$, θ) равен числу перемен знака на интервале ее существования.

Функции Уолша можно сформировать, например, с помощью матриц Адамара. *Матрицей Адамара* $H_N$ порядка $N = 2^n$ называется квадратная матрица размера $N \times N$ с элементами ±1 такая, что функция Уолша, упорядоченная по Адамару [had($n$,T) с номером $n$], является последовательностью прямоугольных импульсов с единичными амплитудами и полярностями, соответствующими знакам $n$-й строки матрицы. Под длительностью подразумевается $(1/N) - n$ доля интервала ортогональности [0, $T$], или при введении безразмерного времени, безразмерного интервала [0, 1].

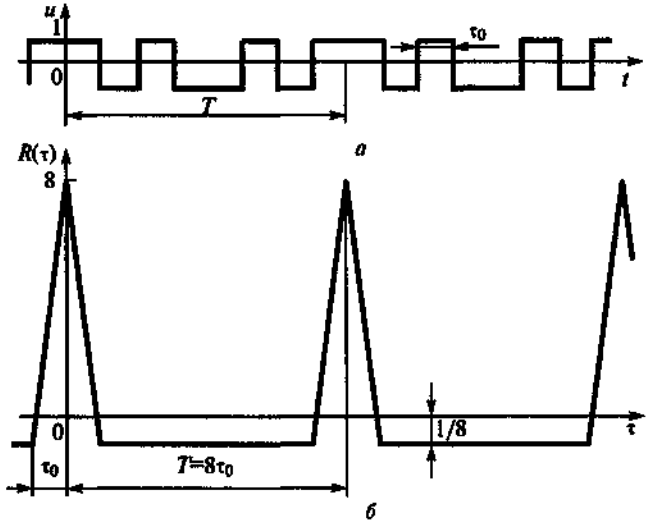



**Рис. 3.28.** Функция Уолша, упорядоченная по Адамару:
*а* — диаграмма восьмой функции; *б* — АКФ

Для примера на рис. 3.28 представлена восьмая функция Уолша, упорядоченная по Адамару и ее АКФ. Видно, что за счет корреляционной обработки выходной сигнал значительно превышает по уровню амплитуду исходной последовательности. В общем случае по рассмотренной системе базисных функций Уолша любую интегрируемую на интервале $0 \le \theta \le 1$ функцию $u(\theta)$ можно представить рядом Фурье:

$$u(\theta) = U(0) + U(1)\text{wal}(1, \theta) + U(2)\text{wal}(2, \theta) + \ldots + U(k)\text{wal}(k, \theta) \qquad (3.117)$$

с коэффициентами

$$U(k) = \int_0^1 u(\theta)\text{wal}(k,\theta)d\theta. \qquad (3.118)$$

**Пример 3.7.** Задан импульсный сигнал в форме прямоугольного треугольника с единичной амплитудой и длительностью $\tau_и$ (рис. 3.29, *а*). Вычислить два первых коэффициента разложения этого сигнала в ряд Фурье по ортонормированной системе функций Уолша и аппроксимировать ими заданный импульсный сигнал.

Р е ш е н и е. На интервале времени $[-\tau_и/2, \tau_и/2]$ заданный треугольный импульс аналитически описывается функцией вида $u(t) = t/\tau_и + 1/3$.

Вычислим коэффициенты $c_k$ обобщенного ряда Фурье по формуле (2.27)

$$c_0 = \int_{-1/2}^{1/2} u(\theta)\text{wal}(0,\theta)d\theta = \int_{-1/2}^{1/2} (\theta + 1/2)d\theta = 1/2;$$

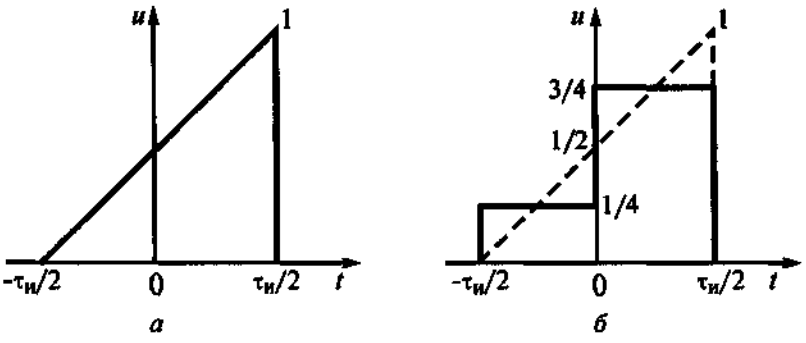



**Рис. 3.29.** Разложение треугольного сигнала по функциям Уолша:
*а* — исходный сигнал; *б* — две аппроксимирующие функции

$$c_1 = \int_{-1/2}^{1/2} u(\theta)\,\mathrm{wal}(1,\theta)\,d\theta = -\int_{-1/2}^{0} (\theta + 1/2)\,d\theta = \int_{0}^{1/2} (\theta + 1/2)\,d\theta = 1/4.$$

В результате представления заданного сигнала двумя первыми членами по ортонормированной системе функций Уолша получена приближенная аппроксимация ступенчатой формы (рис. 3.29, *б*). Увеличение суммируемых членов ряда Фурье приведет к повышению точности аппроксимации.

**Пример 3.8.** На интервале разложения $T_0$ определить спектр отрезка синусоиды единичной амплитуды $u(t) = \sin[(2\pi t)/T]$ (рис. 3.30) в базисе функций Уолша.

Р е ш е н и е. Интервал разложения $T_0$ целесообразно приравнять периоду синусоиды $T$. Как это принято в теории функций Уолша, перейдем к безразмерному времени $\theta = t/T$ и запишем исходное колебание $u(t)$ в форме $u_1(\theta) = \sin 2\pi\theta$. Для сокращения выкладок ограничимся шестнадцатью функциями, причем сначала выберем упорядочение по Уолшу.

Поскольку заданная функция $u_1(\theta)$ нечетна относительно точки $\theta = 1/2$, то все коэффициенты $U(k)$ при четных функциях Уолша в ряде (3.117), т. е. при $\mathrm{cal}(k, \theta)$, равны нулю. Остальные восемь функций $\mathrm{wal}(k, \theta)$, которые совпадают с функциями Радемахера и имеют периодичность внутри интервала [0, 1], кратную периоду функции $u_1(\theta)$, также приводят к нулевым коэффициентам $U(k)$. К таким функциям относятся wal(3, θ), wal (7, θ) и wal (15, θ). Наконец, функция wal(11, θ), нечетная не только относительно точки $\theta = 1/2$, а также относительно точек $\theta = 1/4$ и $\theta = 3/4$ (внутри интервалов [0, 1/2] и [1/2, 1]), имеет нулевой коэффициент $U(11)$ из-за четности $u_1(\theta)$ в указанных интервалах. Таким образом получено, что лишь четыре коэффициента из 16 не равны нулю: $U(1)$, $U(5)$, $U(9)$ и $U(13)$. Определим эти коэффициенты по формуле (3.118). Подынтегральные функции $u_1(\theta)\mathrm{wal}(k, \theta)$, являющиеся произведениями исследуемого сигнала и соответствующей функции Уолша, представлены на рис. 3.30, *а*. В результате интегрирования этих произведений находим:

$$U(1) = 2\int_{0}^{1/2} \sin 2\pi\theta\,d\theta = 2/\pi = 0{,}636;$$

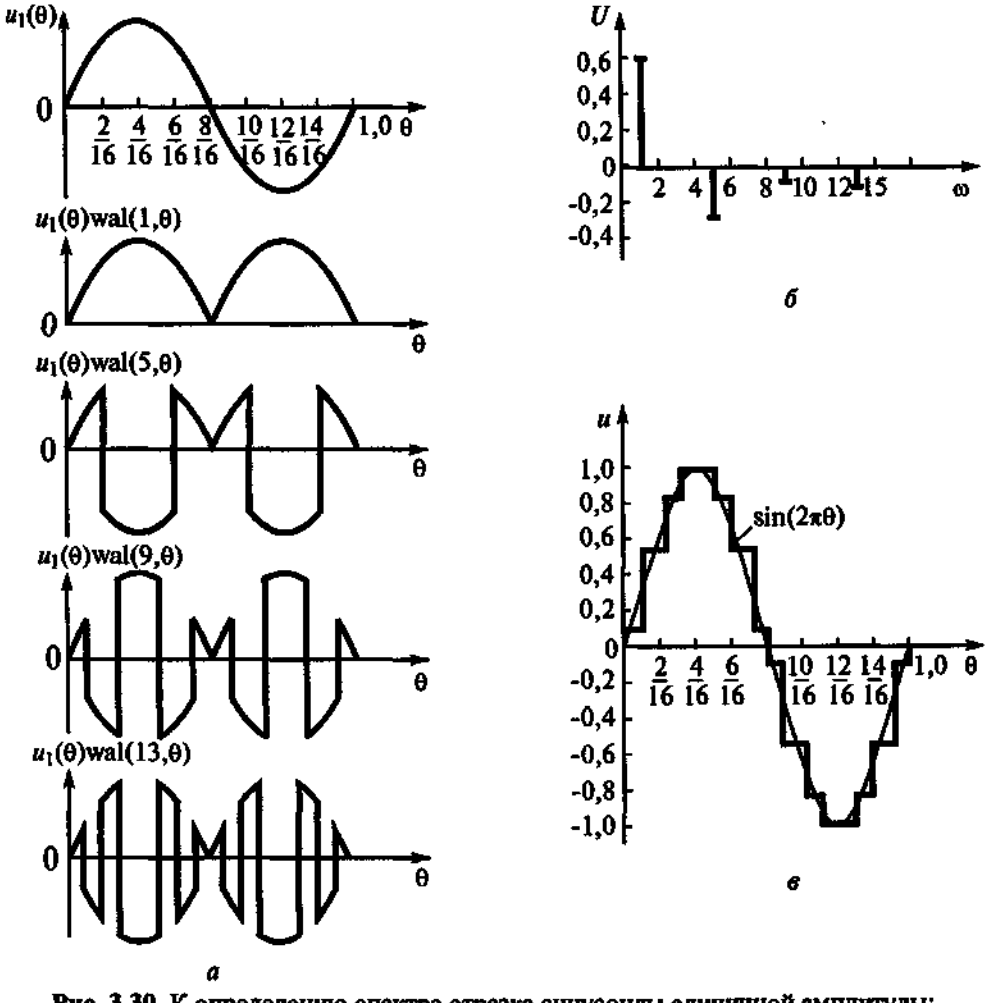



**Рис. 3.30.** К определению спектра отрезка синусоиды единичной амплитуды:
*а* — перемножение отрезка синусоиды и функций Уолша; *б* — спектр;
*в* — восстановление исходного сигнала

$$U(5) = 4\int_0^{2/16}\sin 2\pi\theta d\theta - 2\int_{2/16}^{6/16}\sin 2\pi\theta d\theta = \frac{2}{\pi}\left(1 - 2\cos\frac{\pi}{4}\right) = -0{,}265;$$

$$U(9) = 4\int_0^{1/16}\sin 2\pi\theta d\theta - 4\int_{1/16}^{3/16}\sin 2\pi\theta d\theta + 2\int_{3/16}^{5/16}\sin 2\pi\theta d\theta = -0{,}052;$$

$$U(13) = 4\int_0^{1/16}\sin 2\pi\theta d\theta - 4\int_{1/16}^{2/16}\sin 2\pi\theta d\theta + 4\int_{2/16}^{3/16}\sin 2\pi\theta d\theta - 2\int_{3/16}^{5/16}\sin 2\pi\theta d\theta = -0{,}128.$$

Спектр исследуемого сигнала $u_1(\theta)$ в базисе функций Уолша показан на рис. 3.30, *б*. Восстановление исходного сигнала (см. рис. 3.30, *а*) шестнадцатью функциями Уолша показано на рис. 3.30, *в* (двенадцать спектральных коэффициентов обращаются в нуль). Очевидно, что для более удовлетворительной аппроксимации синусоидального колебания в базисе Уолша требуется существенное увеличение числа спектральных компонентов.

## Контрольные вопросы

1. Какие сигналы называются случайными?
2. Можно ли считать реализацию случайного процесса случайным сигналом?
3. Какие основные функции характеризуют случайные процессы?
4. Перечислите основные свойства функции и плотности вероятности распределения случайной величины.
5. Какие основные числовые характеристики описывают случайные процессы?
6. Как определить АКФ случайного процесса?
7. Приведите общее выражение для дифференциальной функции распределения нормального случайного процесса.
8. Какие случайные процессы относятся к стационарным и эргодическим?
9. Запишите преобразование Винера–Хинчина и назовите основные свойства.
10. Как по спектру мощности вычислить дисперсию стационарного случайного процесса?
11. Каковы основные свойства спектральной плотности мощности стационарного случайного процесса?
12. Как определяется понятие одностороннего спектра мощности?
13. Для каких целей в радиотехнике вводят понятие «белый шум»?
14. Почему случайный процесс типа белого шума называется дельта-коррелированным случайным процессом?
15. Чем характерны шумоподобные (секвентные) сигналы?
16. Назовите основные свойства сигналов Баркера.
17. Постройте систему функций Радемахера.
18. Почему удобно в практической радиоэлектронике и технике связи разлагать сигналы по системе функций Уолша?

# ЧАСТЬ II

# РАДИОЭЛЕКТРОННЫЕ ЦЕПИ И УСТРОЙСТВА

## Глава 4. ЛИНЕЙНЫЕ ЦЕПИ

Чтобы проанализировать процессы в цепях, необходимо представить сигналы удобными математическими формулами или использовать их идеализированные модели. Положения, которые будут рассмотрены, традиционно называют теорией *линейных цепей с постоянными параметрами*. В книге же наряду с понятием «цепь» используется термин «система», чтобы подчеркнуть более высокий уровень рассмотрения. По функциональному назначению линейные цепи делят на *интегрирующие и дифференцирующие цепи* и *устройства*; *частотно-избирательные цепи*; *линейные усилители* и *фильтры*. В радиотехнических устройствах обычно происходит уменьшение мощности сигнала вследствие потерь энергии в цепях. Эти потери компенсируют с помощью усилителей. В радиотехнике с теорией построения усилителей неразрывно связана теория их устойчивости.

Чтобы выделить полезный сигнал из смеси различных сигналов, помех и шумов, нужны частотно-избирательные линейные цепи, которые выполняют на основе резонансных контуров или их аналогов. Для этого используют *фильтры*, способные выделять (пропускать) или подавлять (ослаблять) сигналы с заданным спектром частот.

### 4.1. Общие сведения о методах анализа линейных цепей

В устройствах радиоэлектроники и систем связи специалистам приходится иметь дело с разнообразными радиотехническими сигналами и электрическими цепями. В электрических цепях протекают те или иные процессы, многие из которых имеют сугубо специфический характер, обусловленный целевым назначением радиоэлектронных устройств. Прежде чем рассмотреть и проанализировать основные процессы, протекающие в линейных радиоэлектронных цепях, кратко сформулируем для этого исходные понятия и определения, принятые в радиоэлектронике и теории связи.

*Радиотехнической* или *электрической цепью* (системой) называют совокупность соединенных определенным образом элементов, предназначенных для производства, передачи, приема, преобразования и использования электрического тока. Различают активные и пассивные цепи, участки и элементы цепей.

Активными называют электрические цепи, содержащие источники энергии, пассивными — электрические цепи, не содержащие источников энергии.

Электрические цепи отображают электрическими схемами. *Электрическая схема* — графическое изображение электрической цепи, включающее в себя условные обозначения элементов и устройств и отражающее их соединения. При анализе электрическую цепь заменяют схемой замещения. *Схема замещения* — графическое изображение электрической цепи с помощью идеальных элементов, параметрами которых служат параметры замещаемых элементов.

*Активные элементы* — биполярные и полевые транзисторы, аналоговые и цифровые микросхемы, электронные лампы, магнетроны, клистроны, лазеры и пр. Отличительной особенностью активных элементов является их потенциальная способность к усилению мощности колебаний, подводимых к ним. При этом энергия колебаний увеличивается за счет энергии внешних источников питания. Таким образом, активные элементы можно рассматривать как преобразователи энергии источников питания в энергию колебаний на выходе.

При использовании активных элементов в электрических цепях различают режимы малого и большого сигналов. В режиме малого сигнала, когда амплитуда колебаний достаточно мала, активные элементы можно считать линейными, а в режиме большого сигнала — нелинейными. В соответствии с этим различают модели малого и большого сигнала.

*Пассивные элементы* (рис. 4.1) — резисторы (сопротивления, измеряются в омах, Ом; Георг Ом — Georg Ohm; 1787–1854; в 1826 г. открыл закон Ома), катушки индуктивностей (индуктивности, измеряются в генри, Гн), конденсаторы (емкости, измеряются в фарадах, Ф), соединительные проводники — передают, потребляют или накапливают электрическую (электромагнитную) энергию. Соединенные определенным образом они фактически могут лишь изменить форму электрических сигналов.

*Активное сопротивление* — сопротивление электрической цепи или ее участка, обусловленное необратимыми превращениями электрической энергии в другие виды энергии, чаще в тепловую (поэтому сопротивление принято называть активным, поскольку существуют и реактивные сопротивления). Тепловая энергия, выделяемая в активном сопротивлении, полезно используется или рас-

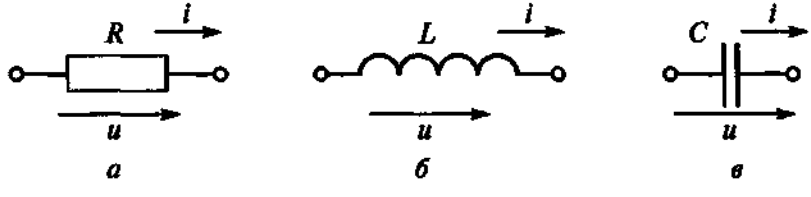



**Рис. 4.1.** Обозначения пассивных элементов:
*а* — резистор; *б* — катушка индуктивности; *в* — конденсатор

сеивается в пространстве. Величина, обратная сопротивлению, называется *проводимостью*: $g = 1/R$ (измеряется в сименсах, См). В схеме замещения во всех случаях, когда надо учесть необратимое преобразование энергии, включают сопротивление.



Г. Ом

Заметим, что использование терминов «сопротивление», «индуктивность» и «емкость» не в качестве характеристик реальных элементов, а их названий, является не вполне удачным (не стандартизированным) и оправдывается только компактностью и упрощением изложения.

*Индуктивностью* называют идеальный элемент схемы замещения, характеризующий способность цепи накапливать магнитное поле. Полагают, что индуктивностью обладают только индуктивные катушки. Индуктивностью других элементов электрической цепи пренебрегают.

*Емкостью* называют идеальный элемент схемы замещения, характеризующий способность участка электрической цепи накапливать электрическое поле.

Мгновенные значения напряжений $u(t)$ и токов $i(t)$ в этих элементах связаны соотношениями (аргумент $t$ в функциях напряжения и тока для удобства здесь опущен):

$$u_R = iR;$$

$$u_L = L\frac{di}{dt}; \qquad (4.1)$$

$$u_C = \frac{1}{C}\int i dt.$$

Активные и пассивные элементы радиотехнической цепи могут быть соединены в двухполюсники, четырехполюсники и многополюсники.

*Двухполюсник* — электрическая цепь, имеющая два вывода. Конкретизируя это понятие, отметим, что *двухполюсник* (одиночный элемент или сложная электрическая цепь; например источник питания, резистор, диод, последовательный контур) имеет всего два внешних вывода — полюса, или зажима (рис. 4.2, *а*).

Двухполюсники могут быть активными и пассивными.

Каждый пассивный двухполюсник характеризуется одним параметром, устанавливающим связь между потребляемым от источника током и падением на-

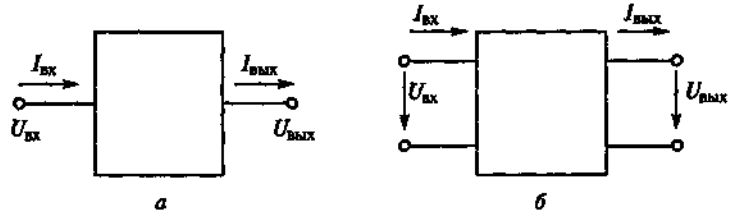



**Рис. 4.2.** Радиоэлектронные цепи:
*а* — двухполюсник; *б* — четырехполюсник

пряжения на нем. В общем случае такая связь может иметь сложную интегрально-дифференциальную форму. Однако при синусоидальной форме тока (или напряжения) линейный двухполюсник обеспечивает такую же синусоидальную форму с той же частотой и для напряжения на нем (соответственно тока).

*Четырехполюсник* содержит по паре входных и выходных выводов и имеет четыре полюса (рис. 4.2, *б*). Четырехполюсники (и двухполюсники) могут быть как активными, так и пассивными. На практике часто требуется определить связь между сигналами на входе и выходе четырехполюсника, не описывая внутренние процессы, протекающие в нем. В этом случае в теории цепей четырехполюсник принято называть «черным ящиком». Четырехполюсники и цепи, состоящие из нескольких четырехполюсников, являются основой тракта передачи и преобразования сигналов, несущих информацию. Теория четырехполюсников дает возможность единым методом анализировать системы, самые различные по структуре и принципу действия.

*Многополюсник*, как один из узлов цепи, имеет более четырех выводов.

С точки зрения соотношения размеров цепей и рабочей длины волны электрических колебаний, протекающих по ним или имеющих в них место, различают цепи с сосредоточенными и распределенными параметрами.

Радиоэлектронные цепи, физические размеры которых гораздо меньше рабочей длины волны, называют цепями с *сосредоточенными параметрами*. Свойства данных цепей практически не зависят от конфигурации выводов активных и пассивных элементов и соединительных проводов.

Радиоэлектронные цепи, физические размеры которых соизмеримы с рабочей длиной волны, относят к цепям с *распределенными параметрами*. Каждый элемент цепи обладает активным сопротивлением, индуктивностью и емкостью.

**По признаку зависимости параметров элементов** от приложенных напряжений и протекающих в них токов цепи делятся на три класса:

- линейные цепи с постоянными параметрами (линейные цепи);
- линейные цепи с переменными параметрами, или параметрические (*parametric*) цепи;
- нелинейные цепи.

*Линейные цепи* состоят из пассивных и активных элементов, параметры которых не зависят от протекающих в них токов и приложенных к ним напряжений. В общем случае речь ведут о цепях (системах), у которых связь между входным $u_{вх}(t) = u_{вх}$ и выходным $u_{вых}(t) = u_{вых}$ сигналами (здесь аргумент $t$ для упрощения опущен) устанавливают с помощью дифференциального уравнения

$$a_n \frac{d^n u_{вых}}{dt^n} + a_{n-1} \frac{d^{n-1} u_{вых}}{dt^{n-1}} + \cdots + a_1 \frac{du_{вых}}{dt} + a_0 u_{вых} =$$

$$= b_m \frac{d^m u_{вх}}{dt^m} + b_{m-1} \frac{d^{m-1} u_{вх}}{dt^{m-1}} + \cdots + b_1 \frac{du_{вх}}{dt} + b_0 u_{вх}. \qquad (4.2)$$

Если цепь (далее часто четырехполюсник) линейна, то в (4.2) все коэффициенты $a_0, a_1, \ldots, a_n$ и $b_0, b_1, \ldots, b_m$ — постоянные вещественные числа.

Положим, что входной сигнал $u_{вх}(t)$ задан. Тогда правая часть уравнения (4.2), которую условно обозначим через $i(t)$, является известной функцией. Анализ отклика линейной цепи на известное входное воздействие сводится при этом к известной в математике задаче решения линейного дифференциального уравнения $n$-го порядка с постоянными коэффициентами

$$a_n \frac{d^n u_{вых}}{dt^n} + a_{n-1} \frac{d^{n-1} u_{вых}}{dt^{n-1}} + \ldots + a_1 \frac{du_{вых}}{dt} + a_0 u_{вых} = i(t). \qquad (4.3)$$

Порядок $n$ этого уравнения в радиотехнике принято называть *порядком линейной цепи* (*системы*).

Один из простейших признаков линейности электрической цепи состоит в том, что при прохождении через линейную систему синусоида остается синусоидой: могут измениться лишь ее амплитуда и фаза (сдвиг во времени), а частота остается неизменной. Это свойство особенно принципиально, поскольку оно указывает на важнейший метод анализа линейных систем с помощью разложения входных и выходных сигналов на гармоники (Фурье-анализ). Что означает «прохождение синусоиды через линейную систему»? Это значит, что синусоида подается на вход системы бесконечно долго, т. е. от $t = -\infty$ до $t = +\infty$. Если же синусоида начала поступать лишь в некоторый конкретный момент времени (а до этого подавалось какой-то другой сигнал), то после начала подачи синусоиды на вход мы можем получить синусоиду на выходе не сразу. Выходной сигнал постепенно начнет приобретать синусоидальную форму. Скорость «стремления к правильной синусоиде» на выходе зависит от конкретной линейной системы.

К линейным цепям (системам) применим принцип *суперпозиции* (*наложения*), суть которого можно сформулировать так: отклик (выходной сигнал) линейной цепи на сложное (суммарное) воздействие нескольких входных источников равен алгебраической сумме откликов на воздействие (входной сигнал) каждого источника в отдельности. Говорят, что реакция системы на входные сигналы аддитивна. Можно сказать, что «целое равно сумме своих частей». В математической форме этот принцип выражается следующим простым равенством

$$B[u_1(t) + u_2(t) + \ldots] = B[u_1(t)] + B[u_2(t)] + \ldots, \qquad (4.4)$$

где $B$ — линейный оператор, характеризующий вид воздействия линейной цепи на входной сигнал.

Линейным системам свойственна еще и *однородность* (гомогенность), т. е. отклик системы на входной сигнал, усиленный в определенное число раз, будет усилен в то же число раз, при этом «размер не имеет значения». Линейность позволяет рассматривать систему исследования по частям, а однородность — в удобном масштабе, в частности, подавая на ее вход простые пробные (тестовые) сигналы и изучая реакцию на них системы, а сложные реальные сигналы, представляя в виде суммы простых.

Очевидно, для реальных объектов свойство линейности может иметь место лишь приближенно и в определенном интервале входных сигналов. Вместе с тем

огромное количество реальных систем по преобразованию сигналов можно считать линейными. Например микрофон является линейной системой (с достаточной степенью точности), так как если в него будут говорить одновременно два человека с разной громкостью, то электрический сигнал на выходе будет взвешенной суммой сигналов (от каждого человека в отдельности) на входе, а коэффициенты будут означать громкость разговора первого и второго человека.

Несмотря на то что линейность — удивительное свойство природы, оно отнюдь не редкое явление. Все электрические цепи, состоящие из сопротивлений, емкостей, индуктивностей и соединительных проводов, линейны. Так как и физические линии связи в основном линейны, то по одним и тем же проводам можно передавать одновременно два сигнала в противоположных направлениях, и при этом не будет возникать взаимных помех. То же самое можно сказать и о земном и космическом пространстве — существует великое множество различных источников электромагнитных колебаний (в том числе и передатчиков радиотехнических систем), которые распространяются во всевозможных направлениях и при этом практически не подавляют друг друга. Но хотя линейность довольно обычна в электрических цепях, ее нельзя считать универсальным свойством природы. В частности, два поезда не могут пройти по одной и той же железнодорожной колее навстречу друг другу, не столкнувшись.

Заметим, что каждому из классов сигналов (см. раздел 2.1) можно поставить в соответствие аналоговую, дискретную или цифровую цепи. Прохождение по этим цепям соответствующих сигналов имеет определенное различие. При рассмотрении данного раздела может возникнуть естественный вопрос: чем линейные (и другие) радиоэлектронные (радиотехнические) цепи отличаются от линейных электрических? Это деление довольно условно, и традиционно принято считать, что радиотехнические цепи предназначены для работы с сигналами радиочастотного диапазона и могут содержать усилительные элементы. Итак, радиотехнические линейные цепи, в отличие от электрических, способны не только передавать, формировать, делить и фильтровать, но одновременно с этим и линейно усиливать входные сигналы.

### Представление сигналов цепей динамическими моделями

Как уже отмечалось ранее, ряд задач радиотехники требует специфической формы представления сигналов и цепей. К таким задачам, в частности, относится определение отклика электрической цепи на заданное входное воздействие. При этом необходимо не только иметь информацию о мгновенном значении сигнала, но и представлять его поведение на всей временно́й оси как до подключения к цепи («в прошлом»), так и после окончания воздействия («в будущем»).

Метод получения подобных моделей можно трактовать следующим образом. Реальный сигнал приближенно заменяют суммой идеализированных элементарных сигналов (например, коротких импульсов, или функций включения), поступающих в заданные последовательные моменты времени. Если длительности элементарных сигналов устремить к нулю, то в пределе их сумма в точности воспроизведет исходный сигнал. Этот метод описания реального сигнала часто стали называть *динамическим представлением*, имея в виду *изменяющийся во*

*времени характер процесса*. В связи с динамическим представлением сигналов и линейные радиотехнические цепи, на которые воздействуют подобные сигналы, характеризуют как *линейные динамические системы*.

Под *динамической системой* понимают любой объект или процесс, для которого однозначно определено понятие состояния как совокупности некоторых величин в данный момент времени и задан закон, описывающий изменение начального состояния с течением времени. Закон (его называют законом эволюции) позволяет по начальному состоянию прогнозировать будущее состояние динамической системы. По существу динамическая система — математический объект, соответствующий реальным системам, эволюция которых однозначно определяется начальным состоянием. В радиотехнике динамические системы — физические объекты, вычислительные процессы и процессы преобразования сигналов и информации, совершаемые в соответствии с конкретными алгоритмами. Описания динамических систем весьма разнообразны: с помощью дифференциальных, разностных, интегральных уравнений, дискретных отображений, теории графов и т.д. Выбор способа описания задает конкретный вид математической модели динамической системы.

К линейным динамическим системам относят устройства, обладающие следующим свойством: параметры выходного сигнала определяются не только уровнем входного сигнала в рассматриваемый момент времени, но и «предысторией» этого воздействия. Линейная динамическая система обладает «памятью», от характера которой зависят особенности преобразования входного сигнала. Такой оказывается динамическая связь между мгновенными значениями входного и выходного сигналов в цепи с сосредоточенными параметрами. Типичные примеры — колебательный контур, дифференцирующая и интегрирующая цепь, усилитель напряжения. Часто линейные системы классифицируют по принципу постоянства или непостоянства их характеристик во времени. Если произвольная задержка входного сигнала приводит к такой же задержке выходного сигнала не меняя его формы, систему называют *стационарной*, или *системой с постоянными параметрами* (*аналоговые* линейные системы имеют дополнительную задержку реакции относительно воздействия на время $\tau_{зап}$).

Все стационарные линейные системы *инвариантны* (инвариантность — подобие, или неизменность) *к временно́му сдвигу*. Это означает, что форма выходного сигнала зависит только от входного сигнала и не зависит от времени начала подачи входного сигнала. Говорят, что линейная система инвариантна относительно времени (иногда ее называют *инвариантной относительно временно́го сдвига*), если ее выходной сигнал не зависит от времени приложения входного сигнала. Например, если входной сигнал $u(t)$ дает выходной сигнал $i(t)$, то воздействие $u(t-\tau)$ даст отклик $i(t-\tau)$, т. е. задержка входного сигнала $u(t)$ на $\tau$ приведет к такой же задержке выходного сигнала $i(t)$. В противном случае система называется *нестационарной, параметрической* или *системой с переменными параметрами*. Параметры стационарной системы неизменны во времени. Далее речь пойдет только о линейных стационарных системах.

Обратим внимание на следующий аспект, который по-разному трактуют в технической литературе. В электрических цепях исследуют прохождение через

них сигналов, поэтому речь идет об *анализе процессов*. В результате анализа процессов устанавливают и влияние на них параметров цепей. Поэтому очень часто говорят об *анализе электрических цепей*, хотя это не вполне корректно.



Г. Кирхгоф

**Методы анализа процессов в линейных цепях (системах)**

При анализе процессов в электрических цепях необходимо определить отклик цепи на входной сигнал в виде сигнала заданной формы. Отклик выражают в значениях напряжений $u(t)$ и токов $i(t)$ в разные моменты времени. Из теоретических основ электротехники известно, что для анализа прохождения гармонических сигналов через линейные цепи используют *законы Кирхгофа* (*Кирхгоф Густав — Kirchhoff Gustav*, знаменитый немецкий физик; 1824—1887), *методы контурных токов*, *методы узловых потенциалов*, *метод эквивалентного генератора* и другие несложные методы. Эти методы применимы и для анализа при произвольном воздействии на вход линейной цепи. Однако в радиотехнике приходится иметь дело с импульсными сигналами, которые более разнообразны по форме и спектральному составу и описываются значительным числом параметров. Кроме того, радиотехнические цепи сложны и по структуре.

При анализе воздействия сигналов на сложные по структуре цепи применяют классический, операторный, частотный методы и метод интеграла наложения. Классический метод основан на составлении и решении дифференциальных уравнений и наиболее удобен для анализа прохождения импульсных сигналов через линейные цепи. Метод достаточно прост, нагляден, хорошо отражает физическую суть происходящих в линейной цепи процессов. Однако становится очень сложным при анализе процессов и цепей, описываемых дифференциальными уравнениями выше третьего порядка. Если на сложные по структуре цепи воздействуют сложные по спектральному составу сигналы, то для анализа прохождения сигналов удобнее применять спектральный и операторный методы, а также относящийся к временным методам метод интеграла наложения.

**Спектральный метод**. Свойства линейных цепей (линейных четырехполюсников) можно определить с помощью такого параметра, как *частотный коэффициент передачи*. Для этого необходимо рассмотреть отклик линейного четырехполюсника на входное воздействие и оценить их связь между собой.

Введем понятия комплексных амплитуд входного и выходного гармонических напряжений с угловой (круговой) частотой ω и запишем их через общепринятые обозначения:

$$\mathbf{U}_{вх} = \dot{\mathbf{U}}_{вх}(\omega) = \mathrm{U}_{вх}(j\omega); \qquad \mathbf{U}_{вых} = \dot{\mathbf{U}}_{вых}(\omega) = \mathrm{U}_{вых}(j\omega).$$

Отношение комплексных амплитуд выходного и входного гармонических напряжений одной частоты определяет частотный коэффициент передачи (чаще просто *коэффициент передачи*) линейной цепи (линейного четырехполюсника)

$$\mathbf{K}(\omega) = \dot{\mathbf{K}}(\omega) = \mathbf{K}(j\omega) = \frac{\mathbf{U}_{вых}}{\mathbf{U}_{вх}} = K(\omega)\mathrm{e}^{j\varphi(\omega)}. \qquad (4.5)$$

Модуль коэффициента передачи $K(\omega) = |\mathbf{K}(\omega)|$ называют *амплитудно-частотной характеристикой* (АЧХ), а аргумент $\varphi(\omega)$ — *фазочастотной характеристикой* (ФЧХ) линейного четырехполюсника. Как правило, АЧХ имеет один максимум, а ФЧХ изменяется монотонно в зависимости от частоты (рис. 4.3). В области некоторой полосы частот отклик линейной цепи на входное воздействие начинает уменьшаться. В связи с этим используют понятие *полосы пропускания* (*рабочей полосы*) — области частот, где модуль коэффициента передачи $K(\omega)$ становится не менее $1/\sqrt{2}$ своего максимального значения. Наиболее же удобен при практических расчетах нормированный модуль коэффициента передачи $K/K_{макс}$, максимальное значение которого равно единице. Значение $1/\sqrt{2} = 0{,}707$, по которому определяют полосу пропускания линейной цепи, введено не случайно. Дело в том, что на границах полосы пропускания модуль коэффициента передачи по мощности, равный отношению выходной и входной мощностей, уменьшается в два раза.

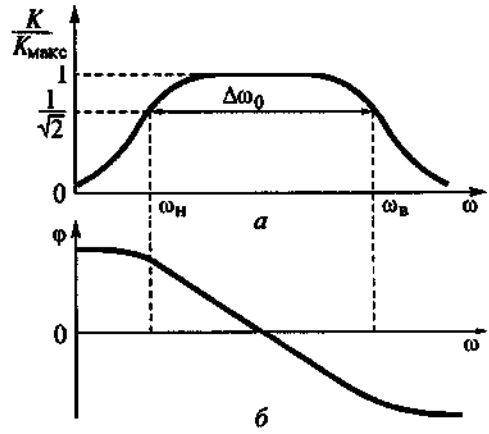



**Рис. 4.3.** Характеристики линейной цепи: *а* — амплитудно-частотная; *б* — фазочастотная

На рис. 4.3 полоса пропускания линейной цепи заключена в области от *нижней* $\omega_н$ до *верхней* $\omega_в$ круговой частоты, и поэтому ее ширина определяется как $\Delta\omega_0 = \omega_в - \omega_н$. При практических расчетах часто пользуются не круговой, а *циклической* частотой $f = \omega/(2\pi)$. В этом случае полоса пропускания цепи

$$\Delta f_0 = f_в - f_н, \tag{4.6}$$

где $f_н$ — нижняя, а $f_в$ — верхняя граничные циклические частоты.

К вопросу о частотном коэффициенте передачи можно подойти и с другой точки зрения. Если на вход линейной цепи подается гармонический сигнал единичной амплитуды, имеющий комплексную аналитическую модель вида $u_{вх}(t) = e^{j\omega t}$, то сигнал на ее выходе запишется как $u_{вых}(t) = \mathbf{K}(\omega)e^{j\omega t}$. Подставляя эти выражения в (4.2), после несложных преобразований запишем частотный коэффициент передачи в форме дифференциального уравнения

$$\mathbf{K}(\omega) = \frac{b_m(j\omega)^m + b_{m-1}(j\omega)^{m-1} + \cdots + b_1(j\omega) + b_0}{a_n(j\omega)^n + a_{n-1}(j\omega)^{n-1} + \cdots + a_1(j\omega) + a_0}. \tag{4.7}$$

Согласно (4.7), частотный коэффициент передачи линейной электрической цепи, у которой связь между входным и выходным сигналами описывается дифференциальным уравнением с постоянными коэффициентами, представляет собой дробно-рациональную функцию переменной $j\omega$. При этом коэффициенты этой функции совпадают с коэффициентами дифференциального уравнения.



С. Лаплас

С помощью частотного коэффициента передачи $\mathbf{K}(\omega)$ можно определить сигнал на выходе линейного четырехполюсника. Пусть на входе линейного четырехполюсника с частотным коэффициентом передачи $\mathbf{K}(\omega)$ действует непрерывный сигнал произвольной формы в виде напряжения $u_{вх}(t)$. Применив прямое преобразование Фурье (2.54), определим спектральную плотность входного сигнала $\mathbf{S}_{вх}(\omega)$. Тогда спектральная плотность сигнала на выходе линейного четырехполюсника

$$\mathbf{S}_{вых}(\omega) = \mathbf{S}_{вх}(\omega)\mathbf{K}(\omega). \qquad (4.8)$$

Проведя обратное преобразование Фурье (2.55) от спектральной плотности (4.8), запишем выходной сигнал как

$$u_{вых}(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathbf{S}_{вх}(\omega)\mathbf{K}(\omega)e^{j\omega t}d\omega. \qquad (4.9)$$

**Операторный метод**. Наряду со спектральным методом в теоретической радиоэлектронике и технике радиосвязи широко применяется очень близкий к нему по логической структуре операторный метод, который базируется на представлении преобразованиями Лапласа входных и выходных сигналов линейного четырехполюсника. По существу спектральный метод является разновидностью операторного метода, в котором операторным изображением сигналов по Фурье служат их спектры. Однако в отличие от реальных спектров операторные изображения сигналов в общем случае являются абстрактными математическими моделями и понятиями, которые только упрощают анализ процессов в радиотехнических цепях.

Термин «операторный метод» связан с известными работами О. Хевисайда. Он предложил символический способ решения линейных дифференциальных уравнений, описывающих переходные процессы в линейных цепях. Метод Хевисайда основан на замене оператора дифференцирования $d/dt$ комплексным параметром $p$, который переводит анализ сигналов из временно́й области в область комплексных величин.

Кратко напомним элементы основных положений преобразований Лапласа (Симон Лаплас — *Simon Laplace*, знаменитый французский математик, физик и астроном; 1749 – 1827).

Рассмотрим некоторый комплексный или вещественный аналоговый сигнал $u(t)$, определенный при $t \geq 0$ и равный нулю в момент времени $t = 0$. *Преобразование Лапласа* этого сигнала есть функция комплексной переменной $p$, выраженная интегралом

$$\mathbf{U}(p) = \int_{0}^{\infty}u(t)e^{-pt}dt. \qquad (4.10)$$

Аналитическую запись сигнала $u(t)$ называют *оригиналом*, а функцию $\mathbf{U}(p)$ его *изображением по Лапласу* (проще, *изображением*). Интеграл (4.10) внешне напоминает прямое преобразование Фурье (2.54). Однако между ними имеется принципиальное различие. В интеграл прямого преобразования Фурье (2.54) входит *мнимая частота* $j\omega$, а в интеграл Лапласа (4.10) — комплексный оператор, который можно рассматривать как *комплексную частоту* $p = \alpha + j\omega$ ($\alpha$ — вещественная составляющая), при этом рассматривают только положительные значения времени $t$. За счет множителя $e^{-\alpha t}$ под интегралом в (4.10) для $\mathbf{U}(p)$ преобразование Лапласа возможно и для неинтегрируемых функций $u(t)$.

Понятие комплексной частоты уже встречалось при анализе спектральных функций сигналов. Однако использование этого понятия непосредственно в интегральном преобразовании делает его более эффективным по сравнению с преобразованием Фурье. Например, по формуле (2.54) невозможно непосредственно определить спектр («изображение») функции включения $\sigma(t) = 1(t)$. Однако для того же сигнала непосредственно по формуле (4.10) легко отыскать его операторное изображение

$$\mathbf{U}(p) = \int_0^\infty \sigma(t) e^{-pt} dt = \int_0^\infty 1 \cdot e^{-pt} dt = -\frac{1}{p} e^{-pt}\Big|_0^\infty = -\frac{1}{p} e^{-\alpha t} e^{-j\omega t}\Big|_0^\infty,$$

или, поскольку $e^{-\alpha \cdot \infty} = 0$, получим

$$\sigma(t) = 1(t) \leftrightarrow \mathbf{U}(p) = \frac{1}{p}. \qquad (4.11)$$

Из приведенного примера видно, что повышение эффективности преобразования (4.10) обусловлено наличием множителя $e^{-\alpha t}$, который обеспечивает сходимость данного интеграла даже для сигналов, не удовлетворяющих условию сходимости интеграла $\int_{-\infty}^{\infty} |u(t)| dt < \infty$. Наличие этого множителя позволяет интерпретировать преобразование Лапласа (4.10) как представление сигнала в виде «спектра» из затухающих колебаний $e^{\alpha t} e^{j\omega t}$ (при $\alpha < 0$) в отличие от преобразования Фурье, представляющего сигнал совокупностью незатухающих гармонических колебаний $e^{j\omega t}$ (в символической форме). Очевидно, что свойства преобразования Лапласа аналогичны свойствам преобразования Фурье.

Преобразование Лапласа (4.10) обладает линейными свойствами, аналогичными свойству линейности преобразования Фурье

$$u_c(t) = u_1(t) + u_2(t) + \ldots \leftrightarrow \mathbf{U}_c(p) = \mathbf{U}_1(p) + \mathbf{U}_2(p) + \ldots \qquad (4.12)$$

Из других свойств отметим более простое преобразование изображений при дифференцировании и интегрировании сигнала по сравнению с аналогичными преобразованиями Фурье. Упрощение связано не только комплексностью оператора $p$, но и тем, что оригиналы анализируют на бесконечном интервале $[0, \infty]$.

По аналогии с обратным преобразованием Фурье при спектральном анализе сигналов в математике, радиотехнике, теории связи и других дисциплинах вводят *обратное интегральное преобразование Лапласа* (*интегральное преобразование Риммана–Меллина*), которое легко осуществляют с помощью вычетов

$$u(t) = \frac{1}{2\pi j}\int_{\alpha_1 - j\infty}^{\alpha_1 + j\infty} \mathbf{U}(p)\mathrm{e}^{pt}\,dp, \tag{4.13}$$

где $\alpha_1$ — вещественная переменная, отражаемая на комплексной плоскости.

**Решение дифференциальных уравнений операторным методом.** Преобразование Лапласа позволяет достаточно просто путем известных в математике процедур решать линейные дифференциальные уравнения с постоянными коэффициентами.

Пусть необходимо найти решение дифференциального уравнения (4.2), описывающего состояние линейной электрической цепи. Исходя из специфики работы радиотехнических устройств, установим ряд очевидных допущений:

- входной сигнал $u_{вх}(t) = 0$ при $t < 0$;
- входной сигнал содержит в себе только те функции, для которых существуют преобразования Лапласа;
- начальные условия нулевые, т. е. $u_{вых}(0) = 0$ (математически нулевые начальные условия означают, что до момента возникновения входного воздействия линейная цепь не содержит запасенной энергии).

Введем общепринятые в математике соответствия между оригиналами входного и выходного сигналов и их изображениями по Лапласу:

$$u_{вх}(t) \leftrightarrow \mathbf{U}_{вх}(p); \quad u_{вых}(t) \leftrightarrow \mathbf{U}_{вых}(p).$$

Осуществив преобразование Лапласа обеих частей дифференциального уравнения (4.2), получим

$$(a_n p^n + a_{n-1}p^{n-1} + \cdots + a_1 p + a_0)\mathbf{U}_{вых}(p) = (b_m p^m + b_{m-1}p^{m-1} + \cdots + b_1 p + b_0)\mathbf{U}_{вх}(p). \tag{4.14}$$

В теории автоматических систем сомножитель перед $\mathbf{U}_{вых}(p)$ в (4.14) обозначают через $Q(p)$, называя *собственным оператором* системы, а сомножитель перед $\mathbf{U}_{вх}(p)$ — через $R(p)$ и называют *оператором воздействия*.

В радиотехнике операторный метод базируется на важнейшей характеристике, являющейся отношением изображений выходного и входного сигналов

$$K(p) = \frac{\mathbf{U}_{вых}(p)}{\mathbf{U}_{вх}(p)} \tag{4.15}$$

и называемой *передаточной функцией* (*операторным коэффициентом передачи*) линейной цепи.

Воспользовавшись уравнением (4.14), находим

$$K(p) = \frac{R(p)}{Q(p)} = \frac{b_m p^m + b_{m-1}p^{m-1} + \cdots + b_1 p + b_0}{a_n p^n + a_{n-1}p^{n-1} + \cdots + a_1 p + a_0}. \tag{4.16}$$

Сравнение соотношений (4.7) и (4.16) показывает, что передаточная функция $K(p)$ отражает результат аналитического переноса комплексного частотного коэффициента передачи $\mathbf{K}(\omega)$ с мнимой оси $j\omega$ на всю область комплексных частот $p = \alpha + j\omega$.

Если известна передаточная функция $K(p)$, то выходную реакцию электрической цепи на заданное входное воздействие $u_{вх}(t)$ можно определить по следующей схеме:

- записать изображение входного сигнала $u_{вх}(t) \rightarrow \mathbf{U}_{вх}(p)$;
- найти изображение выходного сигнала $\mathbf{U}_{вых}(p) = K(p)\mathbf{U}_{вх}(p)$;
- вычислить выходной сигнал $u_{вых}(t) \rightarrow \mathbf{U}_{вых}(p)$.

Корни знаменателя $p_1, p_2, ..., p_n$ в формуле (4.16), т. е. корни функции

$$\ddot{И}(з) = a_n p^n + a_{n-1}p^{n-1} + \cdots + a_1 p + a_0 = 0б$$

называют *полюсами* передаточной функции $K(p)$.

Соответственно корни числителя $z_1, z_2, ..., z_m$ функции $K(p)$

$$R(p) = b_n p^n + b_{n-1}p^{n-1} + \cdots + b_1 p + b_0 = 0$$

характеризуют как *нули* передаточной функции.

В реальных электрических цепях $n \geq m$.

Нахождение корней полиномов — хорошо разработанная задача алгебры. Значения нулей и полюсов позволяют определить ряд свойств линейной цепи.

При делении числителя на знаменатель в (4.16) появляется постоянный множитель $K_0$ и это уравнение принимает так называемое *нуль-полюсное* представление передаточной функции

$$K(p) = \frac{R(p)}{Q(p)} = K_0 \frac{(p - z_1)(p - z_2)\cdots(p - z_m)}{(p - p_1)(p - p_2)\cdots(p - p_n)}. \qquad (4.17)$$

Действительные значения коэффициентов $a_n$ и $b_m$ дифференциального уравнения (4.16) обусловливает следующее свойство полюсов и нулей передаточной функции линейного четырехполюсника: либо все эти числа вещественные, либо образуют комплексно-сопряженные пары.

Очень часто используют наглядный прием отображения нулей и полюсов передаточной функции на комплексной плоскости $\alpha, j\omega$. При этом полюса принято обозначать крестиками, а нули — кружками. Например, на рис. 4.4 кружками в начале координат показаны нули, а крестиками 1 и 2 — полюсы передаточной функции некоторой колебательной цепи. Полюсы 1 и 2 отрицательны, вещественны и определяют разность двух затухающих экспонент. Комплексно-сопряженные полюсы 3 и 4 определяют колебательный характер пере-

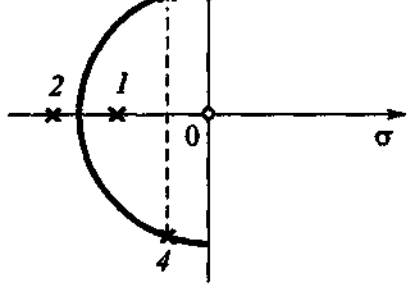



**Рис. 4.4.** Представление нулей и полюсов на комплексной плоскости

даточной функции $K(p)$ с тем бо́льшим затуханием, чем левее они расположены, и с тем бо́льшей частотой затухающих колебаний, чем дальше они отходят вверх и вниз от вещественной оси $\alpha$. Расположение полюсов в левой полуплоскости соответствует затухающему характеру передаточной функции. Нули передаточной функции могут располагаться как в левой, так и в правой полуплоскостях.

### Динамическое представление линейных цепей. Метод интеграла наложения (интеграла Дюамеля)

Наряду с частотным коэффициентом передачи свойства линейных четырехполюсников часто проще оценить видом их отклика на воздействие ряда элементарных сигналов. Широкое применение нашло два вида динамического представления линейных цепей. Согласно первому из них для анализа отклика цепи в качестве элементарных сигналов служат прямоугольные импульсы, длительностью $\Delta$, в пределе стремящиеся к дельта-функции. Эти импульсы непосредственно примыкают друг к другу и образуют последовательность, вписанную в кривую или описанную вокруг нее. При втором способе в качестве элементарных сигналов используются ступенчатые функции, возникающие в виде функций включения через равные промежутки времени $\Delta$. Высота каждой ступеньки равна приращению сигнала на интервале времени $\Delta$. Именно такой оказывается динамическая связь между мгновенными значениями входного и выходного сигналов в электрической цепи с сосредоточенными параметрами.

Одним из элементарных электрических сигналов, применяемых при анализе прохождения различных колебаний через радиоэлектронные цепи (четырехполюсники), является дельта-функция $\delta(t)$. Другим элементарным электрическим сигналом в радиотехнике служит функция включения $\sigma(t)$. Свойства и характеристики этих функций были подробно проанализированы в 2.2.

Дельта-функция и функция включения связаны между собой аналитически. Результатом дифференцирования единичной функции является дельта-функция

$$\frac{d\sigma(t)}{dt} = \delta(t). \tag{4.18}$$

Соответственно

$$\sigma(t) = \int_{-\infty}^{\infty} \delta(t)dt. \tag{4.19}$$

**Пример 4.1.** Найти производную от произведения экспоненциального импульса и функции включения $u(t) = e^{-\alpha t}\sigma(t)$.

Р е ш е н и е. Для функции $e^{-\alpha t}$ в момент времени $t = 0$ $e^{-\alpha t} = 1$. Производная

$$\frac{d}{dt}e^{-\alpha t} = -\alpha e^{-\alpha t}.$$

В результате вычислений получим следующее выражение

$$\frac{d}{dt}e^{-\alpha t}\sigma(t) = \delta(t) - \alpha e^{-\alpha t}\sigma(t).$$

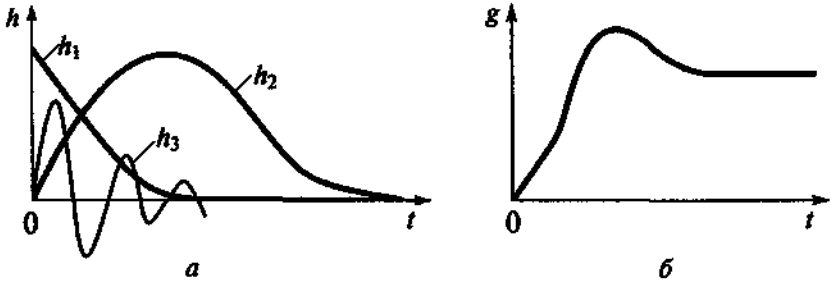



**Рис. 4.5.** Характеристики линейной цепи:
*а* — различные виды импульсных; *б* — переходная

**Импульсная и переходная характеристики линейной цепи.** Линейность и стационарность позволяют легко найти реакцию линейной системы теоретически на любой входной сигнал, зная всего одну функцию — реакцию системы на поданную на вход дельта-функцию $\delta(t)$. Эту реакцию называют *импульсной характеристикой* (*impulse response*) линейной цепи (системы) и обозначают $h(t)$. Различные виды реальных импульсных характеристик линейных цепей $h_1$, $h_2$, $h_3$ показаны на рис. 4.5, *а*.

Откликом линейной цепи на единичную функцию является *переходная характеристика* $g(t)$ (рис. 4.5, *б*). Кстати заметим, что если входной и выходной сигналы линейной цепи имеют одинаковую размерность, то импульсная характеристика, как и дельта-функция времени, имеет размерность частоты.

Положим, что требуется определить выходной сигнал $u_{вых}(t)$ линейной цепи (линейного четырехполюсника), если известны ее импульсная характеристика $h(t)$ и входной сигнал $u_{вх}(t)$. Заменим приближенно кривую входного сигнала $u_{вх}(t)$ ступенчатой линией в виде совокупности достаточно коротких прямоугольных импульсов, имеющих одинаковую длительность $\Delta\tau$ (рис. 4.6, *а*). Формирование выходного сигнала можно пояснить следующим образом. Достаточно малый «кусочек» входного сигнала длительностью $\Delta\tau$ подается на вход анализируемой цепи. Подачу такого импульса в систему можно рассматривать как «мгновенный удар» по ней. Как отреагирует (откликнется) линейная система на такое применение импульса на входе?

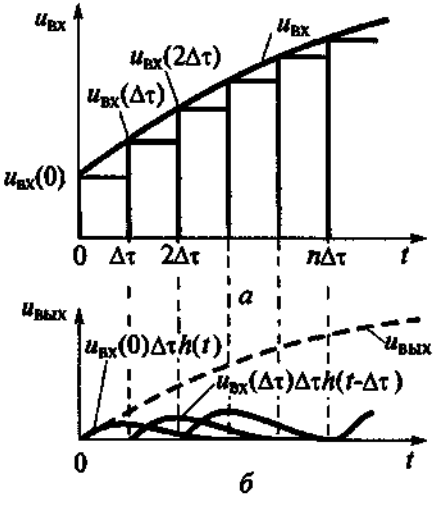



**Рис. 4.6.** К интегралу Дюамеля:
*а* — входной сигнал;
*б* — отклики на импульсы и выходной сигнал

Если выбрать длительность прямоугольных импульсов $\Delta\tau$ бесконечно малой, то отклик линейной


Ж. Дюамель

цепи на первый по счету прямоугольный импульс будет приближенно равен отклику той же цепи на дельта-функцию (а это будет импульсная характеристика), умноженному на площадь ($u_{вх}(0)\Delta\tau$) первого импульса, т. е. $u_{вх}(0)\Delta\tau h(t)$ (рис. 4.6, *б*). Откликом линейной цепи на второй импульс с достаточной точностью является произведение $u_{вх}(\Delta\tau)\Delta\tau h(t-\Delta\tau)$, где $u_{вх}(\Delta\tau)\Delta\tau$ — площадь этого импульса, а величина $h(t-\Delta\tau)$ — импульсная характеристика линейной цепи, соответствующая моменту времени $t=\Delta\tau$.

Следовательно, для некоторого произвольного момента времени $t=n\Delta\tau$ ($n$ — число условно сформированных импульсов, приходящихся на интервал времени $0...t$) отклик линейной цепи приближенно выразится такой суммой (см. штриховые линии на рис. 4.6, *б*)

$$u_{вых}(t)\approx\sum_{k=0}^{n}u_{вх}(k\Delta\tau)\Delta\tau h(t-k\Delta\tau).$$

Если длительность импульсов $\Delta\tau$, отражающих входной сигнал, последовательно приближается к нулю, то малое приращение времени $\Delta\tau$ превращается в $d\tau$, а операция суммирования трансформируется в операцию интегрирования по переменной $\tau=k\Delta\tau$

$$u_{вых}(t)=\int_{0}^{t}u_{вх}(\tau)h(t-\tau)d\tau. \qquad (4.20)$$

Для реальных линейных цепей всегда $h(t)=0$ при $t<0$. Поэтому выражение (4.20) можно записать в более общей форме

$$u_{вых}(t)=\int_{-\infty}^{\infty}u_{вх}(\tau)h(t-\tau)d\tau=u_{вх}(t)*h(t). \qquad (4.21)$$

Это соотношение, имеющее фундаментальное значение в теории линейных цепей, представляет собой *интеграл наложения*, или *интеграл Дюамеля* (Жан Дюамель — *Jean Duhamel*; 1797–1872; французский математик). Напомним, что данный интеграл в математике называют *сверткой* двух функций (см. гл. 2). Итак, линейная система осуществляет свертку входного сигнала со своей импульсной характеристикой, в результате чего получается выходной сигнал. Формула (4.21) имеет ясный физический смысл: линейная стационарная цепь, выполняя обработку входного сигнала, проводит операцию взвешенного суммирования всех его мгновенных значений, существовавших «в прошлом».

В теории электрических цепей часто применяют другую, эквивалентную форму интеграла Дюамеля

$$u_{вых}(t)=\int_{-\infty}^{\infty}u_{вх}(t-\tau)h(\tau)d\tau. \qquad (4.22)$$

К интегралу Дюамеля можно прийти и аналитическим путем. Известно, что любой сигнал может быть представлен в виде свертки самого себя с дельта-функцией (2.78)

$$u_{вх}(t) = \int_{-\infty}^{\infty} u_{вх}(\tau)\delta(t-\tau)d\tau.$$

Линейная система преобразует относительно переменной $t$ все функции, входящие в это выражение. При этом входной сигнал $u_{вх}(t)$ преобразуется в выходной сигнал $u_{вых}(t)$, а дельта-функция $\delta(t-\tau)$ — в импульсную характеристику $h(t-\tau)$. Функция $u_{вх}(\tau)$ не зависит от переменной $t$, и поэтому остается без изменений. В результате получается формула, показывающая, что выходной сигнал линейной системы с постоянными параметрами равен свертке входного сигнала с импульсной характеристикой системы

$$u_{вых}(t) = \int_{-\infty}^{\infty} u_{вх}(\tau)h(t-\tau)d\tau. \qquad (4.23)$$

Определим связь импульсной характеристики с частотным коэффициентом передачи линейной цепи. Воспользуемся гармоническим сигналом единичной амплитуды, записанным в комплексной форме $u_{вх}(t) = \exp(j\omega t)$. Подставив это выражение в (4.22) и вынеся его за знак интеграла, находим отклик цепи

$$u_{вых}(t) = \int_{-\infty}^{\infty} e^{j\omega(t-\tau)}h(\tau)d\tau = \left[\int_{-\infty}^{\infty} h(\tau)e^{-j\omega\tau}d\tau\right]e^{j\omega t}.$$

Интеграл в скобках является комплексной функцией частоты

$$K(j\omega) = \dot{K}(\omega) = \mathbf{K}(\omega) = \int_{-\infty}^{\infty} h(\tau)e^{-j\omega\tau}d\tau = \int_{-\infty}^{\infty} h(t)e^{-j\omega t}dt \qquad (4.24)$$

и представляет частотный коэффициент передачи (здесь сделана формальная замена переменной $\tau$ на $t$).

Выражение (4.24) устанавливает чрезвычайно важный факт — частотный коэффициент передачи и импульсная характеристика линейной цепи связаны прямым преобразованием Фурье. Очевидно и наличие обратного преобразования Фурье для коэффициента передачи и импульсной характеристики

$$h(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathbf{K}(\omega)e^{j\omega t}d\omega, \qquad (4.25)$$

с помощью которого можно легко определить импульсную характеристику цепи по ее частотному коэффициенту передачи.

Поскольку существует простая связь между дельта-функцией и функцией включения по формулам (4.18) и (4.19), все выводы для линейной цепи, сделанные при помощи дельта-функции, легко переносятся на функцию включения.

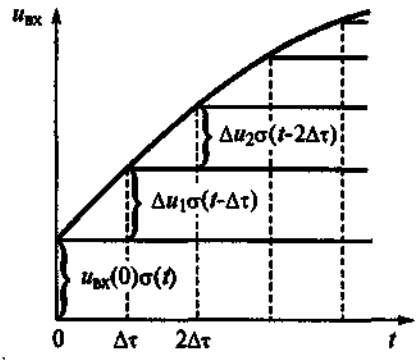



**Рис. 4.7.** Представление сигналов суммой скачков напряжений

Проведя аналогичные рассуждения и расчеты, можно показать возможность простого представления входных и выходных сигналов при помощи функции включения $\sigma(t)$ и переходной характеристики линейной электрической цепи $g(t)$.

Разбив входной сигнал (рис. 4.7) на элементарные функции включения $\Delta u\sigma(t)$ (здесь $\Delta u$ — амплитуда элементарного скачка входного напряжения) и поступая так же, как и при выводе соотношения (4.20), получаем еще одну форму интеграла Дюамеля, позволяющую определить сигнал на выходе линейной цепи,

$$u_{вых}(t) = u_{вх}(0)g(t) + \int_0^t \frac{du_{вх}(\tau)}{d\tau} g(t-\tau)d\tau. \qquad (4.26)$$

В теории линейных цепей установлена определенная связь между импульсной и переходной характеристиками. Поскольку переходная характеристика цепи $g(t)$ есть отклик на единичную функцию $\sigma(t)$, которая, в свою очередь, представляет собой интеграл от дельта-функции $\delta(t)$ [см. (4.19)], то и между функциями $h(t)$ и $g(t)$ существует интегральное соотношение

$$g(t) = \int_{-\infty}^{t} h(\tau)d\tau. \qquad (4.27)$$

Следует непременно отметить, что АЧХ $K(\omega)$ и ФЧХ $\varphi(\omega)$, а также импульсная $h(t)$ и переходная $g(t)$ характеристики позволяют оценивать влияние линейной цепи на изменение формы и параметров входного сигнала.

Экспериментально импульсную характеристику линейной цепи можно построить, подавая на ее вход короткий импульс единичной площади и уменьшая длительность импульса при сохранении площади до тех пор, пока сигнал на выходе перестанет изменяться. Это и будет импульсная характеристика цепи.

Импульсная и переходная характеристики позволяют в ряде случаев сравнительно просто найти сигнал на выходе линейной цепи при воздействии на ее вход сложных по структуре колебаний.

### Условие физической реализуемости линейной цепи

Формула (4.20) имеет четкий физический смысл — выходной сигнал $u_{вых}(t)$ линейной цепи в любой момент времени является результатом взвешенного суммирования мгновенных значений входного сигнала $u_{вх}(t)$, поступивших за предыдущее время. Роль *весовой функции* выполняет импульсная характеристика цепи, взятая с аргументом $t - \tau$, т. е. $h(t - \tau)$. Это является одним из основных условий *физической реализуемости* линейной цепи: *каков бы ни был конкрет-*

*ный вид импульсной характеристики цепи, выходной сигнал не может возникнуть до момента появления входного сигнала.* В этом состоит важнейший принцип *причинности* (*каузальности*) линейных цепей, показывающий, что для физически реализуемой системы импульсная $h(t)$ и переходная $g(t)$ характеристики должны быть равны нулю при $t < 0$.

Некаузальные системы реализовать физически невозможно, так что установив некаузальность системы с заданными (частотными, фазовыми и т. п.) характеристиками, не следует пытаться добиться невозможного. Если же требуются именно эти показатели (такие важные практически устройства, как дифференциатор или преобразователь Гильберта, увы, оказываются некуазальны) — приходится огрублять и искажать требуемые характеристики, чтобы некаузальность не простиралась далее какого-то момента времени в прошлом, и затем вводить задержку на это время, с тем чтобы следствие не наступало раньше причины.

Сформулированные требования могут быть использованы в качестве критерия физической реализуемости и частотной характеристики линейной цепи (приводится без доказательства). Однако не всякая функция $\mathbf{K}(\omega)$ физически реализуема. Согласно теореме Пэли и Винера должно выполняться условие

$$\int_0^\infty \frac{|\ln \mathbf{K}(\omega)|}{1+\omega^2} d\omega < \infty.$$

В частности, невозможно реализовать идеальный ФНЧ, частотная характеристика которого имеет вид строгого прямоугольника.

### Многомерные сигналы

Все полученные соотношения между характеристиками относятся к сравнительно простым, одномерным входному и выходному сигналам. В более общем случае при анализе прохождения многомерных сигналов через многополюсные линейные цепи с $n$ входами и $m$ выходами вводят *парциальные* импульсные характеристики $h_{ij}(t)$; $i = 1, 2, ..., m$; $j = 1, 2, ..., n$, каждая из которых представляет собой отклик на $i$-м выходе при подаче дельта-функции на $j$-й вход.

Совокупность парциальных функций $h_{ij}(t)$ можно представить в виде векторной матрицы импульсных характеристик

$$h_{ij}(t) = \mathbf{h}(t) = \begin{bmatrix} h_{11} & h_{12} & \dots & h_{1j} & \dots & h_{1n} \\ h_{21} & h_{22} & \dots & h_{2j} & \dots & h_{2n} \\ \dots & \dots & \dots & & \dots & \dots \\ h_{m1} & h_{m2} & \dots & h_{mj} & \dots & h_{mn} \end{bmatrix}. \qquad (4.28)$$

При этом интеграл Дюамеля для многомерного сигнала записывается в виде

$$\mathbf{U}_{\text{вых}}(t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty} \mathbf{U}_{\text{вх}}(\tau)\mathbf{h}(t-\tau)d\tau, \qquad (4.29)$$

где $\mathbf{U}_{\text{вх}}$ — $n$-мерный, $\mathbf{U}_{\text{вых}}$ — $m$-мерный векторы входных и выходных напряжений.

## 4.2. Дифференцирование и интегрирование сигналов

В различных радиоэлектронных устройствах и системах передачи информации широкое применение находят электрические цепи, осуществляющие дифференцирование и интегрирование сигналов. Из курса математики известно, что подобные операции относятся к линейным, поэтому они реализуются линейными электрическими цепями.

### Дифференцирующие цепи

Обратимся к одной из схем простейшего линейного четырехполюсника — последовательной электрической *RC*-цепи, на входе которой действует напряжение $u_{вх}(t)$, а выходное напряжение $u_{вых}(t)$ снимается с резистора $R$ (рис. 4.8, *а*). Выясним, в каких условиях эта цепь будет дифференцировать входной сигнал.

Воспользовавшись известным в физике вторым законом Кирхгофа, запишем мгновенное значение напряжения для всей линейной цепи

$$u_{вх}(t) = \frac{1}{C}\int i(t)dt + u_{вых}(t), \tag{4.30}$$

где $i(t)$ — мгновенное значение тока.

Продифференцируем по времени обе части этого соотношения:

$$\frac{du_{вх}(t)}{dt} = \frac{1}{C}i(t) + \frac{du_{вых}(t)}{dt}.$$

Умножив и поделив первое слагаемое в правой части на $R$ и учитывая, что $u_{вых}(t) = i(t)R$, получим

$$u_{вых}(t) + \tau_a \frac{du_{вых}(t)}{dt} = \tau_a \frac{du_{вх}(t)}{dt}. \tag{4.31}$$

Здесь параметр $\tau_a = RC$ — *постоянная времени цепи.*

Если постоянная времени цепи $\tau_a$ настолько мала, что

$$u_{вых}(t) >> \tau_a \frac{du_{вых}(t)}{dt}, \tag{4.32}$$

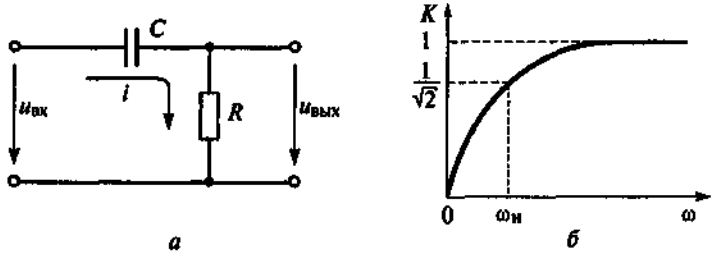



**Рис. 4.8.** Дифференцирующая цепь:
*а* — схема; *б* — АЧХ

то выражение (4.31) примет вид

$$u_{\text{вых}}(t) \approx \tau_a \frac{du_{\text{вх}}(t)}{dt}. \qquad (4.33)$$

Итак, анализируемая *RC*-цепь (рис. 4.8, *а*) при соблюдении неравенства (4.32) может осуществлять линейную операцию дифференцирования поданного на нее сигнала.

Чтобы определить частотный коэффициент передачи дифференцирующей цепи, запишем комплексную амплитуду тока:

$$\mathbf{I} = \frac{\mathbf{U}_{\text{вх}}}{R + 1/(j\omega C)}.$$

Выразив комплексную амплитуду выходного напряжения через ток $\mathbf{U}_{\text{вых}} = \mathbf{I}R$, находим частотный коэффициент передачи

$$\mathbf{K}(\omega) = \frac{\mathbf{U}_{\text{вых}}}{\mathbf{U}_{\text{вх}}} = \frac{\mathbf{I}R}{\mathbf{U}_{\text{вх}}} = \frac{R}{R + 1/(j\omega C)} = \frac{j\omega\tau_a}{1 + j\omega\tau_a}. \qquad (4.34)$$

Теперь, вычислив модуль частотного коэффициента передачи, определяем АЧХ дифференцирующей цепи

$$K(\omega) = |\mathbf{K}(\omega)| = \frac{1}{\sqrt{1 + 1/(\omega\tau_a)^2}}. \qquad (4.35)$$

Из графика АЧХ (рис. 4.8, *б*) следует, что полоса пропускания дифференцирующей цепи ограничена только со стороны нижних частот. Приравняв правую часть (4.35) к $1/\sqrt{2}$, вычислим нижнюю граничную частоту полосы пропускания

$$\omega_{\text{н}} = 1/\tau_a. \qquad (4.36)$$

Проанализируем прохождение импульсных сигналов через дифференцирующую цепь, подав на ее вход напряжение в виде прямоугольного импульса с амплитудой $E$ и длительностью $\tau_{\text{и}}$ (рис. 4.9, *а*). В радиотехнике и теории связи верхнюю частоту спектральной плотности прямоугольного видеоимпульса принято приближенно ограничивать значением $\omega_{\text{в}} = 2\pi/\tau_{\text{и}}$. Тогда условие для соотношения частот $\omega_{\text{в}} > \omega_{\text{н}}$ можно записать в виде $\tau_a << \tau_{\text{и}}/(2\pi)$, или как $\tau_a << \tau_{\text{и}}$.

На рис. 4.9, *б* показаны графики выходного напряжения при различных отношениях постоянной времени $\tau_a$ и длительности входного импульса $\tau_{\text{и}}$. Из графических построений видно, что при $\tau_a/\tau_{\text{и}} = 10$ формы выходного напряжения и входного импульса практически совпадают. С уменьшением отношения $\tau_a/\tau_{\text{и}}$ увеличиваются завал вершины выходного импульса и его отрицательный выброс при $t > \tau_{\text{и}}$. При малых отношениях $\tau_a/\tau_{\text{и}}$ форма выходного напряжения представляет собой два разнополярных остроконечных импульса. Считают, что при отношении постоянной времени и длительности $\tau_a/\tau_{\text{и}} < 0{,}1$ *RC*-цепь дифференцирует входной импульс.

Отметим, что в радиоэлектронных устройствах (например, усилительных каскадах) при больших отношениях $\tau_a/\tau_н$ линейную электрическую *RC*-цепь применяют как *разделительную*, разделяющую цепи переменного и постоянного токов, а при малых $\tau_a/\tau_н$ — как *дифференцирующую*.

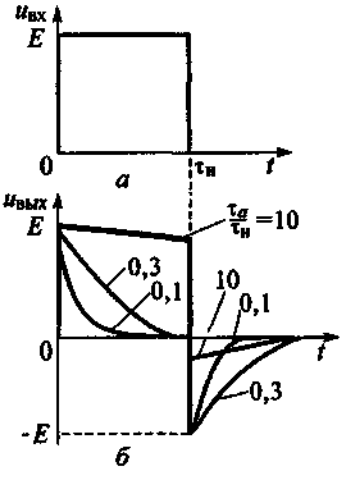



**Рис. 4.9.** Операция дифференцирования: *а* — входной сигнал; *б* — выходной сигнал

**Пример 4.2.** Положим, что в момент времени $t = 0$ (см. рис. 4.8, *а*) к *RC*-цепи подключается гармоническое напряжение $u_{вх}(t) = U_н\cos\omega_0 t$. С помощью спектрального метода определить выходное напряжение.

Р е ш е н и е. Обратившись к прямому преобразованию Фурье (2.54), найдем спектральную плотность входного сигнала

$$\mathbf{S}_{вх}(\omega) = \int_{-\infty}^{\infty} U_н\cos\omega_0 t\, e^{-j\omega t}dt = U_н\frac{j\omega}{\omega_0^2 - \omega^2}.$$

Подставив это значение спектральной плотности и коэффициент передачи $\mathbf{K}(\omega)$ из формулы (4.34) в (4.9), находим выходное напряжение

$$u_{вых}(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathbf{S}_{вх}(\omega)\mathbf{K}(\omega)e^{j\omega t}d\omega = \frac{U_н}{2\pi}\int_{-\infty}^{\infty}\frac{j\omega}{\omega_0^2 - \omega^2}\frac{j\omega\tau_a}{1 + j\omega\tau_a}e^{j\omega t}.$$

Вычисление данного интеграла дает следующий результат:

$$u_{вых}(t) = U_н\left[\frac{1/\tau_a^2}{\omega_0^2 + 1/\tau_a^2}e^{-t/\tau_a} + \frac{\omega_0}{\sqrt{\omega_0^2 + 1/\tau_a^2}}\right]\cos(\omega_0 t + \varphi_a),$$

где $\varphi_a = \operatorname{arctg}(\omega_0\tau_a)$ — фазовый сдвиг выходного сигнала по отношению к начальной фазе входного сигнала, вносимый дифференцирующей *RC*-цепью.

В реальных цепях $1/\tau_a << \omega_0$, поэтому по истечении некоторого времени первое слагаемое в скобках стремится к нулю, и в цепи устанавливается гармоническое колебание $u_{вых}(t) = U_н\cos(\omega_0 t + \varphi_a)$, сдвинутое относительно входного на угол $\varphi_a$.

### Интегрирующие цепи

Рассмотрим *RC*-цепь, в которой выходной сигнал снимается с емкости *C* (рис. 4.10, *а*). Определим условие, при котором цепь интегрирует сигнал. Используя аналогию с формулой (4.30), запишем

$$u_{вх}(t) = \tau_a\frac{du_{вых}(t)}{dt} + u_{вых}(t). \qquad (4.37)$$

Пусть постоянная времени цепи настолько велика, что

$$\tau_a \frac{du_{вых}(t)}{dt} >> u_{вых}(t),$$

тогда

$$u_{вх}(t) \approx \tau_a \frac{du_{вых}(t)}{dt}.$$

Интегрирование последнего выражения дает:

$$u_{вых}(t) = \frac{1}{\tau_a} \int_{-\infty}^{t} u_{вх}(t)dt. \qquad (4.38)$$

Итак, *RC*-цепь с большой постоянной времени интегрирует входной сигнал.

Для определения частотного коэффициента передачи интегрирующей цепи запишем комплексную амплитуду тока через комплексное входное напряжение

$$\mathbf{I} = \frac{\mathbf{U}_{вх}}{R + 1/(j\omega C)}.$$

Учитывая, что $\mathbf{U}_{вых} = \mathbf{I}/(j\omega C)$, и используя последнее выражение, запишем

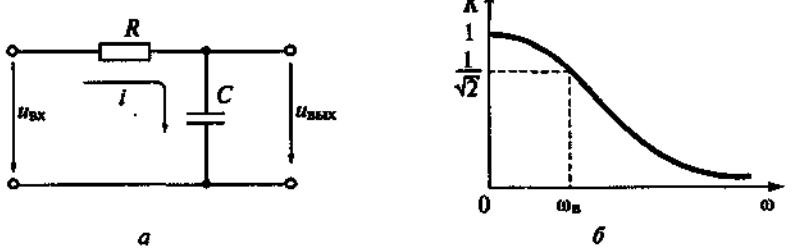
$$\mathbf{K}(\omega) = \frac{\mathbf{U}_{вых}}{\mathbf{U}_{вх}} = \frac{\mathbf{I}/(j\omega C)}{\mathbf{U}_{вх}} = \frac{\mathbf{U}_{вх}/(j\omega C)}{\mathbf{U}_{вх}[R + 1/(j\omega C)} = \frac{1}{1 + j\omega\tau_a}. \qquad (4.39)$$

Вычислив модуль данного частотного коэффициента передачи, находим АЧХ интегрирующей цепи (рис. 4.10, *б*)

$$K(\omega) = \frac{1}{\sqrt{1 + (\omega\tau_a)^2}}. \qquad (4.40)$$

Приравняв правую часть формулы (4.40) к $1/\sqrt{2}$, можно определить верхнюю граничную частоту полосы пропускания интегрирующей цепи

$$\omega_в = 1/\tau_a. \qquad (4.41)$$




*а* *б*

**Рис. 4.10.** Интегрирующая цепь:
*а* — схема; *б* — АЧХ

Из рис. 4.10, *б* следует, что интегрирующая цепь не пропускает высокочастотные составляющие спектра входного сигнала, поэтому в радиоэлектронных устройствах их используют в качестве так называемых *сглаживающих* (*smoothing*), или *низкочастотных, фильтров*. При подаче на вход такой *RC*-цепи импульсных сигналов форма выходного напряжения будет существенным образом зависеть от отношения $\tau_a/\tau_и$.

На рис. 4.11 показаны соответственно входной сигнал в виде прямоугольного импульса напряжения и отклики *RC*-цепи на него при разных значениях отношения постоянной времени и длительности входного импульса $\tau_a/\tau_и$. Из графических построений следует, что при значениях отношения $\tau_a/\tau_и = 0{,}1$ форма выходного сигнала близка к форме входного импульса. Считается, что выходной импульс практически совпадает по форме с входным при $\tau_a/\tau_и = 0{,}03$. С увеличением отношения $\tau_a/\tau_и$ до 0,5 и более выходное напряжение не успевает возрасти до значения $E$, и форма выходного сигнала существенно искажается.

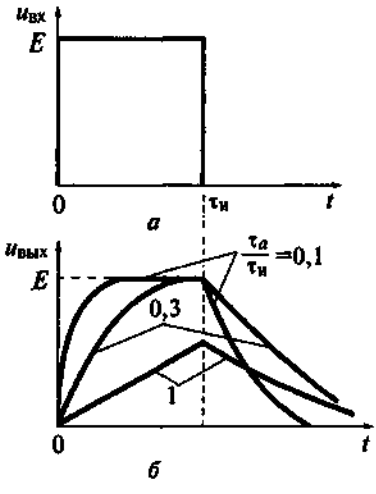



**Рис. 4.11.** Операция интегрирования:
*а* — входной сигнал;
*б* — выходной сигнал

**Пример 4.3.** Пусть в момент времени $t = 0$ на интегрирующую *RC*-цепь действует прямоугольный импульс напряжения с амплитудой $E$ и длительностью $\tau_и$ (рис. 4.11, *а*). С помощью метода интеграла Дюамеля определить закон изменения выходного напряжения.

Р е ш е н и е. Из теоретических основ электротехники известно, что импульсная характеристика рассматриваемой *RC*-цепи имеет следующий вид

$$h(t) = \frac{1}{\tau_a} e^{-t/\tau_a}.$$

Отклик цепи будем анализировать для двух случаев: $0 \le t \le \tau_и$ и $t > \tau_и$. Для вычисления воспользуемся формулой (4.20). При $0 \le t \le \tau_и$ имеем

$$u_{вых}(t) = \int_0^t u_{вх}(\tau) h(t-\tau)\,d\tau = \int_0^t \frac{E}{\tau_a} e^{-(t-\tau)/\tau_a}\,d\tau = E(1 - e^{-t/\tau_a}).$$

Соответственно при $t > \tau_и$

$$u_{вых}(t) = \frac{E}{\tau_a} \int_0^{\tau_и} e^{-t/\tau_a} e^{-\tau/\tau_a}\,d\tau = E e^{-t/\tau_a}(e^{-\tau_и/\tau_a} - 1).$$

Законы изменения выходного напряжения показаны на рис. 4.11, *б*.

## 4.3. Резонансные цепи

Выделение полезного сигнала из смеси побочных сигналов и шумов осуществляют частотно-избирательными цепями, которые строят на основе колебательных контуров, фильтров на ПАВ и других элементах опто- и акустоэлектроники.

**Последовательный колебательный контур.** Последовательный колебательный (часто, резонансный) контур состоит из последовательно соединенных сопротивления $R$, индуктивности $L$ и емкости $C$ (рис. 4.12, *а*). Обычно резистор $R$ определяет сопротивление омических потерь провода, из которого выполнена индуктивность. Сопротивление потерь емкости мало, и его на практике не учитывают. Положим, что на входе последовательного контура действует гармоническое напряжение с амплитудой $U_{вх}$ и частотой ω. Запишем *полное входное сопротивление* (*импеданс* — устаревшее) контура

$$z = R + j\omega L + 1/(j\omega C) = R + j[\omega L - 1/(\omega C)], \qquad (4.42)$$

где $j\omega L$ и $1/(j\omega C)$ — *реактивные сопротивления* соответственно индуктивности и емкости.

Частоту $\omega_р$, на которой реактивная составляющая входного сопротивления равна нулю, называют *резонансной* (от лат. *resono* — откликаюсь) *частотой контура*. Приняв $x(\omega_р) = \omega_р L - 1/(\omega_р C) = 0$, находим (*формула Томсона*)

$$\omega_р = 2\pi f_р = \frac{1}{\sqrt{LC}}, \qquad (4.43)$$

где $f_р$ — *циклическая резонансная частота контура*.

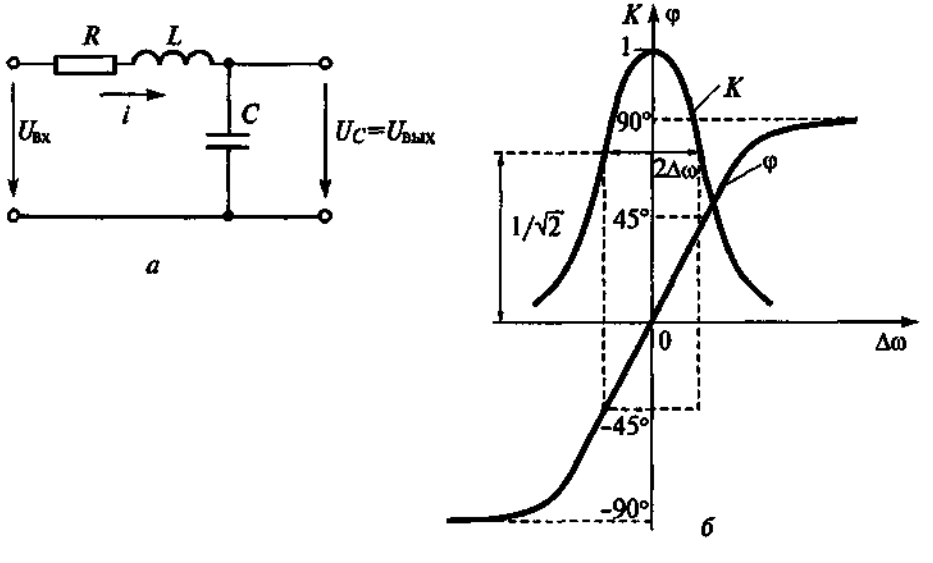



**Рис. 4.12.** Последовательный колебательный контур:
*а* — схема; *б* — АЧХ и ФЧХ

На резонансной частоте входное сопротивление этого контура активно (т. е. определяется омическими потерями) и равно $R$. На любой другой частоте входного сигнала реактивное сопротивление колебательного контура

$$x(\omega) = \omega L - \frac{1}{\omega C} = \omega_p L\left(\frac{\omega}{\omega_p} - \frac{1}{\omega\omega_p LC}\right) = \rho\left(\frac{\omega}{\omega_p} - \frac{\omega_p}{\omega}\right). \qquad (4.44)$$

Реактивное сопротивление индуктивности или емкости на частоте резонанса

$$\rho = \omega_p L = \frac{1}{\omega_p C} = \sqrt{\frac{L}{C}} \qquad (4.45)$$

называют *характеристическим сопротивлением контура.*

Преобразуем полное входное сопротивление контура (4.42) к виду

$$z = R\left[1 + j\frac{\rho}{R}\left(\frac{\omega}{\omega_p} - \frac{\omega_p}{\omega}\right)\right] = R\left[1 + jQ\left(\frac{\omega}{\omega_p} - \frac{\omega_p}{\omega}\right)\right] = R(1 + j\xi), \qquad (4.46)$$

где

$$\xi = Q\left(\frac{\omega}{\omega_p} - \frac{\omega_p}{\omega}\right) \qquad (4.47)$$

— *обобщенная расстройка*;

$$Q = \rho/R \qquad (4.48)$$

— *добротность контура.*

Оценим частотно-избирательные свойства колебательного контура. Ток, протекающий в контуре, описывается выражением:

$$\mathbf{I} = \frac{\mathbf{U}_{вх}}{z} = \frac{U_{вх}}{R}\frac{1}{1 + j\xi} = I_p\frac{1}{1 + j\xi}. \qquad (4.49)$$

Ток $\mathbf{I}$ в последовательном резонансном контуре, имеющий максимальное значение на частоте резонанса $I_p = U_{вх}/R$, называют *резонансным* током.

Введем понятие частотного коэффициента передачи по току:

$$\mathbf{K}_I(\xi) = \frac{\mathbf{I}}{I_p} = \frac{1}{1 + j\xi} = \frac{1}{\sqrt{1 + \xi^2}}e^{j\varphi}, \qquad (4.50)$$

где $\varphi$ — величина сдвига фазы входного сигнала на выходе контура.

При больших добротностях ($Q >> 1$; ее реальные величины 250 ... 400) имеем $\omega \approx \omega_p$ и обобщенную расстройку можно приближенно записать как

$$\xi = Q\frac{(\omega + \omega_p)(\omega - \omega_p)}{\omega\omega_p} \approx Q\frac{2\omega(\omega - \omega_p)}{\omega\omega_p} = 2Q\frac{\Delta\omega}{\omega_p}, \qquad (4.51)$$

где $\Delta\omega = \omega - \omega_p$ — *абсолютная расстройка* контура относительно частоты входного сигнала.

Подставив в (4.50) значение $\xi$ из (4.51), найдем модуль частотного коэффициента передачи для малой расстройки частоты

$$K_1(\Delta\omega) = \frac{1}{\sqrt{1 + (2Q\Delta\omega/\omega_p)^2}}. \qquad (4.52)$$

Данная функция представляет собой АЧХ контура, графически отражаемую *резонансной кривой* (рис. 4.12, *б*).

Так как на границах полосы пропускания АЧХ $K_1(\Delta\omega) = 1/\sqrt{2}$, то ее ширина (для колебательных контуров полосу пропускания исторически принято записывать через $2\Delta\omega_p$; см. рис. 4.12, *б*):

$$2\Delta\omega_p = \omega_p/Q. \qquad (4.53)$$

В выражении (4.50) аргумент функции определяется формулой

$$\varphi = \operatorname{arctg}(2Q\Delta\omega/\omega_p), \qquad (4.54)$$

и представляет собой ФЧХ контура для малых расстроек (см. рис. 4.12, *б*).

Так как на частоте резонанса напряжения на контуре $U_{вх} = I_p R$, $U_C = I_p \rho$, то

$$K(\omega_p) = \frac{U_{вых}}{U_{вх}} = \frac{U_C}{U_{вх}} = \frac{\rho}{R} = Q. \qquad (4.55)$$

Итак, при настройке контура в резонанс амплитуда напряжения на конденсаторе (или индуктивности) в $Q$ раз больше амплитуды входного напряжения. Поэтому резонанс в последовательном контуре называют *резонансом напряжений.*

**Пример 4.4.** В последовательном контуре частота резонанса $f_0$ = 465 кГц, а добротность $Q$ = 100. Найти полосу пропускания контура по циклической частоте.

Р е ш е н и е. Используя формулу (4.53) для циклической частоты, находим:

$$2\Delta f_p = f_p/Q = 465/100 = 4{,}65 \text{ кГц}.$$

**Параллельный колебательный контур.** Такой контур состоит из параллельно соединенных индуктивности $L$ и емкости $C$, а в цепь индуктивности включено сопротивление ее потерь $R$ (рис. 4.13, *а*). Полное входное сопротивление контура

$$z = \frac{(R + j\omega L)/(j\omega C)}{R + j\omega L + 1/(j\omega C)}. \qquad (4.56)$$

Для резонансной частоты $\omega_p = 1/\sqrt{LC}$ и большой добротности $Q >> 1$ справедливо неравенство $\omega_p L >> R$. В этом случае выражение (4.56) существенно упрощается. Воспользовавшись формулами (4.43) — (4.46), запишем

$$z \approx \frac{j\omega L/(j\omega C)}{R + j[\omega L - 1/(\omega C)]} = \frac{L/C}{R(1 + j\xi)} = \frac{\rho^2}{R(1 + j\xi)} = \frac{R_0}{1 + j\xi}. \qquad (4.57)$$

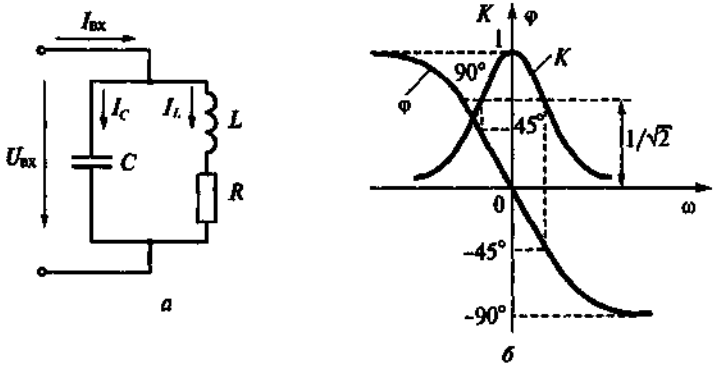



**Рис. 4.13.** Параллельный колебательный контур:
*а* — схема; *б* — АЧХ и ФЧХ

Здесь параметр $\xi$ — обобщенная расстройка, определяемая по (4.51), а

$$R_0 = \rho^2/R = \rho Q = RQ^2 \tag{4.58}$$

*— резонансное сопротивление параллельного контура.*

Аналитически АЧХ параллельного контура отражается зависимостью нормированного по резонансному сопротивлению модуля входного сопротивления от величины абсолютной расстройки

$$K_z(\Delta\omega) = \frac{z}{R_0} = \frac{R_0}{R_0\sqrt{1+\xi^2}} = \frac{1}{\sqrt{1+(2Q\Delta\omega/\omega_p)^2}}. \tag{4.59}$$

ФЧХ параллельного контура определяется следующим выражением

$$\varphi = -\operatorname{arctg}(2Q\Delta\omega/\omega_p). \tag{4.60}$$

Графики АЧХ и ФЧХ параллельного контура показаны на рис. 4.13, *б*.

Частотный коэффициент передачи контура по току нетрудно определить, вычислив отношение тока емкости (индуктивности) к входному току. На резонансной частоте этот параметр выразится простой формулой

$$K_I(\omega_p) = Q. \tag{4.61}$$

Итак, на резонансной частоте ток в параллельном контуре в $Q$ раз больше входного тока. Поэтому говорят о *резонансе токов* в параллельном контуре.

Полоса пропускания параллельного контура, как и последовательного, определяется формулой (4.53).

**Пример 4.5.** К источнику гармонического тока подключен параллельный колебательный контур с добротностью $Q = 150$, настроенный на резонансную частоту $f_p = 5$ МГц. Определить, во сколько раз ослабляется значение напряжения на контуре при расстройке его по частоте на $\Delta f = 0{,}05$ МГц по сравнению с значением резонансного напряжения.

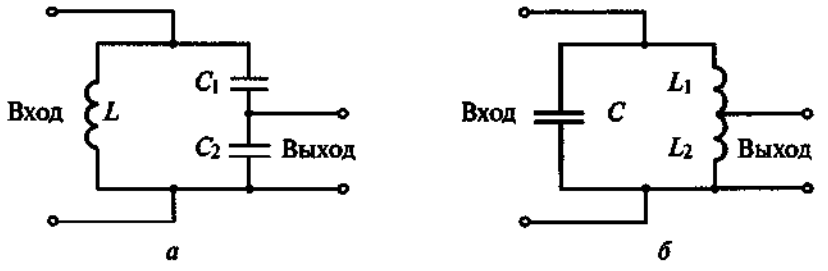



**Рис. 4.14.** Неполное включение контуров:
*а* — через емкости; *б* — через индуктивности

Решение. Воспользовавшись формулой (4.51), определяем обобщенную расстройку контура

$$\xi = 2Q\Delta f / f_p.$$

Амплитуда напряжения на контуре пропорциональна модулю его входного сопротивления. Из соотношения (4.57) имеем

$$\frac{|z|}{R_0} = \frac{1}{\sqrt{1+\xi^2}}.$$

Подставив ξ, находим, что амплитуда напряжения на контуре при расстройке в 0,05 МГц уменьшается в 3,2 раза по сравнению с резонансным напряжением.

Часто, с целью согласования с другими (как правило, низкоомными) цепями, требуется уменьшить выходное сопротивление контура, для чего используют его *неполное*, или *частичное*, включение. Внешние цепи при этом могут подключаться либо к средней точке соединения двух емкостей, либо к отводу части витков индуктивности (рис. 4.14). Возможно также согласование контуров с высокоомными цепями. В этом случае входы и выходы контуров с неполным включением необходимо поменять местами.

В общем случае величина уменьшения входного сопротивления определяется *коэффициентами включения* контуров, которые соответственно равны:

$$p_C = C_1/(C_1 + C_2);$$
$$p_L = L_2/(L_1 + L_2). \qquad (4.62)$$

Резонансное сопротивление таких контуров вычисляется с учетом коэффициента включения (он обозначен через $p$)

$$R_{0p} = p^2 \rho Q = p^2 R_0. \qquad (4.63)$$

**Связанные контуры**. Если требуется существенно повысить частотную избирательность радиотехнических устройств, то используют многоконтурные линейные цепи, в которых удается получать близкую к *идеальной* (*прямоугольной*) форму АЧХ. Простейшими многоконтурными частотно-избирательными цепями являются два *связанных колебательных контура*.

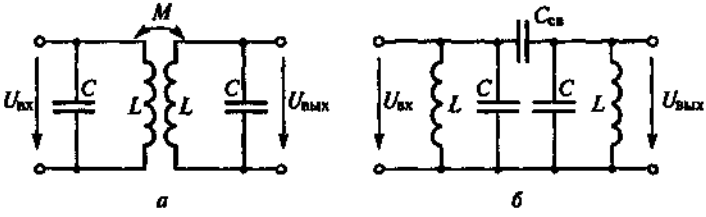



**Рис. 4.15.** Виды связи в контурах:
*а* — индуктивная; *б* — емкостная

Не проводя подробных аналитических выкладок, напомним основные положения, которые во многом базируются на анализе одиночных колебательных контуров. Для удобства реализации связанные контуры составляют из одинаковых элементов, а связь между ними чаще всего бывает индуктивной или емкостной (рис. 4.15). При индуктивной связи между катушками индуктивностей образуется *взаимная индуктивность*, определяемая *коэффициентом взаимной индуктивности M* (не путать с глубиной модуляции).

Одним из основных параметров связанных контуров является *коэффициент связи* $k_{св}$. Для связанных контуров с индуктивной связью $k_{св} = M/L$, а с емкостной — $k_{св} = C/(C + C_{св})$. Наиболее же важным параметром обычно считают *фактор связи* $A_c = k_{св}Q$. При $A_c < 1$ связь называют *слабой*, а при $A_c > 1$ — *сильной*.

С помощью несложных выкладок можно показать, что для индуктивно-связанных контуров частотный коэффициент передачи по напряжению

$$\mathbf{K}(\omega) = \frac{A_c Q}{1 - \xi^2 + A_c^2 - 2j\xi}, \tag{4.64}$$

где $Q$ — добротность; $\xi$ — обобщенная расстройка, определяемые соответственно формулами (4.48) и (4.51).

АЧХ связанных контуров определяется модулем коэффициента передачи $\mathbf{K}(\omega)$:

$$K(\omega) = \frac{A_c Q}{\sqrt{\left(1 - \xi^2 + A_c^2\right)^2 + 4\xi^2}}. \tag{4.65}$$

В качестве наглядного примера на рис. 4.16 показаны АЧХ индуктивно-связанных контуров при разных факторах связи, построенные по (4.65) для циклической частоты.

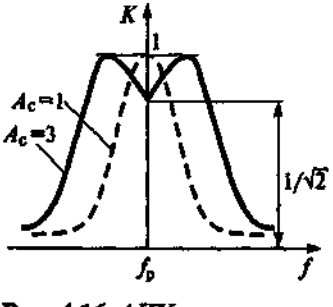



**Рис. 4.16.** АЧХ индуктивно-связанных контуров

Колебательные системы из большого числа связанных контуров называются *фильтрами сосредоточенной селекции*. С их помощью удается получить амплитудно-частотную характеристику, еще более приближающуюся к прямоугольной форме. Фильтры сосредоточенной селекции находят особенно широкое применение в аналоговых (высокочастотных) частях схем приемников различных радиотехнических систем.

**Неискажающая передача сигналов через линейные цепи.** Рассмотрим идеальный линейный четырехполюсник, частотный коэффициент передачи которого теоретически определяется функцией вида

$$\mathbf{K}(\omega) = K(\omega)e^{j\varphi(\omega)} = K_{н}e^{-j\omega t_c}, \qquad -\infty < \omega < \infty, \tag{4.66}$$

где $K_н = K(\omega)$ — постоянный коэффициент; $t_c = \varphi(\omega)/\omega$ — некоторый момент времени. Представленная функция показывает, что АЧХ линейного четырехполюсника равномерна, а ФЧХ — линейна в бесконечной полосе частот (рис. 4.17).

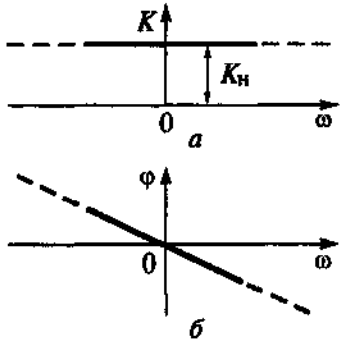



**Рис. 4.17.** Идеальный линейный четырехполюсник:
*а* — АЧХ; *б* — ФЧХ

Определим форму сигнала (напряжения) на выходе линейного четырехполюсника, если на его входе действует аналоговый сигнал (напряжение) $u_{вх}(t)$, имеющий спектральную плотность $\mathbf{S}_{вх}(\omega)$.

В соответствии с известной формулой (4.9) запишем выражение для выходного сигнала

$$u_{вых}(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathbf{S}_{вх}(\omega)K_{н}e^{-j\omega t_c}e^{j\omega t}d\omega = \frac{K_{н}}{2\pi}\int_{-\infty}^{\infty}\mathbf{S}_{вх}(\omega)e^{j\omega(t-t_c)}d\omega. \tag{4.67}$$

Как следует из теоремы запаздывания (2.63), произведение $\mathbf{S}_{вх}(\omega)e^{-j\omega t_c}$ является спектральной плотностью сигнала, сдвинутого относительно входного сигнала на интервал времени $t_c$. Значит, колебание на выходе идеального линейного четырехполюсника

$$u_{вых}(t) = K_{н}u_{вх}(t - t_c) \tag{4.68}$$

и с точностью до постоянного коэффициента $K_н$ повторяет смещенный на определенное время входной сигнал.

Таким образом, идеальный линейный четырехполюсник, обладающий равномерной АЧХ и линейной ФЧХ в бесконечной полосе частот, теоретически осуществляет передачу радиотехнических сигналов без искажений. В практических линейных цепях даже в полосе пропускания АЧХ не всегда равномерна, а ФЧХ — не строго линейна. Такие изменения могут привести к тому, что амплитуды не всех спектральных составляющих входного сигнала пропорционально (линейно) изменятся при прохождении через линейную цепь, а фазовые соотношения сигналов на входе и выходе будут нарушены. Вследствие этого формы выходного и входного сигналов могут отличаться, т. е. появятся искажения. Но важной особенностью линейных цепей является то, что при прохождении через них сигналов не нарушается форма ни одной гармонической составляющей, а может изменяться лишь их амплитуда и начальная фаза. Поэтому такие искажения в линейных цепях относят к классу *линейных* (иначе, *частотных*). Возможны и фазовые искажения.

## 4.4. Схемотехника усилительных устройств

Электронное устройство, предназначенное для увеличения мощности входных электрических колебаний с сохранением их формы и частоты за счет использования энергии внешнего источника питания, называется *электронным усилителем* (проще, *усилителем*). Минимальную часть усилителя, способную повышать мощность электрических колебаний, принято называть усилительным каскадом. Усилитель может содержать несколько последовательно соединенных усилительных каскадов, расположенных на одном полупроводниковом кристалле.

Структурная схема усилительного каскада (рис. 4.18) содержит усилительный элемент, к входным зажимам которого подключен источник усиливаемого сигнала $E_г$ с внутренним сопротивлением $R_г$, а к выходным — нагрузка (как правило, активная) $R_н$. Источником входного сигнала усилителя может быть предыдущий каскад, а нагрузкой — последующий каскад.

**Классификация и параметры усилителей.** Электронные усилители сигналов принято классифицировать по назначению, используемым усилительным элементам и режимам их работы, форме усиливаемых сигналов и полосе рабочих частот.

Все практические усилительные схемы усиливают мощность входного сигнала, однако в ряде случаев основным показателем усилителя являются величины выходных напряжения $U_{вых}$ (или $U_н$) или тока $I_{вых}$ (или $I_н$). Поэтому по назначению различают *усилители мощности, напряжения* и *тока*.

Усиление электрических сигналов небольшой мощности производится в основном усилителями на биполярных и полевых транзисторах, маломощных высокочастотных сигналов на транзисторах и туннельных диодах, а мощных высокочастотных сигналов усилителями на транзисторах, электронных лампах, лампах бегущей волны (ЛБВ) и амплитронах. Используются еще электромагнитные, магнитные и диэлектрические усилители, а для усиления слабых сигналов сантиметрового диапазона волн — малошумящие параметрические усилители. Следует сказать об оптических усилителях, а также молекулярных усилителях (ма-

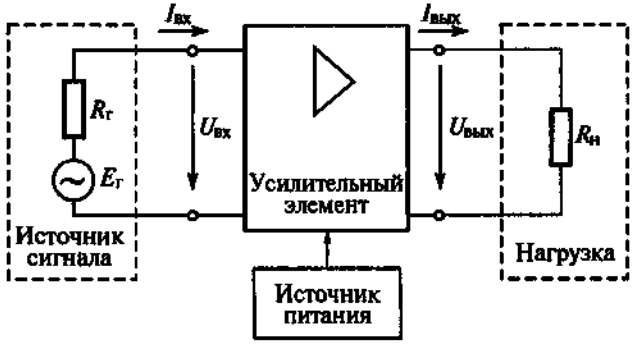



**Рис. 4.18.** Структурная схема усилительного каскада

зерах), создающие самые низкие уровни шумов из всех разработанных к настоящему времени малошумящих усилителей сигналов СВЧ-диапазона. Новейшая же электронная усилительная техника базируется в основном на линейных интегральных аналоговых и цифровых микросхемах.

Усилительные элементы в усилителях могут работать в линейном и нелинейном режимах. Если усилительный элемент работает в линейном режиме, то усилитель относят к классу линейных, считают *активным линейным четырехполюсником*, и на него распространяются все свойства линейных цепей. Если же усилительный элемент работает в нелинейном режиме, то усилитель является *нелинейным* устройством, хотя по форме выходной и входной сигналы практически всегда совпадают. Один из нелинейных режимов работы — *ключевой*.

По форме усиливаемых входных сигналов различают усилители *гармонических* и *импульсных сигналов*.

В зависимости от значения нижней граничной частоты усиливаемых сигналов усилители подразделяются на *усилители постоянного тока* (УПТ) и *переменного тока*. Современные УПТ способны усиливать и переменные сигналы, верхняя частота спектра которых достигает значения 100 МГц. Усилители переменного тока усиливают лишь гармонические составляющие сигнала в определенной полосе частот: от *нижней* $f_н$ до *верхней* $f_в$. Среди усилителей переменного тока различают *усилители низких* (*звуковых*) *частот* (УНЧ и соответственно УЗЧ), *усилители промежуточных частот* (УПЧ), *усилители высоких частот* (УВЧ), *усилители сверхвысоких частот* (УСВЧ), *узкополосные* (*избирательные*) и *широкополосные* (*импульсные*) *усилители*.

Усилителям низких частот свойственно усиление сигналов в частотном диапазоне от десятков герц до десятков килогерц. УПЧ применяются в диапазонах от сотен килогерц до десятков мегагерц. УВЧ и УСВЧ усиливают сигналы от сотен мегагерц до десятков гигагерц.

Избирательные усилители работают в узкой полосе частот — они делятся на *резонансные* и *полосовые усилители*. Отношение рабочей полосы частот к несущей частоте $\Delta f/f_0$ составляет в них 0,005...0,1. На выходе резонансных усилителей обычно включены колебательные контуры.

Усилители, ширина полосы *рабочих* частот которых соизмерима с ее средней частотой, относят к *широкополосным*. Они усиливают импульсные сигналы.

*Частотный коэффициент передачи* — основной параметр усилителя. Обычно входное и выходное значения сигнала являются однотипными, поэтому частотный коэффициент передачи называют *коэффициентом усиления*.

В зависимости от характера входной и выходной величин сигналов различают коэффициенты усиления по *напряжению* $K_U = U_{вых}/U_{вх}$, *току* $K_I = I_{вых}/I_{вх}$ и *мощности* $K_P = P_{вых}/P_{вх} = U_{вых}I_{вых}/(U_{вх}I_{вх}) = K_UK_I$. По существу, представленные коэффициенты усиления отражают АЧХ усилителя. Эти и другие параметры и характеристики усилителей будут введены в процессе анализа схем.

Усилительный каскад, работающий непосредственно на нагрузку, принято называть *выходным каскадом* (*усилителем мощности*). Остальные каскады усилителя являются фактически маломощными, и их относят к *входным* и *предварительным каскадам*.

### Усилительные каскады на биполярных транзисторах

**Биполярный транзистор**. В электронике широко используют биполярные транзисторы с двумя *p-n*-переходами, к которым относят этот термин. Биполярный транзистор разработан в 1947 г. в США лауреатами Нобелевской премии по физике (1956 г.) Уильямом Шокли (*William Shockley*; 1910–1989), Уолтером Браттейном (*Walter Brattain*; 1902–1987) и Джоном Бардиным (*John Bardeen*; 1908–1991; Бардин первый ученый, удостоившийся Нобелевской премии дважды; он получил премию вторично в 1972 г. вместе с Л. Купером и Д. Шриффером за создание теории сверхпроводимости, обычно называемой «БКШ-теорией»). Первые транзисторы имели ограниченные возможности (предельные рабочие частоты — сотни килогерц и мощности рассеяния порядка 100...200 мВт) и выполняли лишь некоторые функции электронных ламп. Напомним основные сведения о *биполярном транзисторе* (проще, *транзисторе*; слово «транзистор» образовано от англ. *transconductance* — преобразование проводимости и *varistor* — нелинейный резистор) — полупроводниковом приборе с двумя *p-n*-переходами и тремя электродами: *базой*, *коллектором* и *эмиттером*.

По характеру проводимости внешних слоев переходов транзисторы бывают *p-n-p-* и *n-p-n*-типов. Принцип действия транзисторов обоих типов идентичен. Отличие заключается в противоположности направлений протекания токов и полярности приложенных напряжений. Далее анализируются усилительные устройства, выполненные на транзисторах *n-p-n*-типа.

Из теории полупроводниковых приборов известно, что в биполярном транзисторе протекают токи базы $I_б$, коллектора $I_к$ и эмиттера $I_э$, связанные соотношением: $I_э = I_к + I_б$. Обычно $I_к = (0{,}9 \ldots 0{,}95)I_э$, т. е. практически весь ток, создаваемый эмиттером транзистора, протекает через коллекторный переход.

В зависимости от полярностей приложенных к *p-n*-переходам транзистора напряжений различают три режима работы:

- *отсечки* — оба *p-n*-перехода закрыты, через транзистор протекает сравнительно небольшой тепловой ток;
- *насыщения* — оба *p-n*-перехода полностью открыты, через транзистор протекает максимальный ток;
- *активный* — один из *p-n*-переходов открыт, к другому приложено обратное напряжение; через транзистор протекает управляемый ток.



У. Шокли



У. Браттейн



Д. Бардин

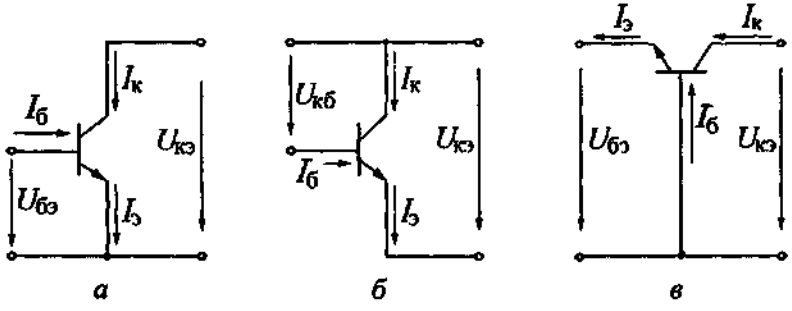



**Рис. 4.19.** Схемы включения транзистора:
*а* — с общим эмиттером; *б* — с общим коллектором; *в* — с общей базой

Активный режим является промежуточным между режимами насыщения и отсечки; он используется для усиления сигналов. В транзисторе возможен еще *инверсный* режим, который на практике не используют. Инверсное включение транзистора происходит при противоположной полярности подключаемых эмиттерного и коллекторного напряжений.

Транзистор эффективно управляется только в активном режиме. При этом управление усилительными свойствами осуществляется путем изменения тока базы. В исходном состоянии при $I_б = 0$ транзистор закрыт; через транзистор протекает тепловой ток коллектора $I_{к0}$, и поэтому считается, что токи $I_э$ и $I_к$ близки к нулю. При увеличении тока базы транзистор «приоткрывается», увеличивается количество носителей (электронов и дырок) в его переходах, в результате чего начинают протекать токи эмиттера $I_э$ и коллектора $I_к$. Если ток базы уменьшается (количество носителей в переходах снижается), то транзистор «призакрывается», и токи эмиттера и коллектора уменьшаются.

Транзисторы могут быть включены в усилительные и другие схемы тремя способами (рис. 4.19): с *общим эмиттером* (ОЭ), с *общим коллектором* (ОК) и с *общей базой* (ОБ). Различие в способах включения зависит от того, какой из трех электродов транзистора является по переменному току общим для входной и выходной цепей усилительного каскада.

Связь между токами и напряжениями в транзисторе характеризуют тремя системами параметров: это системы *z*-, *y*- и *h*-параметров. Выбор удобной для практических расчетов системы параметров зависит от схемы включения транзистора. В радиотехнических схемах широко используют устройства, в которых транзисторы включены по схеме ОЭ. При такой схеме включения для расчетов применяют *h*-параметры, экспериментально определяемые по статическим *входным* (*базовым*) и *выходным* (*коллекторным*) *вольт-амперным характеристикам* (ВАХ) транзистора (ВАХ — зависимость напряжения на зажимах элемента электрической цепи от тока в нем).

На рис. 4.20 представлены типовые экспериментальные входные и выходные статические характеристики транзистора, включенного по схеме ОЭ.

Входные характеристики транзистора отражают зависимость входного тока (тока базы $I_б$) от входного напряжения (напряжения база–эмиттер $U_{бэ}$) при по-

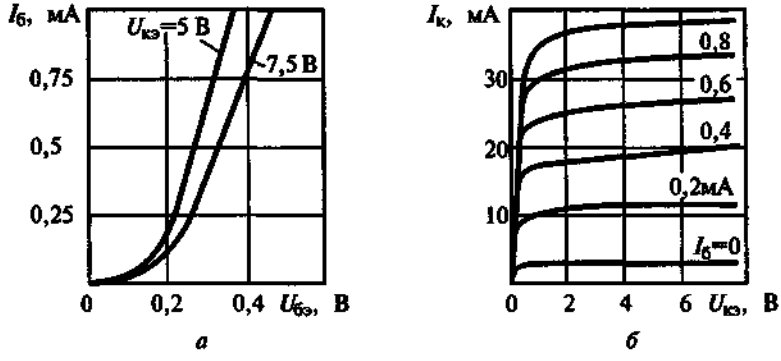



**Рис. 4.20.** Статические характеристики транзистора с ОЭ:
*а* — входные; *б* — выходные

стоянном напряжении коллектор–эмиттер $U_{кэ}$. Выходные характеристики представляют собой зависимость выходного тока (тока коллектора $I_к$) от напряжения коллектор–эмиттер $U_{кэ}$ при постоянном токе базы $I_б$.

В теории усилителей транзистор, включенный по схеме с ОЭ, часто рассматривают как линейный активный четырехполюсник. В этом случае для приращений токов и напряжений в транзисторе справедливы равенства:

$$\Delta U_{бэ} = h_{11}\Delta I_б + h_{12}\Delta U_{кэ}; \quad \Delta I_к = h_{21}\Delta I_б + h_{22}\Delta U_{кэ}, \tag{4.69}$$

где $h_{11} = \Delta U_{бэ}/\Delta I_б$ — входное сопротивление транзистора при $\Delta U_{кэ}$ = const;
$h_{12} = \Delta U_{бэ}/\Delta U_{кэ}$ — коэффициент внутренней обратной связи при $I_б$ = const;
$h_{21} = \Delta I_к/\Delta I_б$ — коэффициент передачи тока при $\Delta U_{кэ}$ = const;
$h_{22} = \Delta I_к/\Delta U_{кэ}$ — выходная проводимость транзистора при $I_б$ = const.

Эти параметры транзистора определяют по его статическим характеристикам.

**Пример 4.6.** Используя базовую (входную) характеристику транзистора при $U_{кэ}$=5 В (рис. 4.20, *а*), определить его входное сопротивление.

Решение. Выбрав величину приращения напряжения база–эмиттер транзистора $\Delta U_{бэ} \approx 0{,}45 - 0{,}25 = 0{,}2$ В, находим по кривым входных ВАХ приращение его тока базы $\Delta I_б \approx 1 - 0{,}25 = 0{,}75$ мА. Тогда входное сопротивление транзистора:

$$h_{11} = \Delta U_{бэ}/\Delta I_б \approx 0{,}2/(0{,}75\cdot 10^{-3}) = 265 \text{ Ом}.$$

Усилительный каскад ОЭ. Полупроводниковая техника на *дискретных* компонентах содержит ряд вариантов усилительного каскада ОЭ. Принцип действия усилительных каскадов ОЭ рассмотрим на примере схемы на рис. 4.21. На входе каскада действуют усиливаемые переменные ток $i_{вх}$ и напряжение $u_{вх}$, а на выходе — усиленные переменные ток $i_н$ и напряжение $u_{вых}$ (здесь и далее аргумент $t$ в функциях токов и напряжений для упрощения опущен). В схеме конденсаторы $C_1$ и $C_2$ — разделительные. Конденсатор $C_1$ препятствует протеканию постоянного тока от источника питания $E_к$ в цепь источника входного сигнала. Конденсатор $C_2$ обеспечивает выделение из коллекторного напряжения переменной составляющей, поступающей на резистор нагрузки $R_н$.

Резисторы базового делителя напряжения $R_1$, $R_2$ задают *режим покоя* транзистора, при котором в нем протекают только постоянные *токи покоя* базы $I_{бп}$, коллектора $I_{кп}$ и эмиттера $I_{эп}$, а на его базе, коллекторе и эмиттере соответственно действуют постоянные *напряжения покоя* $U_{бп}$, $U_{кп}$ и $U_{эп}$. Заметим, что в транзисторных схемах один полюс источника питания часто соединяют с корпусом устройства, который обозначают ⊥. В схеме рис. 4.21 такой полюс — потенциал источника питания $-E_к$.

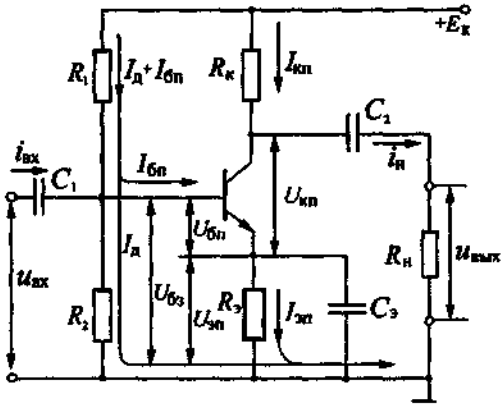



**Рис. 4.21.** Схема усилительного каскада ОЭ

Резистор $R_э$ и делитель $R_1$, $R_2$ составляют цепь *отрицательной обратной связи* (ООС), предназначенную для термостабилизации режима покоя транзистора при изменении его температуры. Под *обратной связью* (ОС) понимают процесс передачи части выходного сигнала на вход усилителя. Действие ООС объясняется следующим образом. При увеличении, например, из-за роста температуры тока коллектора покоя $I_{кп}$ возрастают ток эмиттера покоя $I_{эп}$ и падение напряжения на резисторе $R_э$, поскольку $U_{эп} = I_{эп} R_э$. Поскольку напряжение между базой и землей (база–земля) $U_{бэ}$ фиксировано базовым делителем $R_1$, $R_2$ и $U_{бэ} = U_{бп} + U_{эп}$, то с увеличением напряжения $U_{эп}$ уменьшается напряжение $U_{бп}$. Это приводит к призакрыванию транзистора, уменьшению тока базы покоя $I_{бп}$ и, следовательно, снижению тока коллектора покоя $I_{кп}$. Тем самым производится компенсация первоначального увеличения тока коллектора покоя.

Включение резистора $R_э$ в цепь эмиттера изменяет работу каскада и по переменному сигналу. Переменный ток эмиттера $i_э$ создает на резисторе $R_э$ падение напряжения $u_э = i_э R_э$, которое уменьшает усиливаемое напряжение, подводимое к базе транзистора, ведь $u_{бэ} = u_{вх} - u_э$. При этом снижается и коэффициент усиления каскада, поскольку действует ООС по переменному току. Для ее исключения резистор $R_э$ шунтируют конденсатором $C_э$ достаточно большой емкости. Поскольку реактивное сопротивление конденсатора мало, то переменный ток протекает по нему и не создает потерь напряжения на резисторе $R_э$.

В режиме покоя транзистора расчет каскада по постоянному току (входной сигнал отключен) проводят графоаналитическим методом с использованием статических выходных и входных ВАХ транзистора. Метод удобен при нахождении связи параметров режима покоя ($U_{кп}$ и $I_{кп}$) с амплитудными значениями переменных составляющих — выходного напряжения $U_{выхm}$ и коллекторного тока $I_{кm}$ (рис. 4.22). При расчетах на выходных характеристиках (рис. 4.22, *а*) проводят *линию нагрузки по постоянному току* (линия 1–2), положение которой определяется вторым законом Кирхгофа для коллекторной цепи каскада

$$E_к = U_{кп} + I_{кп}(R_к + R_э). \qquad (4.70)$$

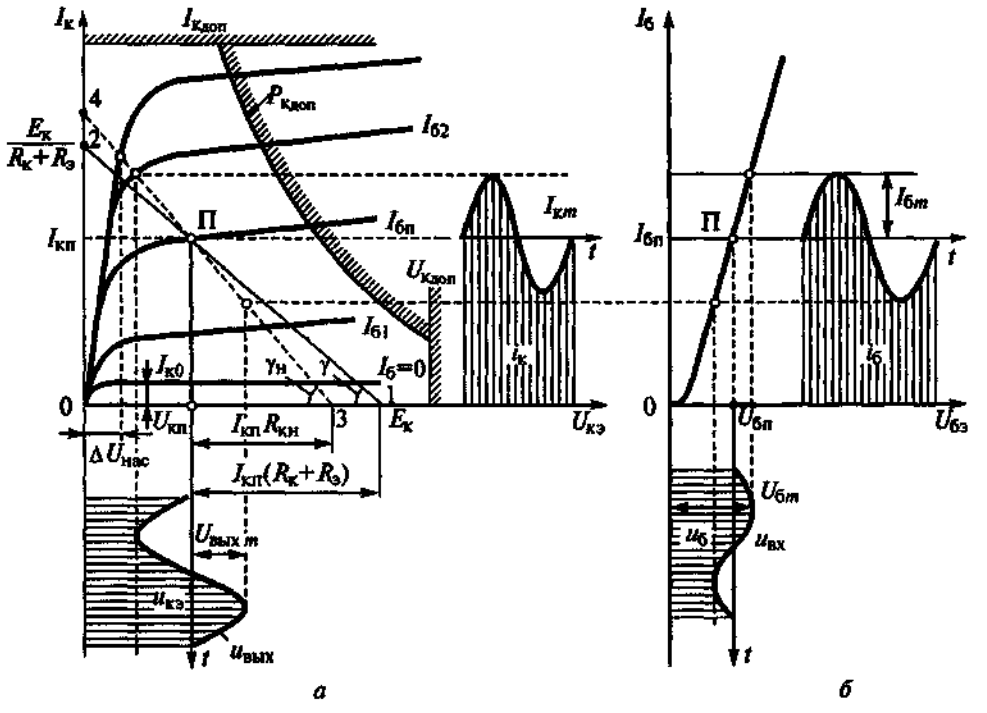



**Рис. 4.22.** Графический анализ каскада ОЭ по характеристикам транзистора: *а* — выходным; *б* — входной

Линию нагрузки можно провести из точки $E_к$ под углом $\gamma = \text{arctg}(R_к + R_э)$ (см. рис. 4.22, *а*), на практике ее строят по двум точкам, характеризующим режимы *холостого хода* (точка 1) и *короткого замыкания* (точка 2) в коллекторной цепи транзистора. Для точки 1: ток и напряжение холостого хода $I_{кх} = 0$, $U_{кх} = E_к$; для точки 2: напряжение и ток короткого замыкания $U_{кз} = 0$; $I_{кз} = E_к/(R_к + R_э)$.

При расчетах любые значения тока $I_{кп}$ и напряжения $U_{кп}$ определяют точками пересечений (*рабочими точками*) выходных характеристик с линией нагрузки по постоянному току. Одну из этих точек, полученную для заданного тока базы покоя $I_{бп}$, называют *точкой покоя* и обозначают буквой П (см. рис. 4.22). Используя координаты точки покоя П, можно определить ток коллектора покоя $I_{кп}$, напряжение коллектора покоя $U_{кп}$ и падение напряжения на резисторе $R_к$, равное $U_{R_к} = I_{кп}R_к$. При этом транзистор работает в активном режиме.

Для определения параметров выходного сигнала в динамическом режиме усиления (с подключенными входным сигналом и нагрузкой) используют *линию нагрузки по переменному току* (*динамическую линию нагрузки*). Если учесть, что сопротивления источника питания $E_к$ и конденсатора $C_2$ по переменному току малы, то сопротивление нагрузки по переменному току будет определяться параллельно включенными резисторами $R_к$ и $R_н$

$$R_{кн} = \frac{R_к R_н}{R_к + R_н}. \qquad (4.71)$$

В режиме усиления сигнала токи и напряжения транзистора состоят из суммы постоянных и переменных составляющих, поэтому линия нагрузки по переменному току также пройдет через точку П. И поскольку $R_{кн} < R_к$, то линия будет находиться под углом $\gamma_н = \text{arctg}R_{кн}$, бóльшим, чем угол $\gamma$. Для ее построения на оси абсцисс отмечают точку 3, где напряжение равно сумме $U_{кп} + I_{кп}R_{кн}$, и через нее и точку П проводят прямую (рис. 4.22, *а*, штриховая линия 3–4).

Вначале принцип действия каскада ОЭ рассмотрим при отключенной нагрузке $R_н$ (*режим холостого хода по переменному току*). При подаче на вход каскада переменного напряжения $u_{вх}$ переменный ток базы $i_б$ будет изменяться в соответствии с входной характеристикой (см. рис. 4.22, *б*). Одновременно и по такому же закону станет менять свои значения переменный ток коллектора. Так, например, при увеличении амплитуды входного напряжения возрастет ток базы $i_б$. Поскольку ток коллектора связан с током базы как $i_к = h_{21}i_б$ ($h_{21}$ составляет 50...75), то он тоже возрастет. В результате увеличивается падение переменного напряжения на резисторе $R_к$ (ведь $U_{R_к} = i_кR_к$), а переменное напряжение на коллекторе $u_{кэ} = u_{вых} = E_к - i_кR_к$ уменьшится. При уменьшении же входного напряжения картина меняется на обратную. Из данного анализа следует, что каскад ОЭ наряду с усилением мощности изменяет (инвертирует; инверсия — такое действие над входным сигналом, при котором все составляющие его спектра изменяют фазу на 180°) фазу входного сигнала на 180° (см. рис. 4.22). Точно так же работает схема и при подключении нагрузки $R_н$, лишь переменный коллекторный ток при этом распределяется между резисторами $R_к$ и $R_н$, что естественно снижает усиление.

При использовании каскада ОЭ для усиления мощности необходимо учитывать параметры *предельно допустимых режимов* работы транзистора. Таких параметров три, и их отмечают на выходных характеристиках (см. рис. 4.22, *а*). Кривая допустимой мощности рассеяния строится по формуле $P_{к\,доп} = U_{кэ}I_к$ и представляет собой гиперболу, а линии допустимых коллекторного тока $I_{к\,доп}$ и напряжения коллектор–эмиттер $U_{к\,доп}$ — прямые, параллельные осям координат. Для исключения искажений формы выходного сигнала надо обеспечить такой режим работы транзистора, чтобы рабочая точка, перемещаясь по линии нагрузки, не выходила за пределы *напряжения насыщения* ($\Delta U_{нас} = 0{,}3...0{,}7$ В).

Основные показатели усилительного каскада ОЭ рассчитывают с помощью *h-параметров* транзистора, используя эквивалентную схему каскада (рис. 4.23), основа которого — схема замещения транзистора без учета емкостей переходов (обведена штриховой линией). В упрощенной схеме замещения транзистор представляют активным линейным четырехполюсником, на входе которого действуют напряжение $U_{вх}$ и ток $I_{вх}$, а на выходе — напряжение $U_{вых}$ и ток $I_н$.

Указанные величины представлены *действующими значениями*, связанными с известными *амплитудными* фор-

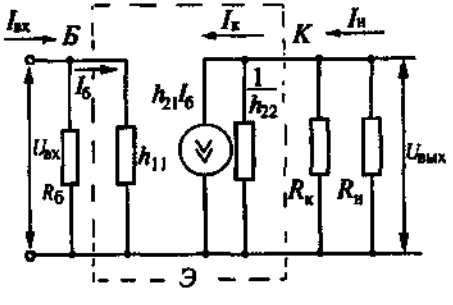



Рис. 4.23. Эквивалентная схема каскада ОЭ

мулами: $U = U_m/\sqrt{2}$; $I = I_m/\sqrt{2}$. В схеме резистор $h_{11}$ отражает входное сопротивление, эквивалентный генератор тока $h_{21}I_б$ — усилительные свойства, а сопротивление $1/h_{22}$ — величину, обратную выходной проводимости транзистора. В эквивалентной схеме не показаны конденсаторы и источник питания, так как их сопротивления по переменному току близки к нулю. Поэтому резисторы $R_к$ и $R_н$ включены непосредственно между эмиттером и коллектором. Сопротивление $R_б = R_1 \| R_2$ отражает наличие базового делителя, резисторы $R_1$, $R_2$ которого по переменному току соединены параллельно. Формулы для расчета сопротивлений $R_1$ и $R_2$ нетрудно получить из схемы на рис. 4.21

$$R_1 = \frac{E_к - U_{бэ}}{I_д + I_{бп}};\ R_2 = \frac{U_{бэ}}{I_д}. \qquad (4.72)$$

*Входное сопротивление* каскада при $R_б >> h_{11}$

$$R_{вх} = U_{вх}/I_{вх} = R_б\, h_{11}/(R_б + h_{11}) \approx h_{11}. \qquad (4.73)$$

*Выходное сопротивление* с учетом неравенства $R_к << 1/h_{22}$

$$R_{вых} = \frac{R_к(1/h_{22})}{R_к + 1/h_{22}} \approx R_к. \qquad (4.74)$$

*Коэффициент усиления по напряжению*

$$K_U = \frac{U_{вых}}{U_{вх}} = h_{21}\frac{R_{кн}}{h_{11}}. \qquad (4.75)$$

*Коэффициент усиления по току*

$$K_I = I_н/I_{вх} \approx I_к/I_б \approx h_{21}. \qquad (4.76)$$

Для усилителя важна *амплитудная характеристика* — зависимость выходного напряжения от входного (рис. 4.24). У идеального усилителя амплитудная характеристика — прямая линия, проходящая через начало координат. Угол ее наклона пропорционален коэффициенту усиления $K_U$. Обычно амплитудная характеристика имеет изгиб и пересекает ось ординат в точке $U_{вых} = U_ш$, определяющейся напряжением собственных шумов усилителя. Участок $U_{вх} < U_{вх\ мин}$ не используют, так как усиливаемый сигнал здесь не различим на фоне внутренних шумов усилителя.

Изгиб амплитудной характеристики при амплитудах $U_{вх} > U_{вх\ макс}$ отражает появление искажений формы выходного сигнала. Без искажений усиливаются входные сигналы с амплитудой не выше $U_{вх\ макс}$ и не ниже $U_{вх\ мин}$, отношение которых представляет *динамический диапазон* усилителя [дБ]

$$D_{дБ} = 20\lg(U_{вх\ макс}/U_{вх\ мин}). \qquad (4.77)$$

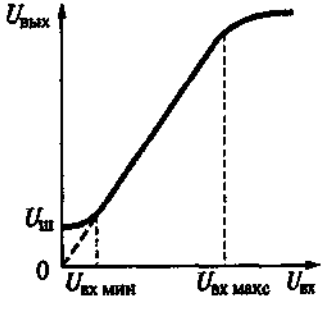



**Рис. 4.24.** Амплитудная характеристика усилителя

Линейное усиление сигнала обеспечивается при небольших амплитудах входного напряжения и выборе точки покоя на линейных участках входной и выходной характеристик транзистора. Тогда имеет место линейная зависимость между переменными токами базы $i_{б1}$ и коллектора $i_{к1}$, а также напряжениями $u_{вх1}$ и $u_{вых1}$ (см. штриховые линии на рис. 4.22). Если же амплитуда входного сигнала велика, то нелинейность характеристик приводит к искажениям формы выходного напряжения.

Искажения формы выходного напряжения относят к *нелинейным*. Уровень нелинейных искажений усиливаемого сигнала оценивают *коэффициентом гармоник* (*коэффициентом нелинейных искажений*)

$$K = \sqrt{\frac{P_2 + P_3 + \ldots + P_n}{P_1}} = \frac{\sqrt{U_2^2 + U_3^2 + \ldots + U_n^2}}{U_1} = \frac{\sqrt{I_2^2 + I_3^2 + \ldots + I_n^2}}{I_1}, \qquad (4.78)$$

где $P_1, P_2, \ldots, P_n, U_1, U_2, \ldots, U_n, I_1, I_2, \ldots, I_n$ — мощности, напряжения и токи соответственно 1-, 2-, ..., $n$-й гармонических составляющих выходного сигнала.

Нелинейные искажения в отличие от линейных сопровождаются появлением на выходах усилителей новых, паразитных гармонических составляющих в спектре усиленного сигнала, частоты которых кратны частотам входного сигнала. Однако (это принципиально) если линейные искажения влияют на качество передачи информации только в «своем» канале, то нелинейные искажения, кроме того, могут исказить информацию, передаваемую по другим каналам. Дело в том, что возникающие при нелинейных искажениях высшие гармоники одного передаваемого сигнала могут попадать в спектры других сигналов и искажать их. В общем случае коэффициент гармоник усилителя не должен превышать 1...5 %. Высококачественные усилители звука имеют коэффициент гармоник менее 1%.

Усилительный каскад ОК. Коллектор транзистора в схеме усилительного каскада ОК по переменному току заземлен (т. е. соединен с корпусом) через источник питания $E_к$ (рис. 4.25). При этом входное напряжение подключено между базой и коллектором, а выходное — снимается непосредственно с эмиттера транзистора. Режим работы усилительного каскада ОК по постоянному току определяется резистором $R_э$. Делитель напряжения $R_1$, $R_2$ и разделительные конденсаторы $C_1$, $C_2$ выполняют те же функции, что и в каскаде ОЭ. При расчете каскада ОК по постоянному току используют графоаналитический метод по аналогии со схемой ОЭ. Линию нагрузки по постоянному току в этом случае проводят также через две точки (см. рис. 4.22): 1 — $I_{кх} = 0$, $U_{кх} = E_к$; 2 — $U_{кз} = 0$, $I_{кз} = E_к / R_э$.

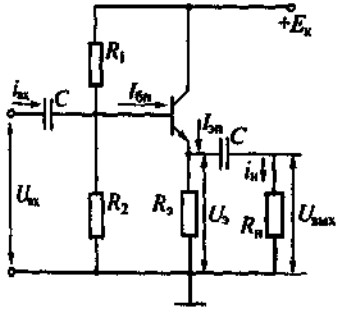



Рис. 4.25. Усилительный каскад ОК

Амплитуда выходного напряжения усилительного каскада меньше амплитуды входного, поскольку $U_{вых} = U_{вх} - U_{бэ}$. Поэтому *коэффициент усиления по напряжению*

$$K_U = \frac{U_{вых}}{U_{вх}} = \frac{U_{вх} - U_{бэ}}{U_{вх}} = 1 - \frac{U_{бэ}}{U_{вх}}. \qquad (4.79)$$

Как правило, $U_{бэ} << U_{вх}$, следовательно, $K_U \approx 1$, а $U_{вых} \approx U_{вх}$.

*Сопротивление нагрузки по переменному току*

$$R_{эн} = R_э R_н /(R_э + R_н). \quad (4.80)$$

Выходное напряжение в каскаде ОК совпадает с входным по фазе. Так, при поступлении положительного приращения входного напряжения ток базы увеличивается, вызывая возрастание токов коллектора и эмиттера. Это приводит к увеличению падения переменного напряжения на сопротивлении $R_{эн}$, с которого снимается выходное напряжение. При подаче же отрицательного приращения входного напряжения выходное напряжение также получит отрицательное приращение. Таким образом, выходной сигнал повторяет входной сигнал и по амплитуде, и по фазе, поэтому усилительный каскад ОК называют *эмиттерным повторителем*.

*Коэффициент усиления по току* эмиттерного повторителя

$$K_I = I_н/I_{вх} \approx I_э/I_б \approx h_{21} + 1. \quad (4.81)$$

*Входное сопротивление*

$$R_{вх} = \frac{U_{вх}}{I_{вх}} = \frac{U_{бэ} + U_{вых}}{I_б} = \frac{U_{бэ}}{I_б} + \frac{I_э R_{эн}}{I_б} = h_{11} + (h_{21} + 1)R_{эн}. \quad (4.82)$$

*Выходное сопротивление* при $h_{21} >> 1$:

$$R_{вых} = U_{вых}/I_н \approx U_{вх}/I_э = h_{11}/(h_{21} + 1) \approx h_{11}/h_{21}. \quad (4.83)$$

Достаточно высокое входное (до 1 МОм и выше) и сравнительно малое выходное (десятки ом) сопротивления позволяют использовать эмиттерный повторитель в качестве усилителя мощности для согласования высокоомного (маломощного) генератора с низкоомной нагрузкой.

Усилительный каскад ОБ. Усиливаемый сигнал в усилительном каскаде ОБ (рис. 4.26) включен между эмиттером и базой, причем последняя по переменному току соединена с корпусом через конденсатор $C_б$. Назначение резисторов $R_1$, $R_2$, $R_к$ и конденсаторов $C_1$, $C_2$ такое же, как в каскаде ОЭ. Усилительный каскад ОБ не инвертирует фазу входного сигнала. Воздействие, например, входной отрицательной полуволны приводит к увеличению и коллекторного тока и падения напряжения на резисторе $R_к$. В результате коллекторное (выходное) напряжение уменьшается.

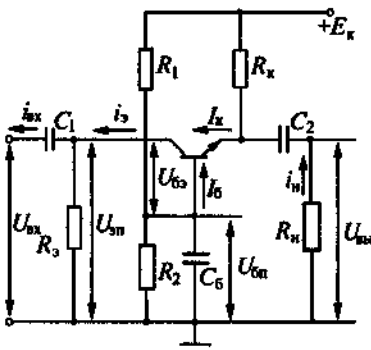



**Рис. 4.26.** Усилительный каскад с ОБ

При воздействии положительной полуволны входного напряжения картина меняется на обратную. Транзисторный каскад ОБ имеет почти такой же коэффициент усиления по напряжению, как и каскад ОЭ, но коэффициент усиления по току меньше единицы, так как входным является эмиттерный, а выходным — коллекторный ток. Каскад ОБ применяют в усилителях и генераторах дециметровых и сантиметровых волн.

Это связано с тем, что при включении транзистора по схеме ОБ его предельная частота усиления в $h_{21}$ + 1 раз больше, чем в схеме с ОЭ. С повышением частоты входного сигнала коэффициент усиления транзисторов снижается.

### Усилительные каскады на полевых транзисторах

**Полевой транзистор.** Первые полевые транзисторы разработали в 1952 г. американцы У. Шокли, С. Мид, Д. Канг, М. Аталлай, хотя патентование и их создание велись с 1926 г. Первый промышленный образец полевого транзистора был внедрен французом С. Тешнером. В отличие от биполярных полевые транзисторы имеют высокое входное сопротивление (до десятков мегаом). Поэтому и усилительные каскады на полевых транзисторах обладают большими входными сопротивлениями. Управление током осуществляют изменением проводимости канала *p*- или *n*-типа, через который под воздействием электрического поля он протекает, т. е. полевые транзисторы управляются по входной цепи напряжением (электрическим полем — отсюда название — «полевые»), а не током. Электропроводность канала обусловлена движением носителей только одного типа, поэтому по принципу действия они являются *униполярными*.

По способу создания проводящего канала различают *полевые транзисторы с p-n-переходом* и *с встроенным* и *индуцированным каналами*. Последние два называют МДП-транзисторами (металл – диэлектрик – полупроводник).

Большинство полевых транзисторов (рис. 4.27, *а*, *б*) имеют три электрода (вывода): *сток*, *исток* и *затвор* (от англ. *gate* — ворота). Включение полевых транзисторов (как и биполярных) в усилительный каскад осуществляют по трем схемам: с *общим истоком* (ОИ), *общим стоком* (ОС) и *общим затвором* (ОЗ).

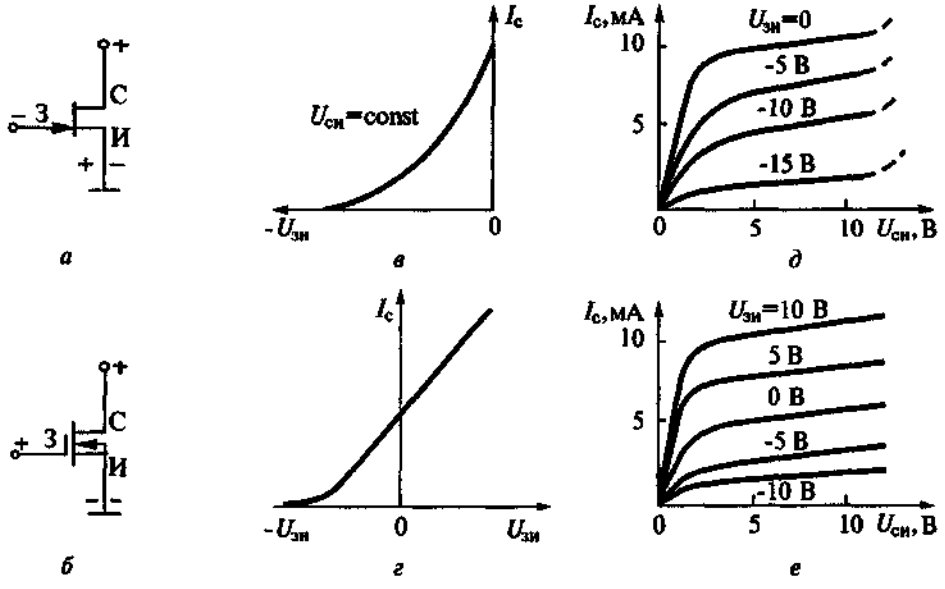



**Рис. 4.27.** Полевые транзисторы с каналами *n*-типа:
*а*, *б* — условные обозначения; *в*, *г* — стоко-затворные характеристики;
*д*, *е* — стоковые характеристики соответственно транзисторов с *p-n*-переходом и МДП-типа

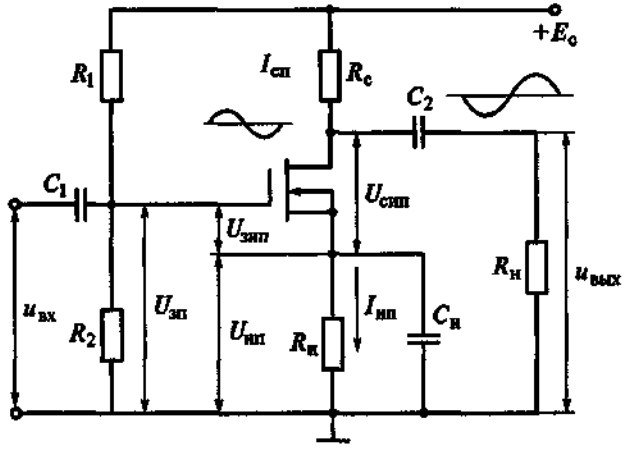



**Рис. 4.28.** Усилительный каскад ОИ на МДП-транзисторе

Запитывают полевые транзисторы по стоковой цепи разными напряжениями, что зависит от типа канала: полевые транзисторы с каналом *n*-типа питают положительным напряжением $+E_c$, *p*-типа — отрицательным напряжением $-E_c$.

Для расчета схем на полевых транзисторах используют статические ВАХ. *Стоко-затворная* (*входная*) характеристика — это зависимость тока стока $I_c$ МДП-транзистора от напряжения затвор–исток $U_{зи}$ при постоянном напряжении сток–исток $U_{си}$ (рис. 4.27, *в*, *г*). *Стоковые* (*выходные*) *характеристики* МДП-транзистора отражают зависимость тока стока $I_c$ от напряжения сток–исток $U_{си}$ при постоянном напряжении затвор–исток $U_{зи}$ (рис. 4.27, *д*, *е*).

Рассмотрим усилители на МДП-транзисторах с каналами *n*-типа.

Усилительный каскад ОИ. На рис. 4.28 показан *усилительный каскад* ОИ на МДП-транзисторе, в которой назначение резисторов $R_1$, $R_2$ и конденсаторов $C_1$ и $C_2$ такое же, как и в каскаде ОЭ. В схеме резистор $R_и$ выполняет функцию термостабилизации режима работы каскада по постоянному току стока $I_{сп}$. Для исключения ООС по переменному току резистор $R_и$ шунтируют конденсатором $C_и$. Нагрузкой по переменному току в каскаде является сопротивление

$$R_{сн} = R_c R_и/(R_c + R_и). \qquad (4.84)$$

Анализ работы каскада ОИ проводят графоаналитическим методом, используя статические стоковые и стоко-затворные характеристики (рис. 4.29). *Линию нагрузки по постоянному току* (линия 1–2 на рис. 4.29, *а*) проводят через две точки в соответствии со вторым законом Кирхгофа для стоковой цепи:

$$E_c = U_{сип} + I_{сп}(R_c + R_и). \qquad (4.85)$$

Тогда координаты определятся следующим образом — для точки 1 — $I_{сх} = 0$, $U_{сх} = E_c$; для точки 2 — $U_{сз} = 0$, $I_{сз} = E_c/(R_c + R_и)$.

*Линия нагрузки по переменному току* проходит через точку покоя П (линия 3 – 4 на рис. 4.29, *а*) и ее наклон к оси напряжений определяется величиной сопротивления $R_{сн}$. В *режиме холостого хода по переменному току* при подаче на вход каскада переменного напряжения $u_{вх}$ переменный ток стока $i_c$ будет изменяться в соответствии с стоко-затворной характеристикой (рис. 4.29, *б*).

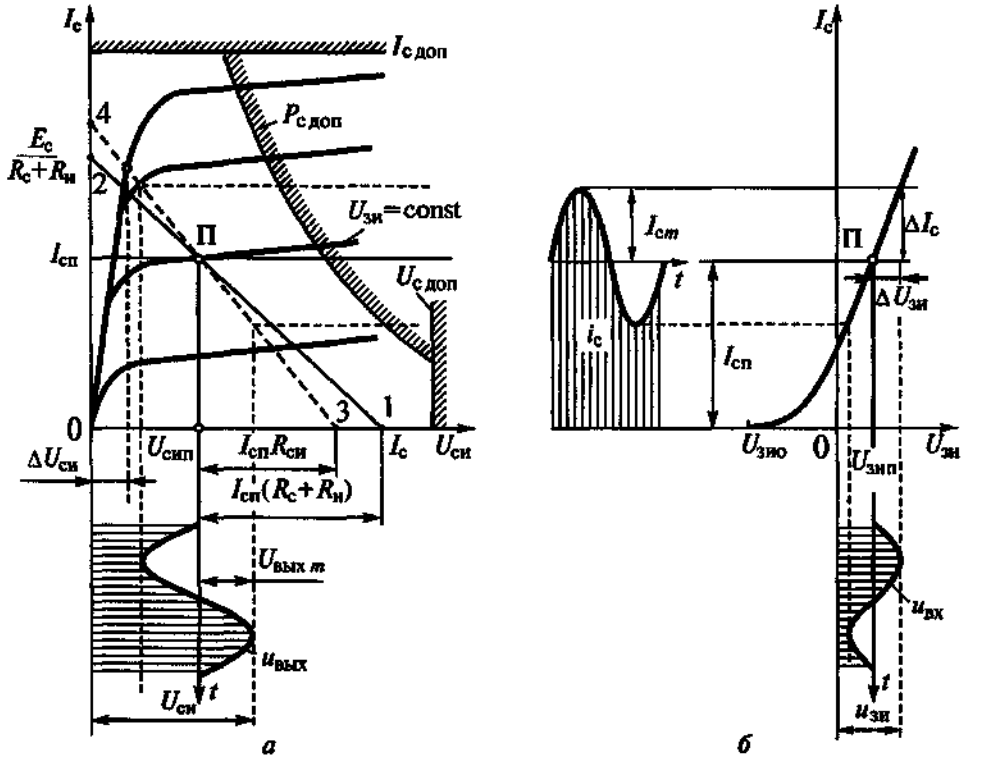



**Рис. 4.29.** Графический анализ каскада ОИ по характеристикам МДП-транзистора: *а* — стоковым; *б* — стоко-затворной

Усилительный каскад ОИ аналогично каскаду ОЭ изменяет фазу входного сигнала на 180°. Основные параметры каскада ОИ в линейном режиме усиления рассчитывают с помощью *эквивалентной схемы*, основа которой — *схема замещения* МДП-транзистора (обведена на рис. 4.30 штриховой линией). В эквивалентной схеме усилительные свойства МДП-транзистора отражены генератором тока $SU_{вх}$ с параллельно включенным внутренним сопротивлением транзистора $r_i$. *Крутизна стоко-затворной характеристики* $S = \Delta I_c/\Delta U_{зи}$, определяемая с помощью кривой на рис. 4.29, *б*. Делитель в цепи затвора представлен сопротивлением $R_з = R_1 \| R_2$, а нагрузка по переменному току — сопротивлением $R_{си}$. Межэлектродные емкости $C_{зи}$ и $C_{зс}$ отражают емкости *p-n*-переходов, а емкость $C_{си}$ — межэлектродную выходную емкость транзистора.

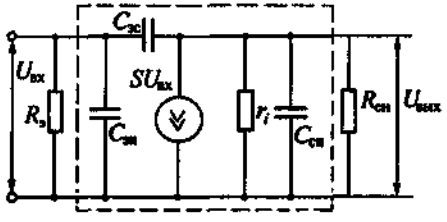



**Рис. 4.30.** Эквивалентная схема каскада ОИ

*Коэффициент усиления по напряжению* находят из эквивалентной схемы

$$K_U = \frac{U_{вых}}{U_{вх}} = \frac{SU_{вх}(r_i \| R_{си})}{U_{вх}} = S(r_i \| R_{си}).$$

Как правило, в полевых транзисторах $r_i >> R_c$ и $r_i >> R_{сн}$, поэтому

$$K_U = SR_{сн}. \quad (4.86)$$

*Входное сопротивление* каскада ОИ определяется в основном делителем в цепи затвора $R_{вх} = R_з$.

*Выходное сопротивление* каскада приблизительно равно стоковому сопротивлению $R_{вых} \approx R_c$.

Усилительный каскад ОС. В простейшей типовой схеме усилительного каскада с ОС (*истоковом повторителе*) на МДП-транзисторе (рис. 4.31) резисторы $R_1$, $R_2$ и $R_и$ задают режим покоя. Нагрузкой усилительного каскада по постоянному току служит резистор $R_и$, а по переменному току — сопротивление параллельно включенных резисторов $R_н$ и $R_и$: $R_{ин} = R_и R_н/(R_и + R_н)$. Как и в каскаде ОК на биполярном транзисторе, выходное напряжение в истоковом повторителе совпадает по фазе с входным и практически равно ему.

*Коэффициент усиления по напряжению*

$$K_U = \frac{U_{вых}}{U_{вх}} = \frac{SR_{ин}}{1 + SR_{ин}}. \quad (4.87)$$

Поскольку $SR_{ин} >> 1$, то коэффициент усиления близок к единице.

Входное сопротивление истокового повторителя достигает сотен мегаом. Во-первых, это связано с малым значением входной емкости МДП-транзистора. Поскольку реактивное сопротивление этой емкости достаточно велико, то оно почти не шунтирует входную цепь каскада. Во-вторых, это обусловлено тем, что между затвором и истоком приложена разность напряжений $U_{зи} = U_{вх} - U_{вых}$, которая невелика. Вследствие высокого входного сопротивления входной ток МДП-транзистора оказывается очень малым и мощность, отбираемая от генератора, невелика.

Выходное сопротивление мало, $R_{вых} \approx 1/S$.

### Усилители мощности (выходные каскады усилителей)

Выходные каскады усилителей предназначены для получения в низкоомной нагрузке требуемой мощности сигнала и поэтому их характеризуют рядом энергетических параметров: выходной мощностью, КПД, коэффициентом усиления по мощности и уровнем нелинейных искажений. Для обеспечения высоких энергетических показателей усилителя мощности амплитуды выходных напряжений и токов, а также мощность усиленного сигнала должны быть близки к предельно допустимым параметрам используемого транзистора.

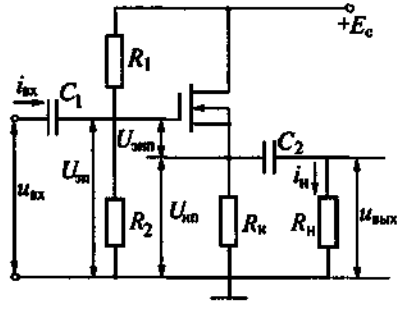



**Рис. 4.31.** Истоковый повторитель

В зависимости от положения точки покоя на линии нагрузки по постоянному току различают три основных *режима* (ранее, *класс*) работы транзисторов в усилителях мощности: *A*, *B* и *AB*.

Применяют также специфические режимы *C*, *D* (близкие к ключевому) и *ключевой* — импульсный, которые здесь не рассматриваются.

В режиме *A* точку покоя транзистора на выходных характеристиках выбирают так, чтобы рабочая точка при перемещении по линии нагрузки не попала в области искажений формы выходного сигнала. Таким образом, все рассмотренные усилительные каскады работают в режиме *A*.

Чтобы определить энергетические параметры усилителя мощности, обратимся к рис. 4.22, *а*. Из графических построений находим:

• *мощность в коллекторной цепи каскада ОЭ*

$$P_к = 0{,}5U_{выхm}I_{кm}; \tag{4.88}$$

• *мощность, потребляемую от источника питания*:

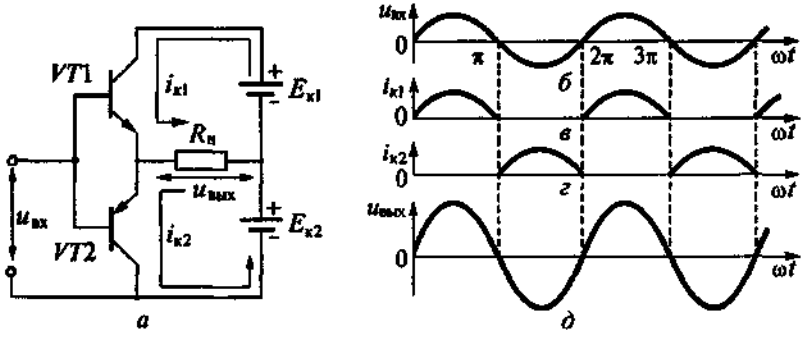
$$P_0 = E_к I_{кп}; \tag{4.89}$$

• *КПД коллекторной цепи*

$$\eta_к = \frac{P_к}{P_0} = 0{,}5\frac{U_{выхm}I_{кm}}{E_к I_{кп}}. \tag{4.90}$$

Как следует из (4.90), даже при максимальных амплитудах напряжения и тока ($U_{выхm} = E_к$ и $I_{кm} = I_{кп}$) КПД транзисторного усилителя мощности, работающего в режиме *A*, не превышает 50 %.

По способу подключения нагрузки выходные каскады усилителей мощности делят на *трансформаторные* и *бестрансформаторные*. Трансформаторные каскады в современных усилительных устройствах почти не применяют. В бестрансформаторных выходных каскадах используют однотипные и разнотипные транзисторы, соединенные по двухтактной схеме с непосредственным подключением нагрузки (рис. 4.32). При этом разнотипные транзисторы имеют идентичные параметры и называются *комплементарными*. Схемы на однотипных транзисторах применяют крайне редко. Их недостаток — один транзистор включен по схеме ОЭ, а другой — по схеме ОК, что требует искусственного выравнивания коэффициентов усиления плеч.




*а*

**Рис. 4.32.** Бестрансформаторный усилитель на комплементарных транзисторах:
*а* — схема; *б* – *д* — временны́е диаграммы токов и напряжений

К входу бестрансформаторного каскада на комплементарных транзисторах (рис. 4.32, *а*) подводят однофазное напряжение $u_{вх}$ (для упрощения опущен аргумент $v = \omega t$). Транзисторы включены по схеме эмиттерного повторителя и запитываются от разнополярных одинаковых источников питания $E_{к1} = |-E_{к2}|$. Нагрузка подключена к общей точке соединения эмиттеров транзисторов.

С помощью временны́х диаграмм токов и напряжений (рис. 4.32, *б*– *д*) рассмотрим принцип действия бестрансформаторного выходного каскада. Транзисторы в схеме работают попеременно в режиме *В*. Например, на угловом интервале 0 – $\pi$ при положительной полуволне входного тестового гармонического напряжения (см. рис. 4.32, *б*) открывается транзистор *VT*1 (*n-p-n*-типа), пропуская в нагрузку импульс коллекторного тока $i_{к1}$. При этом на нагрузке выделяется положительная полуволна выходного напряжения $u_{вых}$ (рис. 4.32, *д*). На угловом интервале $\pi$ – $2\pi$, когда на вход каскада поступает отрицательная полуволна входного напряжения (см. рис. 4.32, *б*), открывается транзистор *VT*2 (*p-n-p*-типа) и через нагрузку протекает импульс тока $i_{к2}$, создавая на ней отрицательную полуволну выходного усиленного напряжения $u_{вых}$ (рис. 4.32, *г*, *д*).

Усилительный каскад мощности рассчитывают графоаналитическим методом, используя статические характеристики (рис. 4.33). Пусть работает одно плечо усилителя, когда на нагрузке выделяется полуволна выходного напряжения. Линию нагрузки по переменному току проводят из точки $E_к$ (рис. 4.33, *а*): угол ее наклона к оси абсцисс определяется сопротивлением $R_н$.

Среднее значение тока, потребляемого транзисторами от источника питания,

$$I_{к0} = \frac{1}{\pi}\int_0^{\pi} I_{кm}\sin\omega t d(\omega t) = \frac{2I_{кm}}{\pi}. \qquad (4.91)$$

Мощность, потребляемая от источника питания,

$$P_0 = E_к I_{к0} = 2E_к I_{кm}/\pi. \qquad (4.92)$$

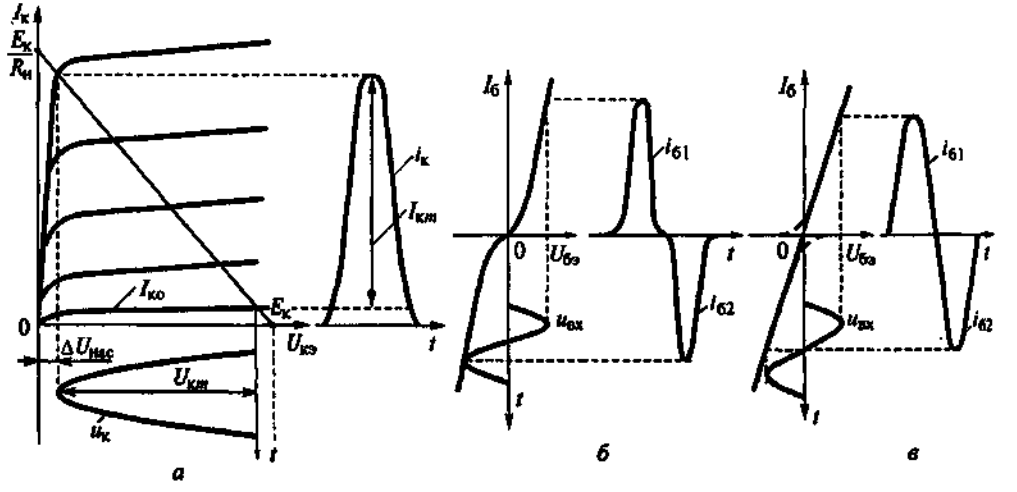



**Рис. 4.33.** Диаграммы работы транзисторов:
*а* — одного в режиме *В*; *б* — двух в режиме *В*; *в* — двух в режиме *АВ*

Мощность, отдаваемая транзисторами в нагрузку,

$$P_к = 0{,}5U_{км}I_{км} = 0{,}5I^2_{км}R_н. \quad (4.93)$$

С помощью (4.92) и (4.93) находим КПД коллекторной цепи каскада

$$\eta_к = \frac{P_к}{P_0} = 0{,}5\frac{U_{км}I_{км}}{2E_кI_{км}/\pi} = \frac{\pi U_{км}}{4E_к}. \quad (4.94)$$

Согласно (4.94) КПД коллекторной цепи растет с увеличением амплитуды выходного напряжения и при $U_{км} = E_к$ достигает предельного значения 78,5 %.

Так как оба транзистора в усилителе включены по схеме эмиттерного повторителя (схему часто называют *двухтактным эмиттерным повторителем*), то значительно упрощается согласование выходного сопротивления усилителя с низкоомной нагрузкой. Однако в этом случае выходное напряжение не превышает входное, и усиление мощности обеспечивается только за счет усиления тока.

Основной недостаток двухтактных усилителей мощности, работающих в режиме *B*, — искажения формы выходного сигнала из-за нелинейности начальных участков входных характеристик транзисторов. Объединенная входная характеристика при этом имеет излом вблизи нуля (рис. 4.33, *б*). Как видно из диаграмм, эта нелинейность искажает базовые токи $i_{б1}$ и $i_{б2}$, вследствие чего искажаются формы коллекторных токов и выходного напряжения. Устраняют недостаток введением транзисторов в промежуточный режим *AB* (рис. 4.33, *в*). Это достигают подачей на их базы небольших отрицательных напряжений смещения, равных напряжению отпирания. Обычно источником базового смещения служат диоды, стабилитроны или транзисторы в диодном включении.

**Пример 4.7.** Определить основные параметры бестрансформаторного выходного каскада на комплементарных транзисторах (см. рис. 4.32, *а*), если выходная мощность, выделяемая на нагрузке $R_н = 10$ Ом, равна 3 Вт.

Р е ш е н и е. Расчет поведем по формулам (4.92) – (4.94). Зададимся с небольшим запасом мощностью, которую должны выделять транзисторы на нагрузке

$$P_{к1} \geq 1{,}1P_к = 3{,}3 \text{ Вт}.$$

Максимальное значение тока коллектора

$$I_{км} = \sqrt{2P_{к1}/R_н} = 0{,}81 \text{ А}.$$

Амплитуда выходного напряжения на нагрузке

$$U_{км} = I_{км}R_н = 8{,}1 \text{ В}.$$

Для исключения нелинейных искажений выходного сигнала напряжения питания выбирают из условия $E_к \geq U_{км} + \Delta U_{нас}$ (см. рис. 4.33, *в*). Поскольку напряжение насыщения $\Delta U_{нас} = 0{,}3...0{,}5$ В, то зададим напряжение питания $E_к = \pm 9$ В.

Исходя из проделанных расчетов выбираем мощную комплементарную пару транзисторов КТ814А (*p-n-p*-типа) и КТ815А (*n-p-n*-типа), у которых среднее значение коэффициента усиления по току $h_{21} = 40$. Тогда входной ток двухтактного усилителя $I_{бм} = I_{км}/h_{21} = 0{,}81/40 = 0{,}02$ А = 20 мА.

## 4.5. Элементы теории обратной связи. Устойчивость систем с обратной связью

В 4.4 был описан один из вариантов осуществления обратной связи, где цепь ОС вводилась в усилительный каскад с ОЭ для термостабилизации режима работы транзистора. Рассмотрим этот вопрос подробнее.

### Элементы теории обратной связи

В реальном усилителе принципиально всегда существует *внутренняя паразитная обратная связь*, обусловленная физическими процессами, протекающими в усилительном элементе вследствие его неидеальности (наличие межэлектродных емкостей и т. д.). Внешние реактивные элементы — емкости и индуктивности — также способствуют возникновению паразитной ОС в усилителях. Внешняя и внутренняя паразитные ОС приводят к ухудшению параметров и нарушениям работы усилителя. Однако с помощью специально введенной ОС удается существенно улучшить характеристики усилителя. Простейший принцип введения «полезной» обратной связи был рассмотрен при анализе усилительного каскада ОЭ с термостабилизацией. В структурную схему устройства с ОС (рис. 4.34) входит собственно усилитель с частотным коэффициентом передачи $\mathbf{K}(\omega) = K(j\omega)$ и цепь (петля) ОС с коэффициентом передачи $\beta(\omega) = \beta(j\omega)$. В общем случае $\mathbf{K}(\omega)$ и $\beta(\omega)$, напряжения $\mathbf{U}_{вх}$ и $\mathbf{U}_{вых}$ и напряжение обратной связи $\mathbf{U}_{ос}$ — величины комплексные, что позволяет учесть возможный фазовый сдвиг, возникший в усилителе и цепи ОС из-за наличия реактивных элементов.

Усилители с ОС функционируют в *рабочей полосе частот*, где влияние паразитных ОС несущественно и ими можно пренебречь. В этом случае вместо частотного коэффициента передачи $\mathbf{K}(\omega)$ используют коэффициент усиления по напряжению. Этот параметр, как и коэффициент передачи петли ОС $\beta(\omega)$, характеризуется уже *вещественными значениями* $K$ и $\beta$ (в теории усилителей с ОС коэффициент усиления усилителя обозначают буквой $K$, а не $K_U$). Токи и напряжения в рабочей полосе частот также являются вещественными величинами.

В усилителях применяют или могут возникать самопроизвольно из-за наличия паразитных элементов различные виды ОС. На рис. 4.35 показаны структурные схемы усилителей с часто встречающимися (классическими) видами ОС. В зависимости от метода получения сигнала ОС различают *обратную связь по напряжению*, когда сигнал ОС пропорционален выходному напряжению, и *обратную связь по току*, если сигнал ОС пропорционален выходному току (сигнал ОС снимается с резистора $R_э$). Возможна также и *комбинированная обратная связь*, когда сигнал ОС пропорционален и напряжению и току выходной цепи усилительного каскада.

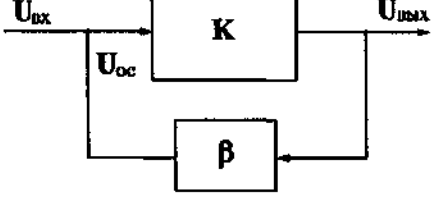



**Рис. 4.34.** Структурная схема усилителя с цепью ОС

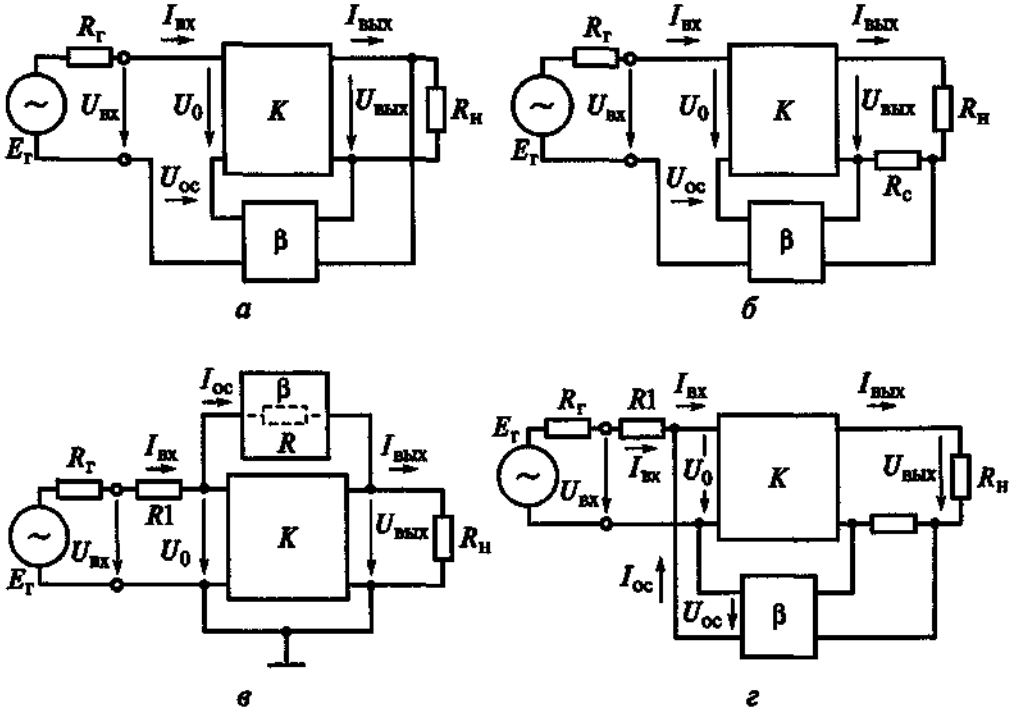



**Рис. 4.35.** Виды обратных связей в усилителях:
*а* — последовательная по напряжению; *б* — последовательная по току;
*в* — параллельная по напряжению; *г* — параллельная по току

По способу введения сигнала ОС во входную цепь усилителя различают *последовательную обратную связь* при включении цепи ОС последовательно с источником усиливаемого сигнала и *параллельную обратную связь*, если цепь ОС включена параллельно усилителю. Используется также и *комбинированная обратная связь* при введении как последовательной, так и параллельной ОС. В приведенных схемах усилителей с ОС указаны основные расчетные параметры: $U_0$ — *напряжение на входе собственно усилителя*; $U_{ос}$ — *напряжение обратной связи*; $K = U_{вых}/U_0$ — *коэффициент усиления собственно усилителя* (усилителя без ОС); $\beta = U_{ос}/U_{вых}$ — *коэффициент передачи петли* ОС.

Влияние обратной связи оценим на примере усилителя с последовательной ОС по напряжению (рис. 4.35, *а*). Введем общепринятый параметр — *коэффициент усиления усилителя с обратной связью* $K_{ос} = U_{вых}/U_{вх}$.

Пусть напряжение на входе собственно усилителя определяется суммой

$$U_0 = U_{вх} + U_{ос}. \tag{4.95}$$

Обе части формулы (4.95) поделим на $U_{вых}$

$$\frac{U_0}{U_{вых}} = \frac{U_{вх}}{U_{вых}} + \frac{U_{ос}}{U_{вых}} \tag{4.96}$$

и запишем это выражение через введенные коэффициенты

$$\frac{1}{K} = \frac{1}{K_{oc}} + \beta . \qquad (4.97)$$

С помощью (4.97) найдем коэффициент усиления усилителя с ОС

$$K_{oc} = \frac{K}{1 - K\beta} . \qquad (4.98)$$

Параметр $K\beta = U_{oc}/U_0$ определяет *фактор обратной связи*, или *коэффициент усиления разомкнутого кольца обратной связи*. Поясним, что для схемы на рис. 4.35, *а* кольцо ОС будет разомкнуто, если разорвана цепь напряжения обратной связи. Величина $(1 - K\beta)$ носит название *глубины обратной связи*.

Как следует из (4.98), при значениях $0 < K\beta < 1$ коэффициент усиления усилителя с ОС $K_{oc}$ становится больше коэффициента усиления собственно усилителя $K$. Это соответствует *положительной обратной связи* (ПОС), при которой напряжение обратной связи $U_{oc}$ поступает на вход усилителя в фазе с входным $U_{вх}$, вследствие чего $U_0 = U_{вх} + U_{oc}$. Значение $K\beta = 1$ характеризует условие *самовозбуждения* усилителя, когда он превращается в автогенератор колебаний широкого спектра частот, независимых от частоты входного сигнала.

Когда напряжение ОС находится в противофазе с входным, формула (4.95) запишется так: $U_0 = U_{вх} - U_{oc}$. В этом случае нетрудно показать, что

$$K_{oc} = \frac{K}{1 + K\beta} , \qquad (4.99)$$

т. е. коэффициент усиления усилителя с подобной ОС снижается в $(1 + K\beta)$ раз.

Такую ОС в теории усилителей и принято называть *отрицательной*.

Наглядно влияние обратной связи на коэффициент усиления усилителя показано на рис. 4.36, на котором можно выделить три характерные области.

1. Фактор обратной связи $K\beta << -1$; $|K\beta| >> 1$; $\beta < 0$. Тогда (4.99) примет вид

$$K_{oc} = \frac{K}{1 + K\beta} \approx -\frac{1}{\beta} = -\left|\frac{1}{\beta}\right| . \qquad (4.100)$$

Данное положение соответствует *глубокой* (100 %-ной) ООС, при которой коэффициент усиления $K_{oc}$ определяется только величиной $\beta$ и не зависит от коэффициента усиления собственно усилителя $K$. Такое свойство ООС широко используют в усилительной технике. Действительно, при стабильном значении $\beta$ изменение $K$ в несколько раз практически не изменяет коэффициент усиления усилителя с ООС $K_{oc}$. Это позволяет выпускать узлы радиоаппаратуры с одинаковыми параметрами даже при значительных разбросах коэффициента усиления собственно усилителя.

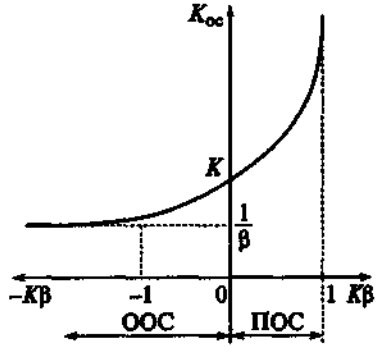



**Рис. 4.36.** Зависимость коэффициента усиления $K_{oc}$ от фактора обратной связи $K\beta$

2. $K\beta = 0$; $\beta = 0$; $K_{oc} = K$. Этот случай характеризует отсутствие ОС.

3. $K\beta \to 1$; $\beta > 0$; из (4.98) имеем $K_{oc} = 1/0 \to \infty$. Физически бесконечно большой $K_{oc}$ означает, что усилитель превращается в автогенератор колебаний.

Оценим влияние отрицательной обратной связи на *стабильность* (неизменность) коэффициента усиления усилителя. Для этого продифференцируем уравнение (4.99) по $K$:

$$\frac{dK_{oc}}{dK} = \frac{1}{(1+K\beta)^2}. \tag{4.101}$$

Перегруппировав в (4.101) переменные и поделив его на (4.99), получим

$$\frac{dK_{oc}}{K_{oc}} = \frac{dK}{K}\frac{1}{(1+K\beta)}. \tag{4.102}$$

Выражение (4.102) показывает, что относительное изменение коэффициента усиления усилителя с ООС в $(1 + K\beta)$ раз меньше относительного изменения коэффициента усиления собственно усилителя. Если, например, относительное изменение коэффициента усиления усилителя $dK/K = 30\%$ и $1 + K\beta = 300$, то относительное изменение коэффициента усиления усилителя с ООС составит $dK_{oc}/K_{oc} = 0{,}1\%$. Итак, усилитель с ООС характеризуется существенно более высокой стабильностью. Однако повышение стабильности параметров усилителя с ООС в $(1 + K\beta)$ раз во столько же раз уменьшает коэффициент усиления.

Специально введенная ООС существенно влияет и на входное сопротивление усилителя $R_{вх} = U_0/I_{вх}$. Проанализируем параметры усилителя с последовательной ООС по напряжению (рис. 4.37, *а*), при которой $U_{вх} = U_0 + U_{oc}$. Разделив члены этого уравнения на $I_{вх}$, найдем входное сопротивление усилителя с ООС:

$$R_{вх\,oc} = \frac{U_{вх}}{I_{вх}} = \frac{U_0}{I_{вх}} + \frac{U_{oc}}{I_{вх}} = R_{вх}\left(1 + \frac{U_{oc}}{U_0}\right) = R_{вх}(1+K\beta). \tag{4.103}$$

Аналогично можно показать, что выходное сопротивление

$$R_{вых\,oc} = \frac{\Delta U_{вых\,oc}}{\Delta I_{вых\,oc}} = \frac{R_{вых}}{1+K\beta}. \tag{4.104}$$

Из формул (4.103) и (4.104) следует, что в усилителе с последовательной отрицательной ОС по напряжению входное сопротивление каскада увеличивается в $(1 + K\beta)$ раз и во столько же раз уменьшается его выходное сопротивление.

**Пример 4.8.** Пусть в схеме усилительного каскада сопротивления $R_{вх} = 5$ кОм, $R_{вых} = 0{,}2$ кОм и коэффициент усиления по напряжению $K = 50$. Как изменятся эти параметры при введении в данный усилительный каскад цепи последовательной отрицательной обратной связи по напряжению (см. рис. 4.35, *а*) с коэффициентом передачи цепи ОС $\beta = 0{,}02$?

**Решение.** Подставив в (4.99) заданные параметры, определим коэффициент усиления усилителя с ООС

$$K_{oc} = \frac{K}{1 + K\beta} = \frac{50}{1 + 0{,}02 \cdot 50} = 25.$$

Воспользовавшись соотношениями (4.103) и (4.104), находим входное и выходное сопротивления усилителя с ООС:

$$R_{вх\,oc} = R_{вх}(1 + K\beta) = 5 \cdot 2 = 10 \text{ кОм},$$

$$R_{вых\,oc} = R_{вых}/(1 + K\beta) = 0{,}2/2 = 0{,}1 \text{ кОм}.$$

Введение ООС позволяет существенно уменьшить нелинейные искажения сигнала и шумы, возникающие в процессе усиления в самом усилителе и, особенно, в его выходном каскаде. Сравним, например, уровни нелинейных искажений в усилителе с ООС и без нее. Пусть при усилении гармонического сигнала на выходе усилителя без ОС в результате нелинейных искажений появилось напряжение паразитной гармоники $U_г$. Обозначим напряжение этой же гармоники уже на выходе усилителя с ООС через $U_{г\,oc}$. Ее значение будет равно разности напряжения $U_г$ и напряжения $K\beta U_{г\,oc}$, появившегося на выходе усилителя за счет введения ООС, т.е. $U_{г\,oc} = U_г - K\beta U_{г\,oc}$. Отсюда следует, что

$$U_{г\,oc} = \frac{U}{1 + K\beta}. \qquad (4.105)$$

Итак, ООС в $(1 + K\beta)$ раз снижает уровень паразитной гармоники, возникающей в усилителе из-за нелинейных искажений. Аналогичное влияние оказывает ООС и на напряжение шумов, создаваемых самим усилителем.

### Устойчивость линейных систем (усилителей) с обратной связью

Общепринятое определение понятия «обратная связь» позволяет более строго обосновать существующее разделение всех устройств с обратными связями на два класса — устройства с положительной ОС, используемые для генерации электрических колебаний, в качества регенеративных усилителей, корректирующих цепей и т. п., и устройства с отрицательной ОС, применяемые как в радиотехнике для стабилизации параметров схем и получения широкой полосы пропускания, так и в устройствах автоматического регулирования и управления.

Практические схемы усилителей с ОС всегда содержат реактивные элементы (элементы с реактивным сопротивлением), накапливающие энергию. Это могут быть межэлектродные емкости транзисторов и микросхем, паразитные емкости монтажа, индуктивности проводов и т. д. Реактивные элементы создают дополнительные фазовые сдвиги усиливаемого сигнала. Если на какой-либо частоте сумма фазовых сдвигов достигает 180°, то ОС из отрицательной переходит в положительную, превращая усилитель в автогенератор электрических колебаний. В этом случае говорят о *неустойчивости усилителя.* Итак, применение обратной связи тесно связано с проблемой обеспечения *устойчивости* усилителей и других устройств и систем, имеющих в своем составе усилительные элементы.

Положим, что некоторый параметр электронной системы $\lambda$ должен изменяться во времени линейно (сплошная линия 1 на рис. 4.37). Очевидно, что этот

режим соответствует устойчивому состоянию системы. Пусть в момент времени $t = t_0$ на систему подействовала возмущающая сила, под действием которой рассматриваемый параметр $\lambda$ изменил свое значение от $\lambda_0$ до $\lambda_1$. После исчезновения возмущения изменение параметра $\lambda$ может происходить так, как показано кривыми 2 или 3 на рис. 4.37. Если *система устойчива*, то возмущенная траектория изменения параметра $\lambda$ (кривая 2) с течением времени будет сколь угодно мало отличаться от невозмущенной линии 1. Кривая 3, у которой амплитуда колебаний нарастает, соответствует *неустойчивой системе*.

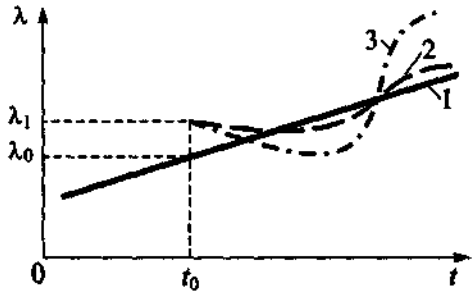



**Рис. 4.37.** К понятию устойчивости электронной системы

**Собственные (свободные) колебания в линейных динамических системах.** Для полной оценки поведения динамической системы, описываемой неоднородным (не равным нулю) дифференциальным уравнением

$$a_n \frac{d^n u_{\text{вых}}}{dt^n} + a_{n-1} \frac{d^{n-1} u_{\text{вых}}}{dt^{n-1}} + \ldots + a_1 \frac{du_{\text{вых}}}{dt} + a_0 u_{\text{вых}} = i(t), \qquad (4.106)$$

необходимо учесть начальные условия, характеризующие внутреннее состояние линейной системы в некоторый заданный момент времени. Оказывается, что наиболее удобно задавать искомую линейную функцию и ее $n-1$ производные для $t = 0$: $u_{\text{вх}}(0)$; $u'_{\text{вых}}(0), \ldots, u^{n-1}_{\text{вых}}(0)$.

Из курса математики известно, что решение неоднородного дифференциального уравнения (4.106), удовлетворяющего любым начальным условиям, в общем случае состоит из суммы двух слагаемых:

- частного решения неоднородного уравнения, правая часть которого $i(t)$ отлична от нуля и которое характеризует вынужденные колебания системы под действием возмущающих сил;
- общего решения однородного уравнения

$$a_n \frac{d^n u_{\text{вых}}}{dt^n} + a_{n-1} \frac{d^{n-1} u_{\text{вых}}}{dt^{n-1}} + \ldots + a_1 \frac{du_{\text{вых}}}{dt} + a_0 u_{\text{вых}} = 0, \qquad (4.107)$$

определяющего свободные (собственные) колебания системы после исчезновения возмущающих сил.

При исследовании устойчивости систем нас интересуют только свободные колебания, так как вся теория устойчивости основана на использовании понятия кратковременных возмущающих сил типа дельта-функции $\delta(t)$.

Решение однородного дифференциального уравнения связано с отысканием корней его характеристического уравнения

$$a_n q^n + a_{n-1} q^{n-1} + \cdots + a_1 q + a_0 = 0. \qquad (4.108)$$

Данное уравнение имеет $n$ корней. Поскольку постоянные коэффициенты $a_n$ являются вещественными величинами, корни уравнения $q_1, q_2, ..., q_n$ могут быть либо вещественными, либо комплексно-сопряженными. Когда все корни различны, общее решение однородного дифференциального уравнения (4.107) характеризует собственные колебания в линейной динамической системе и имеет следующий вид

$$u_{вых}(t) = V_1 e^{q_1 t} + V_2 e^{q_2 t} + \cdots + V_n e^{q_n t}, \qquad (4.109)$$

где $V_1, V_2, ..., V_n$ — постоянные числа, определяемые из начальных условий.

С точки зрения воздействия сигналов на электрическую динамическую систему (электрическую цепь), система *устойчива*, если все ее собственные (свободные) колебания затухают во времени. В этом случае необходимым и достаточным условием устойчивости системы является отрицательное значение действительных (вещественных) частей всех корней характеристического уравнения (4.108). Данные корни не должны быть также и чисто мнимыми. Хотя при этом свободные колебания в линейной динамической системе являются гармоническими функциями вида

$$u_{св}(t) = \cos\omega t, \quad u_{св}(t) = \sin\omega t,$$

небольшие случайные изменения параметров системы могут перевести ее в неустойчивое состояние, когда напряжения

$$u_{св}(t) = e^{\alpha t}\cos\omega t, \quad u_{св}(t) = e^{\alpha t}\sin\omega t$$

являются экспоненциально нарастающими по амплитуде колебаниями.

В принципе возникновение нарастающих собственных колебаний в электрических цепях возможно в случае, если в состав цепи наряду с пассивными элементами $L$, $C$ и $R$ включены активные элементы, вносящие в нее энергию от внешних источников. Наиболее распространенной моделью такого активного элемента является резистор с отрицательным сопротивлением (например, туннельный диод).

Рассмотрим электронную систему (рис. 4.38), аналогичную системе на рис. 4.34, но с комплексной переменной $p$ и включающую в себя линейный усилитель с передаточной функцией $K(p)$, у которого выход соединен с входом через цепь обратной связи с передаточной функцией $\beta(p)$.

Положим, что внешний возмущающий сигнал $U_{вых}(p) = 0$, т. е. электронная система автономна (независима). Аналитически уравнение состояния данной системы записывают на основании того, что изображение выходного сигнала связано с изображениями параметров активного элемента (линейного усилителя) и цепи обратной связи как

$$U_{вых}(p) = K(p)U_{вх}(p) = K(p)\beta(p)U_{вых}(p).$$

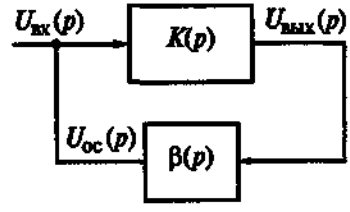



**Рис. 4.38.** Электронная система с обратной связью

Отсюда получим

$$[1 - K(p)\beta(p)]U_{вых}(p) = 0. \quad (4.110)$$

Обычно $U_{вых}(p)$ тождественно не равно нулю (иначе система не могла бы возбуждаться), поэтому равенство (4.110) справедливо лишь при тех значениях аргумента $p$, которые являются корнями характеристического уравнения

$$1 - K(p)\beta(p) = 0. \quad (4.111)$$

Пусть $p_1, p_2, ..., p_n$ — корни этого уравнения. Поскольку исследуется линейная система, то в общем случае по аналогии с (4.109) выходной сигнал

$$u_{вых}(t) = U_1 e^{p_1 t} + U_2 e^{p_2 t} + \cdots + U_n e^{p_n t}. \quad (4.112)$$

Этот сигнал не будет нарастать, если все корни характеристического уравнения имеют отрицательные действительные части, т. е. располагаются в левой полуплоскости переменной $p = \alpha + j\omega$. Линейная система с ОС, обладающая такими свойствами, будет устойчива.

При синтезе и анализе линейных систем с ОС имеют место две задачи. Если синтезируемая система должна быть устойчивой, необходим критерий, который позволил бы по виду функций $K(p)$ и $\beta(p)$ судить об отсутствии корней характеристического уравнения в правой полуплоскости. В случае, когда ОС используется для создания неустойчивой автоколебательной системы, то необходимо знать корни уравнения (4.111), определяющие частоту возникающих колебаний.

При высоком порядке динамической системы прямая проверка устойчивости, связанная с отысканием корней характеристического уравнения, может быть весьма затруднительной. Поэтому был разработан ряд специальных критериев устойчивости, позволяющих определять наличие корней с отрицательными (или положительными) значениями действительных частей непосредственно по виду коэффициентов, не решая самого характеристического уравнения.

**Алгебраические критерии устойчивости**. Задачу решают, анализируя соотношения между коэффициентами уравнения без определения ее корней. Критерий доказан в 1895 г. немецким математиком Адольфом Гурвицем (*Adolf Hurwitz*; 1859–1919). Критерий относится к алгебраическим и по нему исследуют коэффициенты характеристического уравнения замкнутой системы.

Чтобы записать данное уравнение, вначале необходимо найти выражение для передаточной функции замкнутой системы. Пусть и основной элемент и элемент ОС являются цепями с сосредоточенными параметрами и описываются следующими передаточными функциями

$$K(p) = \frac{R_1(p)}{Q_1(p)}, \quad \beta(p) = \frac{R_2(p)}{Q_2(p)}. \quad (4.113)$$

Подставив (4.113) в (4.111), получим характеристическое уравнение системы

$$\frac{Q_1(p)Q_2(p) - R_1(p)R_2(p)}{Q_1(p)Q_2(p)} = 0. \quad (4.114)$$

Отсюда следует, что система с ОС устойчива, если все корни уравнения

$$H(p) = Q_1(p)Q_2(p) - R_1(p)R_2(p) = 0$$

имеют отрицательные действительные части.

В математике многочлены $H(p)$ с называют *многочленами Гурвица.*

Проанализируем один из множества реальных многочленов Гурвица

$$H(p) = (p - p_1)(p - p_2)(p - p_3)$$

с тремя корнями, один из которых $p_1 = -\alpha$ действительный и отрицательный, а два других — комплексно-сопряженные с отрицательными вещественными частями: $p_{2,3} = -\gamma \pm j\,\omega_0$. Прямая подстановка корней показывает, что многочлен

$$H(p) = (p + \alpha)[(p + \gamma)^2 + \omega_0^2] = p^3 + (\alpha + 2\gamma)p^2 + (\gamma^2 + 2\alpha\gamma + \omega_0^2)p + \alpha(\gamma^2 + \omega_0^2) \qquad (4.115)$$

содержит все степени переменной $p$, начиная со старшей, и имеет коэффициенты одного знака. Однако рассмотренный признак характеризует лишь необходимые условия того, чтобы многочлен относился к многочленам Гурвица.

Полное решение этой задачи нашло отражение в известной *теореме Гурвица*, который звучит так — для того чтобы система была устойчивой, необходимо и достаточно иметь положительными следующие величины:

- коэффициенты $a_0$ и $a_n$;
- определители многочлена Гурвица от 1-го до $n$-го порядков, составленные из коэффициентов $a_0, a_1, ..., a_{n-1}, a_n$ характеристического уравнения (4.108);
- все главные миноры этих определителей.

Главный определитель многочлена Гурвица строится по следующей схеме

$$\Delta_n = \begin{vmatrix} a_{n-1} & a_n & 0 & 0 & \dots & 0 & 0 \\ a_{n-3} & a_{n-2} & a_{n-1} & a_n & \dots & 0 & 0 \\ a_{n-5} & a_{n-4} & a_{n-3} & a_{n-2} & \dots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & \dots & a_0 & a_1 \end{vmatrix}. \qquad (4.116)$$

Алгебраический критерий устойчивости принято называть *критерием Рауса–Гурвица* (Эдвард Раус; *Edward Routh*, 1831–1907, английский математик). Раскрывая определители многочлена Гурвица, можно получить условия устойчивости для систем различных порядков. Системы 1-го и 2-го порядков устойчивы всегда, так как условия устойчивости для них сводятся к требованию $a_0 > 0$, $a_1 > 0$, $a_2 > 0$. Для системы 3-го порядка условие устойчивости находят, раскрывая определитель $\Delta_2$:

$$a_1 a_2 - a_0 a_3 > 0.$$



А. Гурвиц



Э. Раус

Если последние неравенства приравнять к нулю, получим условия, при которых система будет находиться на *границе устойчивости*.

**Пример 4.9.** С помощью критерия Рауса–Гурвица проверить устойчивость системы, характеристическое уравнение которой имеет вид

$$p^3 + 4p^2 + 3p + 4 = 0.$$

Р е ш е н и е. В этой формуле $a_3 > 0$ и $a_0 > 0$. Строим определитель

$$\Delta_2 = \begin{vmatrix} 4 & 1 \\ 4 & 3 \end{vmatrix} = 8 > 0.$$

Единственный главный минор $4 > 0$. Таким образом, система устойчива.

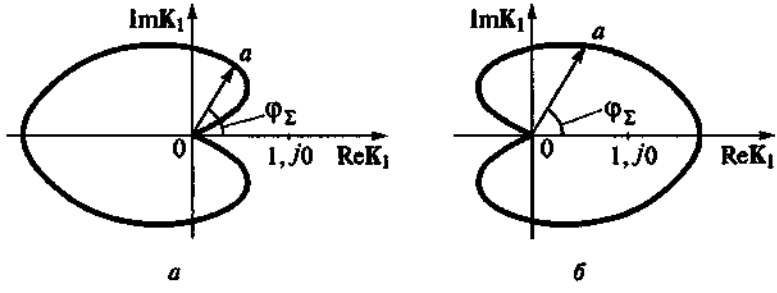
Основное достоинство критерия Рауса–Гурвица — простота вычислений. Недостаток — он применим для цепей с сосредоточенными параметрами, поскольку только в этом случае передаточная функция является частным от деления двух многочленов в (4.114).

**Критерий устойчивости Найквиста.** В радиотехнике широко применяют *критерий Найквиста*, или *частотный критерий* (разработан в 1932 г. для исследования устойчивости усилителей с ОС). При исследовании устойчивости анализируют частотный коэффициент передачи усилителя с разомкнутым кольцом ОС, называемый *амплитудно-фазовой характеристикой* (АФХ):

$$\mathbf{K}_1(\omega) = \mathbf{K}(\omega)\beta(\omega) = K(\omega)e^{j\varphi_K(\omega)}\beta(\omega)e^{j\varphi_\beta(\omega)} = K_1(\omega)e^{j\varphi_\Sigma(\omega)}, \qquad (4.117)$$

где $K_1(\omega) = K(\omega)\beta(\omega)$ — модуль коэффициента передачи такого усилителя; $\varphi_\Sigma(\omega) = \varphi_K(\omega) + \varphi_\beta(\omega)$ — суммарный фазовый сдвиг выходного сигнала, вносимый соответственно собственно усилителем и цепью ОС.

АФХ $\mathbf{K}_1(\omega)$ представляют графически на комплексной плоскости (в координатах $\mathrm{Im}\mathbf{K}_1$, $\mathrm{Re}\mathbf{K}_1$) *годографом* — линией, отражающей зависимость коэффициента передачи от частоты (рис. 4.39). Годограф может быть построен на основании аналитических расчетов, а также снят экспериментально. Это обстоятельство выгодно отличает рассматриваемый критерий от всех других. Для построения годографа амплитудно-фазовой характеристики в диапазоне усили-




*а* *б*

**Рис. 4.39.** Годографы усилителей:
*а* — устойчивого; *б* — неустойчивого

ваемых частот от – ∞ до ∞ вычисляют ее модуль $K_1(\omega)$ и аргумент $\varphi_\Sigma(\omega)$. Определив, например, для частоты $\omega_a$ модуль $K_1(\omega)$ и аргумент $\varphi_\Sigma(\omega)$, проводят из начала координат вектор длиной $K_1(\omega_a)$ под углом $\varphi_\Sigma(\omega_a)$ к действительной оси (см. линии 0 – *a* на рис. 4.39). Проделав аналогичные операции для других частот, лежащих в основном вне полосы пропускания усилителя, строят годограф, соединяя линией концы векторов. Устойчивость усилителя определяют по расположению точки с координатами $(1, j0)$ относительно годографа (рис. 4.39). *Если годограф АФХ усилителя с разомкнутым кольцом ОС охватывает точку с координатами* $(1, j0)$, *то усилитель с замкнутым кольцом ОС неустойчив.* Это условие и является критерием устойчивости Найквиста, или геометрическим критерием устойчивости усилителя.

Во всех случаях для усилителей с ОС, представляющих практический интерес, модуль коэффициента передачи усилителя с разомкнутым кольцом обратной связи $K_1(\omega)$ стремится к нулю с ростом частоты. Поэтому АФХ проходит через начало координат. Кроме того, АФХ замкнута и симметрична относительно вещественной оси, поскольку $\mathbf{K}_1(-\omega) = \mathbf{K}_1^*(\omega)$.

Известны и другие методы исследования устойчивости линейных систем с обратной связью, например критерий Михайлова, корневой критерий.

## 4.6. Интегральные и специальные усилители

Во многих радиоэлектронных устройствах имеют место колебания, частоты которых близки к нулю. Для усиления медленно меняющихся во времени сигналов применяют *усилители постоянного тока* (УПТ). Современные УПТ в основном выполняют в виде интегральных микросхем и их можно условно разделить на дифференциальные и операционные усилители.

### Дифференциальные и операционные усилители

**Дифференциальные усилители.** Практически во всех схемах УПТ между соседними усилительными каскадами используют *непосредственную* (*гальваническую*) связь, что вызывает эффект *дрейфа нуля* — самопроизвольное изменение выходного сигнала при отсутствии входного сигнала. Даже небольшое изменение постоянного тока транзистора в первом каскаде УПТ создает приращение напряжения на его выходе, которое затем усиливается другими каскадами и выделяется на нагрузке как полезный сигнал.

Дрейф нуля отсутствует в схемах *параллельно-балансных* УПТ, называемых *дифференциальными усилителями* (ДУ), которые строят по методу уравновешенного моста. На входы ДУ подают как *дифференциальные* (*разностные*), так и *синфазные* (*одинаковые*) сигналы. При этом ДУ усиливает *дифференциальные* и *подавляет* синфазные сигналы. Один из входов ДУ по отношению к выходу является *инвертирующим* (выходной сигнал противофазен входному), а другой — *неинвертирующим* (выходной сигнал синфазен с входным). Наиболее часто ДУ применяют для сравнения с высокой точностью значений или разности двух напряжений. Это объясняет название «дифференциальный усилитель».

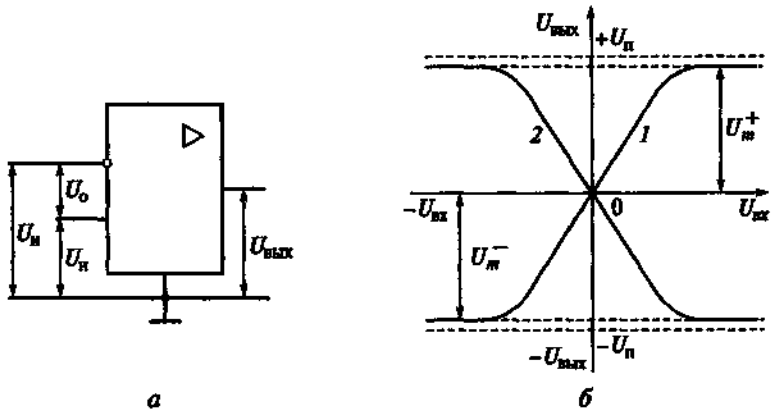



**Рис. 4.40.** Операционный усилитель:
*а* — условное графическое обозначение; *б* — передаточные характеристики

**Операционные усилители.** Первый операционный усилитель на лампах разработал в 1938 г. Д. Филбрик (*G. Philbrick*; США). *Операционным усилителем* (ОУ) называют высококачественный линейный усилитель напряжения, имеющий большой коэффициент усиления ($10^6...10^7$), высокое входное (сотни мегаом) и малое выходное (единицы ом) сопротивления. Входным каскадом ОУ является дифференциальный усилитель, а выходным каскадом — эмиттерный повторитель. На рис. 4.40, *а* показано условное графическое обозначение ОУ.

Один из входов ОУ по отношению к выходу является *неинвертирующим* $U_н$, а другой — *инвертирующим* $U_и$; последний обозначается знаком *инверсии* (кружок на вводе ОУ). Питание ОУ осуществляют от двух одинаковых разнополярных источников $+U_п$ и $-U_п$ (на графических обозначениях источники питания обычно не показывают). При таком питании входные и выходные сигналы могут быть двуполярными, а нулевым входным сигналам соответствует нулевой выходной сигнал. Выходной сигнал ОУ пропорционален дифференциальному входному сигналу — разности входных

$$U_0 = U_н - U_и.$$

Коэффициент усиления по напряжению $K_0$ собственно ОУ равен отношению выходного напряжения к дифференциальному входному напряжению:

$$K_0 = U_{вых}/U_0. \qquad (4.118)$$

Передаточные характеристики (рис. 4.40, *б*) имеют важнейшее значение для ОУ. Если усиливаемый сигнал подан на неинвертирующий вход, а инвертирующий заземлен, то знак выходного напряжения совпадает со знаком входного (линия *1*). При подаче сигнала на инвертирующий вход и заземлении неинвертирующего знак выходного напряжения будет противоположен знаку входного (линия *2*). Угол наклона линейных участков передаточных характеристик пропорционален коэффициенту усиления по напряжению $K_0$. Горизонтальные уча-

стки передаточных характеристик соответствуют режиму насыщения оконечных транзисторов ОУ, поэтому выходное напряжение

$$\pm U_{\text{вых}} = U_m^{\pm} \approx \pm U_{\text{п}}. \tag{4.119}$$

В теории интегральной усилительной техники с целью упрощения анализа и расчета схем на ОУ вводят понятие *«идеальный ОУ»*, для которого справедливы следующие допущения: бесконечно большие коэффициент усиления $K_0 = \infty$ и входное сопротивление $R_{\text{вх0}} = \infty$ и нулевое выходное сопротивление $R_{\text{вых0}} = 0$.

Из этих допущений вытекают два основных свойства (правила анализа) ОУ.

1. Дифференциальный входной сигнал равен нулю

$$U_0 = \frac{U_{\text{вых}}}{K_0} = \frac{U_{\text{вых}}}{\infty} = 0. \tag{4.120}$$

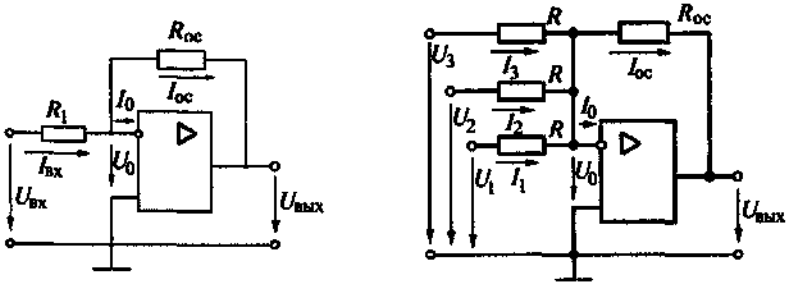
2. Входы ОУ не потребляют ток от источника входного сигнала

$$I_0 = \frac{U_{\text{вх}}}{R_{\text{вх0}}} = \frac{U_{\text{вх}}}{\infty} = 0. \tag{4.121}$$

Понятие идеального ОУ соответствует принципу *«виртуального»* (кажущегося) замыкания его инвертирующего и неинвертирующего входов. При виртуальном замыкании, как и при физическом, напряжение между соединенными зажимами равно нулю. Но в отличие от физического замыкания, ток между виртуально замкнутыми зажимами не течет. Иначе говоря, для тока виртуальное замыкание зажимов эквивалентно разрыву электрической цепи.

В зависимости от условий подачи усиливаемого сигнала на входы ОУ и подключения к нему внешних элементов можно получить две фундаментальные схемы включения — *инвертирующую* и *неинвертирующую*. Фактически любое схемотехническое решение с применением ОУ базируется на этих включениях.

Инвертирующий усилитель. В таком усилителе входное напряжение через резистор $R_1$ подается на инвертирующий вход, который с помощью резистора *обратной связи* $R_{\text{ос}}$ охвачен параллельной ООС по напряжению (рис. 4.41, *а*). Неинвертирующий вход каскада заземлен.




*а* *б*

**Рис. 4.41.** Инвертирующие схемы на ОУ:
*а* — усилителя; *б* — сумматора

Для определения параметров инвертирующего усилителя воспользуемся первым законом Кирхгофа для токов инвертирующего входа: $I_{вх} = I_0 + I_{ос}$. Поскольку по второму свойству идеального ОУ ток $I_0 = 0$, то $I_{вх} = I_{ос}$. Выразив токи через соответствующие им входные напряжения, получим

$$\frac{U_{вх} - U_0}{R_1} = \frac{U_0 - U_{вых}}{R_{ос}}.$$

По первому свойству идеального ОУ напряжение $U_0 = 0$, поэтому $U_{вх}/R_1 = -U_{вых}/R_{ос}$. Тогда коэффициент усиления инвертирующего усилителя

$$K_и = \frac{U_{вых}}{U_{вх}} = -\frac{R_{ос}}{R_1}. \qquad (4.122)$$

Согласно формуле (4.122), изменением значения сопротивления обратной связи $R_{ос}$ можно регулировать коэффициент усиления.

Входное сопротивление инвертирующего усилителя существенно меньше собственного входного сопротивления ОУ. В схеме на рис. 4.41, *а* точка подключения инвертирующего входа ОУ является *виртуальным нулем*, т. е. по входному сигналу она заземлена. Можно показать, что входное и выходное сопротивления инвертирующего усилителя:

$$R_{вх\,и} = \frac{U_{вх} - U_0}{I_{вх}} = \frac{U_{вх}}{I_{вх}} = R_1; \quad R_{вых\,и} = \frac{K_и}{K_0} R_{вых\,0}. \qquad (4.123)$$

Отметим, что при $R_1 = R_{ос}$, $K_и = -1$ схема на рис. 4.41, *а* превращается в *инвертирующий повторитель* (*инвертор*).

Еще одним вариантом инвертирующего усилителя является *преобразователь тока в напряжение*. Это достигают при $R_1 = 0$. Тогда ток $I_{вх} = I_{ос} = -U_{вых}/R_{ос}$ и выходное напряжение $U_{вых} = -I_{вх}R_{ос}$.

Инвертирующий сумматор (суммирующий усилитель). Для суммирования нескольких напряжений можно использовать инвертирующее включение ОУ. На рис. 4.41, *б* в качестве примера показан трехвходовый инвертирующий сумматор. Входные напряжения $U_1$, $U_2$ и $U_3$ через резисторы с обычно одинаковыми сопротивлениями $R$ подаются на инвертирующий вход ОУ. Учитывая, что напряжение на инвертирующем входе ОУ характеризуется виртуальным нулем, токи $I_1$, $I_2$ и $I_3$ будут заданы только соответствующими им входными напряжениями $U_1$, $U_2$ и $U_3$ и сопротивлением $R$:

$$I_1 = U_1/R; \qquad I_2 = U_2/R; \qquad I_3 = U_3/R.$$

Поскольку, согласно второму свойству идеального ОУ, инвертирующий вход почти не потребляет ток, то сумма всех входных токов $I_1$, $I_2$ и $I_3$ протекает только через резистор $R_{ос}$ и общий входной ток создает на нем падение напряжения $U_{вых} = -(I_1 + I_2 + I_3)R_{ос}$. Подставив в эту формулу соответствующие значения токов, выраженные через входные напряжения, и положив $R = R_{ос}$, получим

$$U_{вых} = -(U_1 + U_2 + U_3). \qquad (4.124)$$

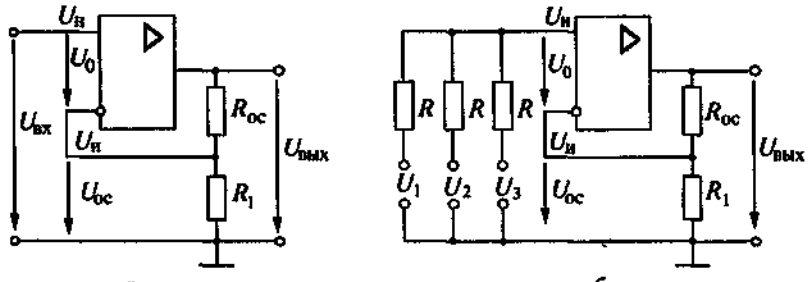



**Рис. 4.42.** Неинвертирующие схемы на ОУ:
*а* — усилителя; *б* — сумматора

Итак, выходное напряжение равно алгебраической сумме входных напряжений, взятых с обратным знаком. В схеме инвертирующего сумматора входные токи полностью протекают через резистор обратной связи $R_{ос}$. При этом токи практически не влияют друг на друга, следовательно, и входные напряжения не взаимодействуют друг с другом, т.е. все три входа усилителя полностью развязаны. Последнее свойство инвертирующего сумматора исключительно полезно для смешивания (микширования) сигналов низкой (звуковой) частоты.

Неинвертирующий усилитель. В таком усилителе входной сигнал поступает на неинвертирующий вход, а инвертирующий — с помощью резистивного делителя $R_1$,$R_{ос}$ охвачен последовательной отрицательной ОС по напряжению (рис. 4.42, *а*).

В схеме $U_{вх} = U_0 + U_{ос}$. Поскольку $U_0 = 0$, то $U_{вх} = U_{ос} = U_{вых}R_1/(R_1 + R_{ос})$. Отсюда коэффициент усиления неинвертирующего усилителя

$$K_н = \frac{U_{вых}}{U_{вх}} = 1 + \frac{R_{ос}}{R_1}. \qquad (4.125)$$

С помощью несложных выкладок можно показать, что входное сопротивление неинвертирующего усилителя велико и равно входному сопротивлению ОУ по неинвертирующему входу, а выходное сопротивление близко к нулю.

Если сопротивление обратной связи $R_{ос} = 0$, то $U_{вых} = U_{вх}$, и неинвертирующий усилитель превращается в *повторитель напряжения*, который используют в радиоэлектронных устройствах для гальванической развязки различных схем.

Неинвертирующий сумматор. Такое включение ОУ можно использовать для суммирования $n$ входных сигналов. Простейшая схема трехвходового неинвертирующего сумматора представлена на рис. 4.42, *б*. Как правило, все входные напряжения источников подключены к неинвертирующему входу ОУ через резисторы $R$ с одинаковым сопротивлением. С помощью несложных вычислений, которые приведены в специальной литературе по ОУ, можно показать, что выходное напряжение $n$-входового сумматора при выборе сопротивлений $R_{ос} = R(n - 1)$ определяется по формуле

$$U_{вых} = (U_1 + U_2 + ... + U_n). \qquad (4.126)$$

Итак, выходное напряжение неинвертирующего сумматора равно алгебраической сумме входных напряжений. Отметим, что для получения минимальных погрешностей при суммировании напряжений необходимо выбирать источники входных сигналов с достаточно малыми выходными сопротивлениями.

**Пример 4.10.** На основе ОУ типа К140УД24 рассчитать усилитель, обеспечивающий коэффициент усиления $K_и = 10$ при работе на нагрузку $R_н = 5$ кОм. Усилитель должен иметь входное сопротивление не менее 15 кОм при амплитуде входного сигнала $U_{вх} = 0{,}1$ В.

Решение. Из справочников находим, что ОУ типа К140УД24 имеет параметры: $K_0 = 10^6$; $I_п = 3{,}5$ мА; $R_{вых\,0} = 50$ Ом. Поскольку требуемое входное сопротивление велико, то используем схему инвертирующего усилителя (см. рис. 4.41, *а*). Выберем сопротивление $R_1 = R_{вх} = 15$ кОм. Используя формулу (4.122), находим сопротивление обратной связи

$$R_{ос} = |K_и|R_1 = 10 \cdot 15 = 150 \text{ кОм}.$$

Ток выходной цепи ОУ определим как сумму токов, протекающих через резисторы $R_{ос}$ и $R_н$:

$$I_{вых} = I_{ос} + I_н = \frac{U_{вых} - U_0}{R_{ос}} + \frac{U_{вых}}{R_н} = \frac{K_и U_{вх}}{R_{ос}} + \frac{K_н U_{вх}}{R_н} = 0{,}21 \text{мА}.$$

Такой ток допустим для заданного типа ОУ.

Выходное сопротивление усилительной схемы, согласно формуле (4.123), составляет $R_{вых\,и} = R_{вых\,0}K_и/K_0 = 0{,}25$ Ом, что много меньше требуемого значения.

Дифференцирующее устройство. Токи (здесь и далее аргумент $t$ в функциях значений токов и напряжений для упрощения опущен) для дифференцирующего устройства (рис. 4.43, *а*) $i_C = i_0 + i_R$, и поскольку ток $i_0 = 0$, то $i_C = i_R$. Записав мгновенные значения токов, протекающих через конденсатор $C$ и резистор $R$ как $i_C = C du_{вх}/dt$ и $i_R = -u_{вых}/R$, получим для выходного напряжения

$$u_{вых} = -RC\frac{du_{вх}}{dt} = -\tau_a \frac{du_{вх}}{dt}, \qquad (4.127)$$

где $\tau_a = RC$ — *постоянная времени цепи.*

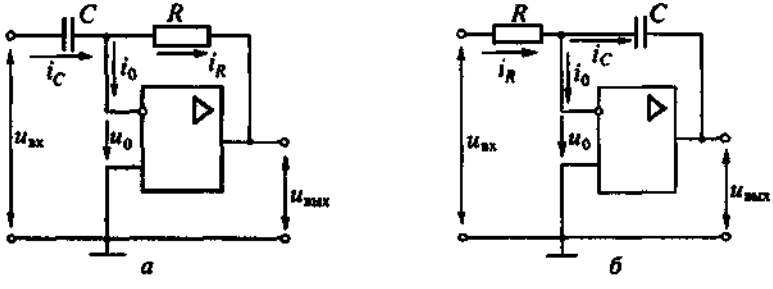



**Рис. 4.43.** Устройства на ОУ:
*а* — дифференцирующее; *б* — интегрирующее

Таким образом, схема на рис. 4.43, *а* производит дифференцирование входного сигнала. Дифференцирующее устройство применяют в интегральных импульсных устройствах.

Интегрирующее устройство (интегратор). Поскольку в схеме рис. 4.43, *б* ток $i_0 = 0$, находим $i_R = u_{вх}/R$, $i_C = -C du_{вых}/dt$. Приравняв эти токи и интегрируя, получим

$$u_{вых} = -\frac{1}{RC}\int_0^t u_{вх}\,dt = -\frac{1}{\tau_a}\int_0^t u_{вх}\,dt, \qquad (4.128)$$

т. е. данное устройство на ОУ осуществляет интегрирование входного сигнала.

На основе интеграторов выполняют *генераторы линейно изменяющегося напряжения*, широко использующиеся в различных радиотехнических устройствах, например в качестве генераторов напряжения разверток электронно-лучевых осциллографов, телевизионных систем и пр.

### Импульсные и избирательные усилители

**Импульсные (широкополосные) усилители** осуществляют усиление импульсных сигналов и сигналов широкого спектра частот. В процессе разработки импульсных (и широкополосных) усилителей важно прежде всего определить величину искажений формы усиливаемого тестового импульса — прямоугольного (рис. 4.44, *а*). Допустимые искажения формы импульса характеризуются максимально возможными длительностями переднего фронта $\tau_ф$ и среза $\tau_c$, измеренными на уровне от 0,1 до 0,9 $U_{вых}$, а также максимально допустимым спадом плоской вершины $\Delta U$ выходного импульса (рис. 4.44, *б*).

Известно, что вершину импульса (медленное изменение напряжения) определяет низкочастотная часть, а его передний и задний фронты (быстрое изменение напряжения) — высокочастотная часть спектра сигнала.

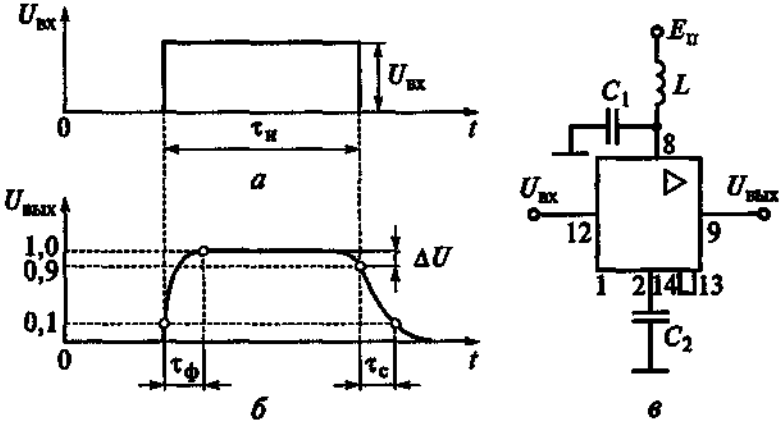



**Рис. 4.44.** Импульсный усилитель:
*а, б* — форма импульсов на входе и выходе;
*в* — схема с коррекцией на быстродействующем ОУ

Следовательно, для неискаженной передачи импульсов прямоугольной формы верхняя граничная частота полосы пропускания усилителя $f_в$ должна стремиться к бесконечности, а низшая $f_н$ — к нулю. Обычно требуемая полоса пропускания импульсных усилителей достигается введением в схему ОУ внешних цепей *низкочастотной* и *высокочастотной коррекций*.

Большинство импульсных усилителей выполняют на основе ОУ. Разделительные конденсаторы в усилителях используют для связи с источником входного сигнала и нагрузкой, поэтому нижняя граничная частота усиления усилителя близка к нулю. Увеличение верхней граничной частоты достигают технологическими методами, обеспечивающими создание высокочастотных интегральных транзисторов и малых паразитных емкостей внутри каскадов усилителя.

Высокочастотная коррекция осуществляется включением в цепь питания ОУ небольших по значениям индуктивности $L$ и конденсатора $C$, образующие вместе с емкостью нагрузки $C_н$ усилителя параллельный колебательный контур. В результате емкостной характер сопротивления нагрузки компенсируется индуктивным характером сопротивления цепи питания. Физическая сущность высокочастотной коррекции заключается во влиянии индуктивности $L$ (от единиц до сотен микрогенри) на скорость изменения тока нагрузки усилителя. В моменты усиления фронтов импульсов емкость нагрузки $C_н$ заряжается или разряжается токами большей величины, чем в отсутствие корректирующей индуктивности $L$. При этом напряжение на емкости $C_н$ (а значит, и на нагрузке $R_н$) меняется более резко, а следовательно, уменьшается длительность фронтов импульса, что приводит к увеличению верхней граничной частоты.

Низкочастотную коррекцию применяют для изменения с помощью шунтирующих *RC*-цепей сопротивления между соответствующими точками схемы широкополосного усилителя на высоких частотах. Как правило, низкочастотную коррекцию осуществляют либо изменением передаточной характеристики одного из элементарных каскадов, либо изменением характера входного импеданса ОУ. В любом случае при низкочастотной коррекции изменяется в некоторых пределах АЧХ усилительного каскада.

На рис. 4.44, *в* в качестве примера представлена принципиальная схема импульсного усилителя на быстродействующем ОУ, в которой для высокочастотной коррекции используются навесные элементы $L$, $C_1$, $C_2$. Возможно включение и цепей низкочастотной коррекции.

**Избирательные усилители** предназначены для усиления узкополосных сигналов. Как правило, отношение граничных частот рабочей полосы избирательного усилителя $f_в/f_н = 1{,}001...1{,}005$. Их АЧХ должна иметь достаточно резкие, близкие к прямоугольным, спады на границах полосы пропускания. По используемому частотному диапазону избирательные усилители делятся на два класса — резонансные и с частотно-зависимой ОС.

В простейшей схеме транзисторного *резонансного усилителя* с ОЭ нагрузкой коллекторной цепи служит параллельный *LC*-контур (рис. 4.45, *а*). Связь с последующим усилительным каскадом или нагрузкой в таком усилителе чаще всего осуществляется через разделительный конденсатор. Может также использоваться и высокочастотная трансформаторная связь.

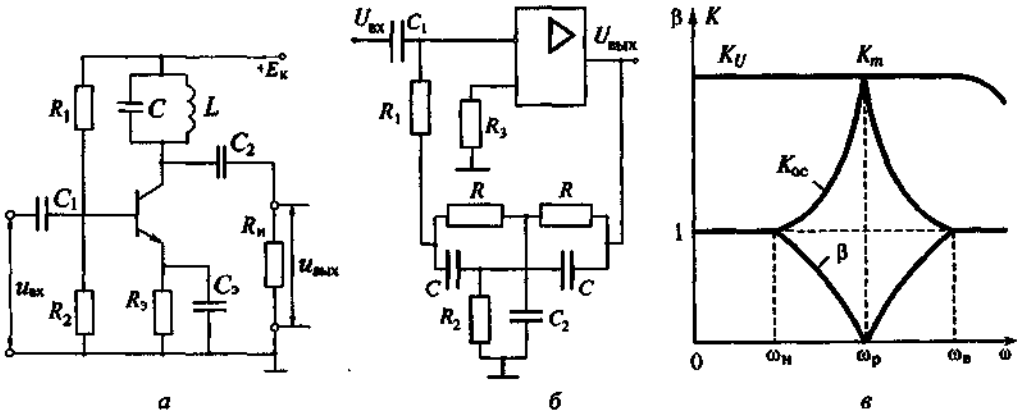



**Рис. 4.45.** Избирательные усилители:
*а* — резонансный; *б* — с частотно-зависимой ОС; *б* — частотные характеристики

Коэффициент усиления резонансного каскада с ОЭ определяется по формуле (4.75) с заменой сопротивления $R_{кн}$ на резонансное сопротивление контура $R_0$: $K_U = h_{21}R_0/h_{11}$. Резонансные усилители применяются на промежуточных и высоких частотах (свыше сотен кГц). Они выполняются обычно на интегральных микросхемах, которые содержат все элементы принципиальной схемы, кроме колебательного контура (на сравнительно низких частотах).

В диапазонах рабочих частот до нескольких десятков килогерц резонансные *LC*-контуры не используют из-за больших габаритов конденсаторов и катушек индуктивностей. Поэтому на достаточно низких частотах применяют избирательные усилители с *частотно-зависимой* ОС, состоящей из *RC*-цепей.

На рис. 4.45, *б* приведена схема избирательного усилителя на микросхеме с частотно-зависимой ООС в виде *двойного Т-образного моста*. Пусть усилитель с коэффициентом усиления $K_m$ имеет АЧХ $K_U(\omega)$ в области низких и средних частот (рис. 4.45, *б*). Известно, что коэффициент передачи двойного *Т*-образного моста $\beta(\omega) = \beta$ существенно зависит от частоты (рис. 4.45, *в*). Так, при частотах входного сигнала $\omega$, отличных от некоторой частоты $\omega_р$, коэффициент передачи $\beta \to 1$ и усилитель оказывается охваченным глубокой ООС, а коэффициент его усиления при больших значениях $K_m$ согласно (4.99) можно записать в виде

$$K_{oc} = \frac{K_m}{1 + K_m\beta} = \frac{K_m}{1 + K_m} \approx 1. \qquad (4.129)$$

При приближении частоты входного сигнала $\omega$ к частоте $\omega_р$ коэффициент передачи $\beta$ цепи ОС уменьшается, вызывая ослабление ООС и увеличение коэффициента усиления $K_{oc}$. На частоте $\omega_р$ влияние ООС на параметры усилителя исчезает и коэффициенты $\beta = 0$, $K_{oc} = K_m$. Частоту $\omega_р$ называют *квазирезонансной*. Данные свойства двойного *Т*-образного моста проявляются при определенных соотношениях параметров, например когда $R_2 = R/2$, $C_2 = 2C$. При этом квазирезонансная частота $\omega_р = 1/(RC)$. В схеме на рис. 4.45, *б* резистор $R_3$ защищает от пробоя неинвертирующий вход ОУ, а $R_1$ задает коэффициент усиления $K_m$.

## 4.7. Электрические фильтры

В радиоэлектронных устройствах широко применяют такой вид линейных цепей (четырехполюсников), как *электрические фильтры* (*filter*; от лат. *filtrum* — войлок), предназначенные для выделения (пропускания) или подавления (ослабления) сигналов с заданным спектром частот. Электрический фильтр — линейная стационарная система, т. е. свойства фильтра не зависят от времени.

Выбор типа и параметров фильтра определяют целями, поставленными при обработке сигнала. Требования к фильтру формируют в процессе анализа исходных данных о спектральном составе полезного сигнала, мешающих сигналов и шумов. Фильтр должен обеспечить максимальное подавление помех при минимально допустимых искажениях полезного сигнала. К основным характеристикам фильтра относят импульсную характеристику, передаточную характеристику (частотный коэффициент передачи), АЧХ, его порядок. В качестве примера на рис 4.46 приведены временны́е диаграммы и спектры сигналов и помех до и после фильтровой обработки.

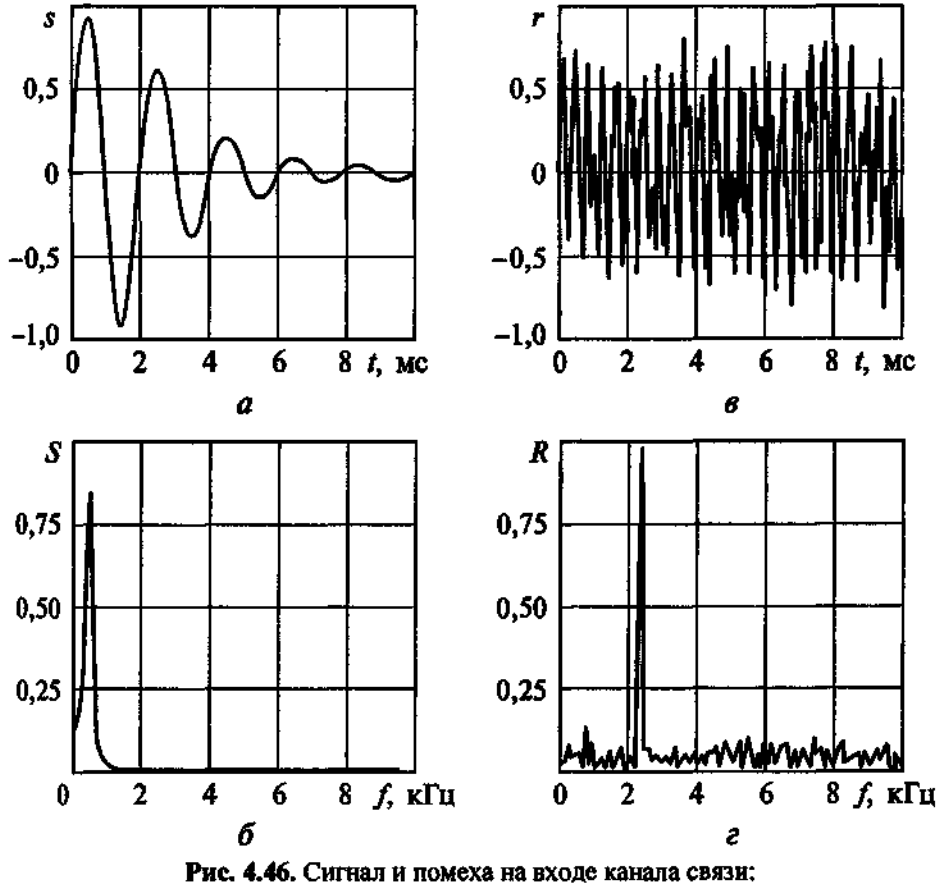



**Рис. 4.46.** Сигнал и помеха на входе канала связи:
*а* — сигнал; *б* — спектр сигнала; *в* — помеха; *г* — спектр помехи

В данном, достаточно типичном случае для системы связи полезный сигнал (временна́я диаграмма сигнала $s(t)$ и его спектр $S(f)$ представлены на рис. 4.46, *а*, *б*) представляет собой сравнительно низкочастотный процесс (порядка одного килогерца). В канале связи на сигнал могут оказать влияние высокочастотные пульсации (частота около 2,5 кГц) с шумом, т. е. помеха (временна́я диаграмма помехи $r(t)$ и ее спектр $R(f)$ показаны на рис. 4.46, *в*, *г*).

Итак, пусть на входной сигнал $s(t)$ аддитивно наложена помеха $r(t)$ (временна́я диаграмма этой смеси $u(t) = s(t) + r(t)$ и структура ее вещественного спектра $U(f) = S(f) + R(f)$ представлены на рис. 4.47, *а*, *б*). Чтобы разделить полезный низкочастотный сигнал и высокочастотные помехи, надо использовать фильтр с соответствующей полосой пропускания. АЧХ анализируемого фильтра показана штриховой линией на рис. 4.47, *б*. Как видно из диаграммы выходного сигнала $y(t)$ и его спектра $Y(f)$ (рис. 4.47, *в*, *г*), фильтр пропускает полезный сигнал и существенно ослабляет помехи, а также практически полностью подавляет высокочастотные пульсации сигнала с частотой 2,5 кГц. Полезный сигнал при этом несколько искажается. В частности, нетрудно заметить, что выходной сигнал $y(t)$ имеет запаздывание относительно полезной составляющей сигнала на входе $s(t)$ — штриховая линия на рис. 4.47, *в*.

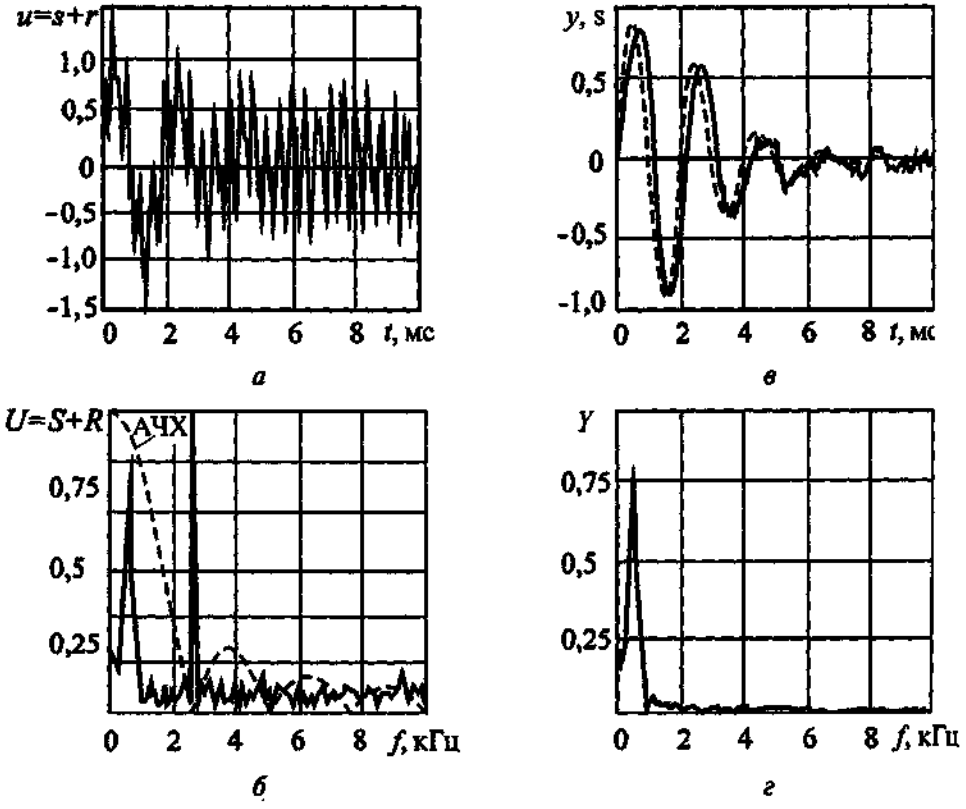



**Рис. 4.47.** Сигнал и помеха на входе и выходе фильтра:
*а* — сигнал в помехах; *б* — спектр сигнала в помехах;
*в* — сигнал после фильтрации; *г* — спектр профильтрованного сигнала

Область частот, в которой фильтры обладают малым ослаблением (затуханием), называют *полосой пропускания* (*pass band*). Область частот, в которой фильтры существенно ослабляют входной сигнал, определяют как *полосу задерживания* (*заграждения*, *подавления*; *stop band*).

Под *идеальным фильтром* понимают линейный четырехполюсник, у которого АЧХ имеет прямоугольную форму (рис. 4.48, *а*). Однако прямоугольные АЧХ *заведомо нереализуемы*. В реальных фильтров они лишь приближаются к идеальным с той или иной степенью точности в зависимости от их структуры.

По характеру расположения полосы пропускания и полосы задерживания фильтры делятся на четыре основных вида.

*Фильтры нижних частот* (*ФНЧ*; *low-pass filters*), полоса пропускания которых расположена в области частот от $\omega = 0$ до некоторой граничной верхней частоты $\omega_в$ (рис. 4.48, *б*).

*Фильтры верхних частот* (*ФВЧ*; *high-pass filters*), полоса пропускания которых простирается от некоторой граничной нижней частоты $\omega_н$ до бесконечности (рис. 4.48, *в*).

*Полосовые фильтры* (*ПФ*; *bandpass filters*), имеющие полосу пропускания в области между граничной нижней частотой $\omega_н$ и граничной верхней частотой $\omega_в$ (рис. 4.48, *г*).

*Режекторные*, или *заграждающие*, *фильтры* (*РФ*; *rejector filter*), полоса задерживания которых расположена в области частот от некоторой граничной нижней частоты задерживания $\omega_{зн}$ до некоторой граничной верхней частоты задерживания $\omega_{зв}$ (рис. 4.48, *д*).

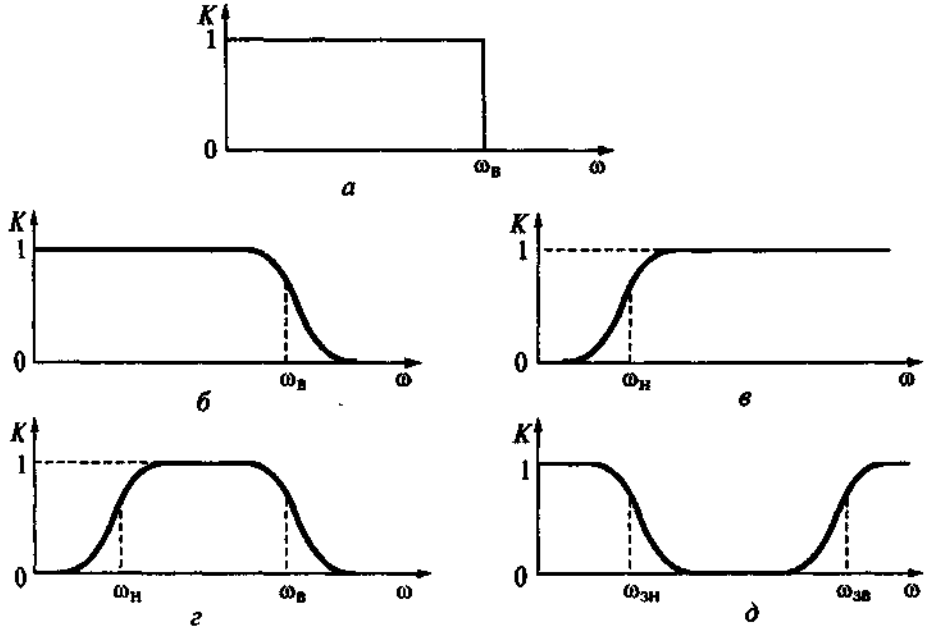



**Рис. 4.48.** Амплитудно-частотные характеристики фильтров:
*а* — идеального; *б* — нижних частот; *в* — верхних частот; *г* — полосового; *д* — режекторного

В теории фильтров граничные частоты называют *частотами среза* (*cutoff frequency*) и обозначают через $\omega_c$.

**Фильтры нижних частот**. Наиболее простейшим типом ФНЧ, уже рассмотренным в ранее, является интегрирующая цепь (см. рис. 4.10, *а*). Однако АЧХ интегрирующей цепи имеет довольно пологий и длительный спад в области верхних частот (см. рис. 4.10, *б*), что часто не обеспечивает заданного ослабления или подавления мешающих сигналов или помех.

Обратимся к частотному коэффициенту передачи по мощности (частотной характеристике), представляющему собой квадрат модуля частотного коэффициента передачи линейного четырехполюсника $K_P(\omega) = |\mathbf{K}(\omega)|^2$. В отличие от комплексного частотного коэффициента передачи $\mathbf{K}(\omega)$ функция $K_P(\omega)$ вещественна, и поэтому удобна при анализе частотных характеристик фильтров.

Частотный коэффициент передачи по мощности идеального ФНЧ для физических (положительных) частот (рис. 4.49, *а*) описывается выражением:

$$K_P(\omega) = \begin{cases} 1, & 0 \le \omega \le \omega_c; \\ 0, & \omega > \omega_c. \end{cases} \qquad (4.130)$$

В теории фильтров идеальные прямоугольные АЧХ аппроксимируют (*аппроксимация* — замена одних функций другими, более простыми и в том или ином смысле близкими к исходным; например кривых линий близкими к ним ломаными; от лат. *approximo* — приближаюсь) различными функциональными зависимостями. По оси абсцисс откладывают безразмерную (нормированную) частоту $x = \omega/\omega_c$, а по оси ординат — коэффициент передачи по мощности $K_P(x)$.

Для аппроксимации идеальной частотной характеристики ФНЧ часто используют известную в математике функцию — *полином Баттерворта* (Стефан Баттерворт; *Stephen Butterworth* — английский инженер; 1885–1958)

$$K_P(x) = \frac{1}{1 + x^{2n}}. \qquad (4.131)$$

ФНЧ, построенные на основе этой функции, называются *фильтрами с максимально плоской характеристикой*, или *фильтрами Баттерворта* (*filtro Butterworth*). Целое число $n$ = 1, 2, 3, ... в формуле (4.131) определяет *порядок* (в данном случае порядок полинома) фильтра.

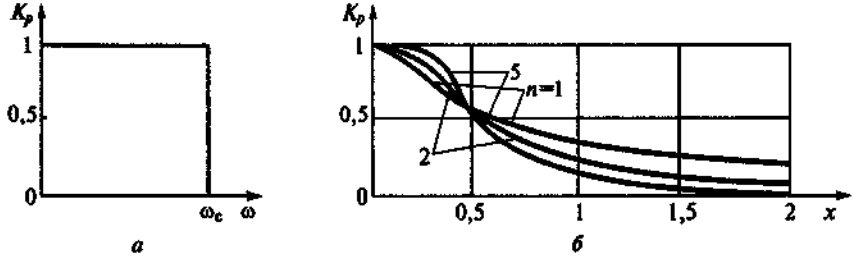



**Рис. 4.49.** Частотные характеристики фильтров:
*а* — идеального; *б* — Баттерворта


П.Л. Чебышев

На нормированной частоте среза ($x = 1$) ослабление сигнала по мощности, вносимое фильтром любого порядка, равно 1/2 (по напряжению $1/\sqrt{2}$). Часто ослабление рассчитывают в логарифмических единицах, тогда на частоте среза оно составит $\Delta = 10\lg 0{,}5 = -3$ дБ. Чем больше порядок фильтра, тем точнее аппроксимируют идеальную (прямоугольную) форму частотной характеристики ФНЧ. На рис. 4.49, *б* показаны графики функций (4.131), построенные для нескольких значений порядка $n$. Ослабление оценивают в специфических точках частотной характеристики, где частота превышает частоту среза вдвое (на *октаву*) и в десять раз (на *декаду*). Пусть частота входного сигнала существенно превышает частоту среза фильтра ($x >> 1$). Тогда из формулы (4.131) получим $K_P(x) \approx 1/x^{2n} = x^{-2n}$. При этом ослабление [дБ]

$$\Delta = 10\lg K_P(x) = -20n\lg x.$$

Значит, при увеличении частоты входного сигнала вдвое ($x = 2$) ослабление [дБ/октаву], вносимое фильтром Баттерворта, составит

$$\Delta_0 = -20n\lg 2 = -20n \cdot 0{,}301 = -6n.$$

Если же частота сигнала увеличивается в 10 раз, то ослабление [дБ/декаду]

$$\Delta_д = -20n\lg 10 = -20n \cdot 1 = -20n.$$

**Пример 4.11.** Низкочастотный фильтр Баттерворта с частотой среза характеристики $\omega_c = 10^5\ с^{-1}$ обеспечивает на частоте $\omega = 3 \cdot 10^5\ с^{-1}$ ослабление –30 дБ по отношению к ослаблению на $\omega = 0$. Определить порядок фильтра.

Р е ш е н и е. Находим нормированную частоту $x = 3 \cdot 10^5/10^5 = 3$. Подставляя соответствующие значения в (4.131), запишем

$$10\lg[1/(1 + 3^{2n})] = -30,$$

или

$$1 + 3^{2n} = 10^3 = 1\,000.$$

Решая это уравнение, находим $2n = \lg 999/\lg 3$. Откуда $n \approx 3$, что и определяет порядок фильтра.

Точнее идеальная характеристика ФНЧ для нормированных частот $x \le 1$ аппроксимируется *полиномом Чебышева* (Пафнутий Львович Чебышев — великий русский математик; 1821–1894) $n$-го порядка

$$T_n(x) = \cos(n\arccos x). \qquad (4.132)$$

В этом случае частотная характеристика ФНЧ имеет следующий вид

$$K_P(x) = \frac{1}{1 + \varepsilon^2 T_n^2(x)}, \qquad (4.133)$$

где параметр $\varepsilon \le 1$ — *коэффициент неравномерности частотной характеристики.*

При аппроксимации полиномами Чебышева необходимо учитывать тот факт, что чем меньше коэффициент ε, тем точнее аппроксимируется частотная характеристика в полосе пропускания ($x \le 1$), однако это снижает крутизну спада частотной характеристики в полосе задерживания ($x > 1$).

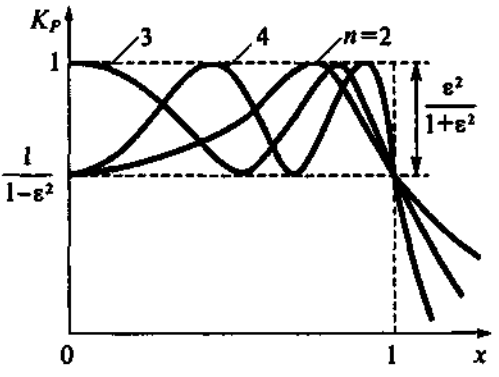



**Рис. 4.50.** Частотные характеристики фильтров Чебышева для $n = 2$, 3 и 4

Полиномы Чебышева низших порядков записываются в виде:

$$T_0(x) = 1; \quad T_1(x) = x; \quad T_2(x) = 2x^2 - 1; \quad T_3(x) = 4x^3 - 3x;$$

$$T_4(x) = 8x^4 - 8x^2 + 1; \quad T_5(x) = 16x^5 - 20x^3 + 5x. \qquad (4.134)$$

На рис. 4.50 показаны частотные характеристики чебышевских ФНЧ для порядков $n = 2$, 3 и 4. Как следует из графиков, в полосе пропускания частотные характеристики ФНЧ чебышевского типа имеют пульсирующий характер. Амплитуда пульсаций при этом $\Delta K_P = \varepsilon^2/(1+\varepsilon^2)$, и тем она ниже, чем меньше коэффициент ε. Такую аппроксимацию частотных характеристик фильтров в радиоэлектронике и технике связи называют *равноволновой.*

Вне полосы пропускания (при нормированных частотах среза $x > 1$, или $\omega > \omega_c$) частотные характеристики фильтра Чебышева монотонно убывают по закону

$$K_P(x) \approx \frac{1}{\varepsilon^2 |T_n(x)|^2}. \qquad (4.135)$$

При этом полином Чебышева $n$-го порядка описывается гиперболическим косинусом

$$T_n(x) = \mathrm{ch}(n \operatorname{arch} x). \qquad (4.136)$$

**Пример 4.12.** На нормированной частоте среза ФНЧ Чебышева 3-го порядка ослабляет мощность в два раза. Определить величину ослабления на частоте, равной трем нормированным частотам среза. Сравнить эффективность ослабления сигналов низкочастотными фильтрами Чебышева и Баттерворта, имеющими одинаковый порядок.

Р е ш е н и е. Согласно формуле (4.132), функция $T_n(1) = 1$ для любого порядка. Поэтому, приравняв в выражении (4.133) $K_P(1) = 0{,}5$, находим $\varepsilon = 1$.

Подставляя в полином Чебышева 3-го порядка заданную нормированную частоту $x = 3$, получим

$$T_3(x) = 4x^3 - 3x = 99.$$

Ослабление, вносимое фильтром Чебышева с коэффициентом неравномерности $\varepsilon = 1$ на частоте $\omega = 3\omega_c$:

$$\Delta_1 = 10\lg\frac{1}{1+99^2} \approx -40 \text{ дБ}.$$

Ослабление, вносимое ФНЧ Баттерворта при таких же параметрах, составит

$$\Delta_2 = 10\lg\frac{1}{1+3^6} \approx -28{,}5 \text{ дБ}.$$

Сравнение результатов показывает, что при одной и той же степени сложности фильтров одного порядка ослабление сигналов фильтром Баттерворта на 11,5 дБ меньше, чем фильтром Чебышева.

**Фильтры верхних частот.** Одним из рассмотренных ранее типов ФВЧ является дифференцирующая *RC*-цепь, форма АЧХ (см. рис. 4.8) которой далека от идеальной (т. е. не прямоугольна). Аппроксимацию идеальной частотной характеристики ФВЧ получают на основе полиномов Баттерворта и Чебышева. Обратимся к функции (4.131) и введем новую нормированную частоту

$$\nu = 1/x = \omega/\omega_c. \tag{4.137}$$

Тогда

$$K_P(x) = \frac{1}{1+(1/\nu)^{2n}} = \frac{\nu^{2n}}{1+\nu^{2n}} = K_P(\nu). \tag{4.138}$$

Алгебраическую функцию $K_P(\nu)$ можно рассматривать как частотную характеристику ФВЧ Баттерворта, обладающую в нормированной полосе частот $1 < \nu < \infty$ такой же неравномерностью ослабления, что и функция $K_P(x)$ фильтра нижних частот в полосе $0 < x < 1$. Аналогичным образом можно получить частотную характеристику ФВЧ Чебышева, заменив аргумент $x$ на $\nu$ в соотношении (4.133) и используя полином гиперболического вида (4.136)

$$K_P(\nu) = \frac{1}{1+\varepsilon^2 T_n^2(\nu)}. \tag{4.139}$$

Полученные с помощью соотношений (4.138) и (4.139) графики частотных характеристик фильтров высоких частот Баттерворта и Чебышева четвертого порядка показаны на рис. 4.51. Сопоставление графиков показывает, что частотная характеристика фильтра высоких частот Чебышева обладает более крутым скатом переднего фронта, чем частотная характеристика фильтра высоких частот Баттерворта. Однако вершина частотной характеристики фильтра высоких частот Чебышева имеет существенно пульсирующий характер, что относится к его недостаткам.

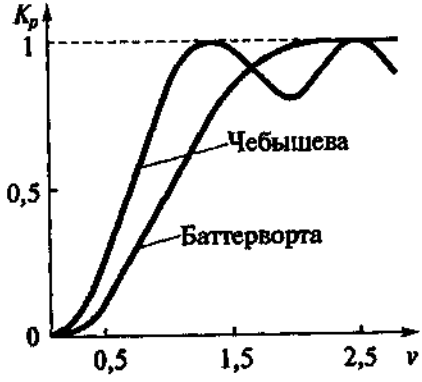



**Рис. 4.51.** Частотные характеристики ФВЧ 4-го порядка

### Фильтры на ПАВ

К используемым в настоящее время в различных устройствах радиоэлектроники и техники связи видам фильтров следует отнести:

- *LC*-фильтры;
- монокристаллические кварцевые и танталато-литиевые фильтры;
- воздушные резонаторные фильтры;
- керамические коаксиальные фильтры;
- полосковые и микрополосковые фильтры;
- программируемые и перестраиваемые фильтры.

Все эти фильтры обладают различными недостатками, которые в той или иной степени не удовлетворяют определенным требованиям фильтрующим устройствам. Поэтому в последние несколько лет широкое применение в приемопередающих устройствах систем связи находят фильтры на поверхностных акустических волнах, свободные от многих недостатков.

*Поверхностные акустические волны* (ПАВ) — это упругие (необъемные) волны, распространяющиеся с невысокой скоростью вдоль свободной поверхности твердого тела или вдоль границы твердого тела с другими средами и затухающие при удалении от границ. В качестве среды распространения используются различные пьезоэлектрические монокристаллы: кварц, ниобат лития, танталат лития, германат висмута и др.

Поверхностные акустические волны занимают диапазон длин волн от $10^{-5}$ до $10^{-1}$ см, а их частоты соответствуют области ультразвука (от 20 кГц до 1 ГГц; частоты до 100 ГГц — *гиперзвук*). Замечательным свойством поверхностных акустических волн является их очень невысокая (от 3 000 до 4 000 м/с, т. е. менее 1/100 000 скорости света) в сравнении с электромагнитными волнами скорость распространения, что позволяет применять к ним практически любые математические способы обработки сигнала.

Внешний вид фильтра на ПАВ изображен на (рис. 4.52). Основой фильтра на ПАВ (это герметизированный электронный прибор) является пьезоэлектрическая подложка.




**Рис. 4.52.** Внешний вид фильтра на ПАВ

Для возбуждения и детектирования ПАВ служат излучающие и приемные пленочные *встречно-штыревые преобразователи* (ВШП; англ. *Inter Digital Transducer — IDT*), нанесенные на пьезоэлектрическую подложку методом фотолитографии (число штырей может быть различным — от единиц до нескольких тысяч) и являющихся по существу дискретной (цифровой) структурой.

Под действием высокочастотного электрического напряжения источника обрабатываемого сигнала $u(t)$ в зазорах между смежными штырями излучающего ВШП (рис. 4.52) возникает переменное электрическое поле, которое вследствие пьезоэффекта материала подложки вызывает механические упругие колебания в ее поверхностном слое. Эти колебания распространяются в тонком приповерхностном слое подложки в направлениях, перпендикулярных электродам в виде поверхностных акустических волн.

Обработка сигнала в заданной полосе частот осуществляется вследствие многократной интерференции ПАВ от большого числа отражателей ВШП, имеющих разные размеры, геометрию и относительное взаимное расположение. Затем обработанные акустические сигналы вновь преобразуются приемным ВШП в электрические и поступают в нагрузку.

Большинство преимуществ фильтров на ПАВ обусловлено непосредственно их физической структурой: практическим отсутствием энергопотребления; возможностью выполнения различных операций обработки сигналов; линейной (или определяемой требованиями) фазой выходного сигнала; очень высокой прямоугольностью АЧХ; исключительным внеполосным подавлением паразитных составляющих; реализацией заданных технических характеристик с высокой точностью; высокой надежностью; малыми габаритными размерами и массой; температурной стабильностью. Поскольку центральная частота и форма частотной характеристики определяются топологией, они не требуют сложной настройки в аппаратуре и не могут расстроиться в процессе эксплуатации. Технология изготовления, совместимая с интегральной технологией, позволяет выпускать их в большом объеме с высокой воспроизводимостью.

Фильтры на ПАВ применяются на частотах от 1 МГц до 3 ГГц с относительной полосой пропускания от 0,1 до 90 %. На очень низких частотах габариты фильтров становятся слишком большими, поэтому вместо них находят применение монолитные фильтры на объемных волнах, выполненные из пьезоэлектрической керамики.

Тем не менее фильтрам на ПАВ свойственны и определенные недостатки:

- наличие паразитных всплесков АЧХ на кратных частотах;
- монотонное снижение коэффициента подавления составляющих по мере повышения частоты; данный недостаток устраняется включением внешних реактивных элементов (индуктивностей);
- довольно значительные потери полезных составляющих в полосе пропускания (до 25…30 дБ);
- повышенная чувствительность к статическим зарядам электричества.

И все же имеющиеся преимущества обеспечивают широкое применение и массовую потребность в фильтрах и других устройствах на ПАВ, и прежде всего, в современных системах подвижной связи.

### Синтез электрических фильтров

Под *синтезом линейной цепи* (здесь — электрического фильтра) понимают определение структуры и параметров элементов ее схемы, обеспечивающих заданные свойства. Как правило, синтез такой цепи — неоднозначная процедура, поскольку заданным параметрам и характеристикам могут отвечать несколько структур. Желательно, чтобы синтезируемая цепь содержала минимально возможное число элементов и была в некотором смысле оптимальна. В задачу курса «Основы радиоэлектроники и связи» не входит общая теория синтеза фильтров, так как она изучается в теоретических основах электротехники. Поэтому рассмотрим лишь ряд вопросов. В частности, проведем синтез фильтров по заданному частотному коэффициенту передачи, разделив его на два этапа.

На первом этапе синтеза выбирают идеализированную функцию, достаточно точно описывающую частотный коэффициент передачи по мощности. Обычно фильтр, отвечающий идеализированной частотной характеристике, практически неосуществим (физически нереализуем). Поэтому этот этап включает также аппроксимацию выбранной идеальной частотной характеристики такой функцией, которая может принадлежать физически реализуемому фильтру.

Второй этап синтеза заключается в практической реализации фильтра (разработке принципиальной схемы), обладающего полученной частотной характеристикой. Поскольку первый этап синтеза фильтров, связанный с подбором аппроксимирующих функций (многочлены Баттерворта, Чебышева) был ранее рассмотрен, перейдем ко второму этапу. Основой реализации служит *передаточная функция* (*операторный коэффициент передачи*) фильтра $K(p)$. Эта функция отражает результат аналитического переноса комплексного частотного коэффициента передачи $\mathbf{K}(\omega)$ с мнимой оси $j\omega$ на всю область комплексных частот $p = \alpha + j\omega$ ($\alpha$ — вещественная часть комплексной частоты)

$$K(p) = \frac{R(p)}{Q(p)} = K_0 \frac{(p - z_1)(p - z_2)\cdots(p - z_m)}{(p - p_1)(p - p_2)\cdots(p - p_n)}, \qquad (4.140)$$

где $K_0$ — постоянная величина; $z_1$, $z_2$, ..., $z_m$ — *нули* функции $K(p)$, т. е. корни уравнения числителя $R(p) = 0$; $p_1$, $p_2$, ..., $p_n$ — *полюсы* функции $K(p)$, т. е. корни уравнения знаменателя $Q(p) = 0$; $m$ и $n$ — соответственно порядки полиномов $R(p)$ и $Q(p)$.

В теории синтеза фильтров рассматривают линейные четырехполюсники, передаточные функции которых имеют ограниченное число нулей и полюсов. Значит, фильтры содержат конечное число сопротивлений, индуктивностей и емкостей. Для упрощения расчетов вводят еще два дополнительных условия: число полюсов передаточной функции должно превышать число нулей; полюсы должны располагаться в левой полуплоскости. Если нули и полюсы передаточной функции заданы, то используют *структурный синтез*, при котором фильтр реализуют в виде каскадного включения $N$-го числа простейших звеньев (цепей) с передаточными функциями $K_1(p)$, $K_2(p)$, ... , $K_N(p)$, разделенных идеальными усилителями. Передаточная функция фильтра определяется произведением

$$K(p) = K_1(p)K_2(p)\cdots K_N(p). \qquad (4.141)$$

**Пример 4.13.** Необходимо синтезировать электрический фильтр, имеющий частотный коэффициент передачи

$$\mathbf{K}(\omega) = \frac{j\omega\tau_1}{1 + j\omega(\tau_1 + \tau_2) - \omega^2\tau_1\tau_2},$$

где $\tau_1$, $\tau_2$ — постоянные времени составляющих схему цепей.

Р е ш е н и е. Представим частотный коэффициент передачи в операторной форме (т. е. в виде передаточной функции), заменив параметр $j\omega$ на $p$:

$$K(p) = \frac{p\tau_1}{1 + p(\tau_1 + \tau_2) - p^2\tau_1\tau_2}.$$

Решая уравнение знаменателя $p^2\tau_1\tau_2 + p(\tau_1 + \tau_2) + 1 = 0$, находим полюсы функции: $p_1 = -1/\tau_1$; $p_2 = -1/\tau_2$. Из этих значений следует, что полюсы функции $K(p)$ находятся на вещественной оси $\alpha$ в левой полуплоскости (рис. 4.53, *а*).

Запишем передаточную функцию фильтра в виде произведения передаточных функций двух звеньев:

$$K(p) = \frac{p\tau_1}{\tau_1\tau_2(p + 1/\tau_1)(p + 1/\tau_2)} = K_1(p)K_2(p),$$

где

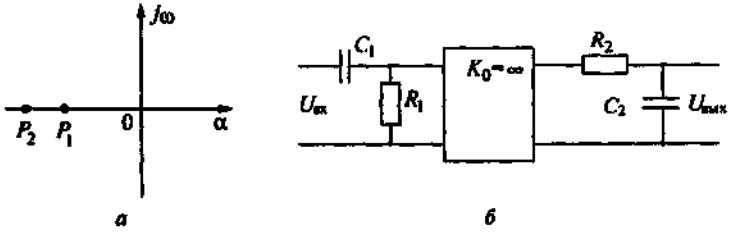
$$K_1(p) = \frac{p\tau_1}{1 + p\tau_1}; \qquad K_2(p) = \frac{1}{1 + p\tau_2}.$$

Перейдем от передаточных функций к частотным коэффициентам передачи:

$$\mathbf{K}_1(\omega) = \frac{j\omega\tau_1}{1 + j\omega\tau_1}; \qquad \mathbf{K}_2(\omega) = \frac{1}{1 + j\omega\tau_2}.$$

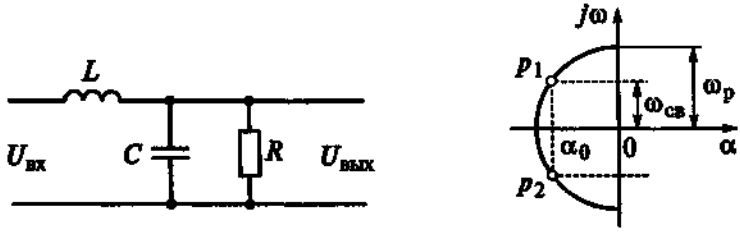
Данные выражения представляют собой частотные коэффициенты передачи соответственно дифференцирующей (4.35) и интегрирующей (4.40) цепей с постоянными времени $\tau_1 = R_1C_1$, $\tau_2 = R_2C_2$. Напомним, что в курсе ТОЭ подобные цепи, содержащие резисторы и реактивные элементы только с сопротивлениями одного знака (или индуктивности, или емкости), называются *звеном первого порядка* (*апериодической цепью*).

Схема фильтра, отвечающая заданному частотному коэффициенту передачи, показана на рис. 4.53, *б*. В фильтре применен идеальный развязывающий усилитель с коэффициентом усиления $K_0 = \infty$, включенный между двумя звеньями.




*а* *б*

**Рис. 4.53.** Синтезируемый фильтр:
*а* — координаты полюсов передаточной функции; *б* — схема




*а* *б*

**Рис. 4.54.** Звено второго порядка:
*а* — схема; *б* — координаты полюсов передаточной функции

Усложненные по структуре электрические фильтры можно реализовать на основе *звеньев второго порядка* (*колебательных цепей*), обязательно содержащих реактивные элементы с сопротивлениями противоположного знака (т. е. и индуктивности, и емкости) и имеющие в передаточной характеристике пару комплексно-сопряженных полюсов. Простейшим звеном второго порядка может служить колебательный контур, представленный на рис. 4.54, *а*.

Записав отношение выходного и входного гармонических напряжений одной частоты, определим передаточную функцию этого звена

$$K(p)=\frac{\omega_p^2}{p^2+2\alpha_0 p+\omega_p^2}, \qquad (4.142)$$

где $\omega_p = 1/\sqrt{LC}$ — резонансная частота; $\alpha_0 = 1/(2RC)$ — коэффициент затухания.

Решая уравнение $p^2 + 2\alpha_0 p + \omega_p^2 = 0$, находим полюсы функции $K(p)$:

$$p_{1,2} = -\alpha_0 \pm j\sqrt{\omega_p^2 - \alpha_0^2} = -\alpha_0 \pm j\omega_{св}, \qquad (4.143)$$

где $\omega_{св} = \sqrt{\omega_p^2 - \alpha_0^2}$ — частота свободных колебаний в контуре.

Простой анализ соотношений между резонансной частотой $\omega_p$ и коэффициентом затухания $\alpha_0$ показывает, что полюсы (рис. 4.54, *б*) могут быть как вещественными ($\omega_p = \alpha_0$), так и комплексно-сопряженными ($\omega_p \neq \alpha_0$).

### Активные фильтры

В области сравнительно низких частот (частота менее 100 кГц) в апериодических и колебательных звеньях электрических фильтров требуются большие емкости конденсаторов и индуктивности катушек, которые по массе и габаритным размерам значительно превышают интегральные микросхемы. Поэтому при построении электрических фильтров указанного радиодиапазона в качестве базовых элементов специалисты часто используют операционные усилители. Электрические фильтры, представляющие собой комбинации определенным образом соединенных *RC*-цепей и операционных усилителей, получили название *активных фильтров*.

Первые активные *RC*-фильтры на ОУ разработал Х. Скотт (США) в 1938 г.

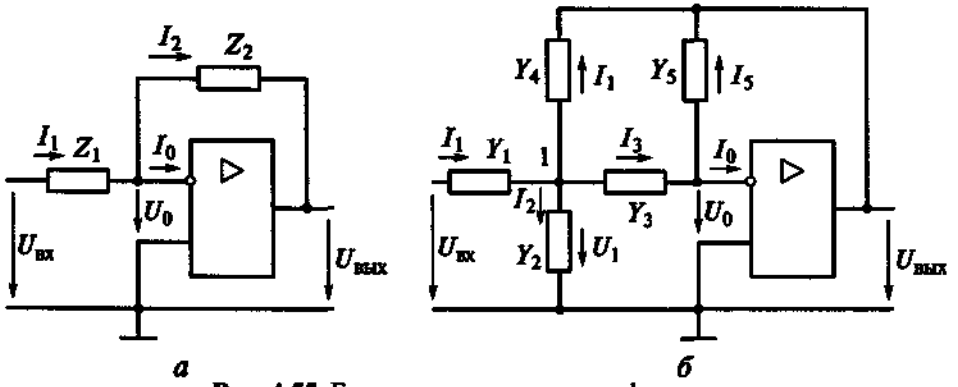



**Рис. 4.55.** Базовые схемы активных фильтров:
*а* — 1-го порядка; *б* — 2-го порядка

Проанализируем базовую схему *активного фильтра* 1-*го порядка* ( фильтра с *однопетлевой отрицательной обратной связью*) на рис. 4.55, *а*. В этой схеме $Z_1$ и $Z_2$ — операторные сопротивления (в них произведение $j\omega$ заменено на $p$) некоторых *RC*-цепей.

По аналогии с (4.122) для коэффициента усиления инвертирующего усилителя нетрудно определить, что передаточная функция схемы

$$K(p) = -Z_2/Z_1. \qquad (4.144)$$

На основе базовой схемы на рис. 4.55, *а* по данной передаточной функции можно строить различные простейшие фильтры.

В радиотехнике широкое применение находят *активные фильтры* 2-*го порядка*, имеющие *двухпетлевую* частотно-зависимую отрицательную обратную связь. Базовая схема такого фильтра приведена на рис. 4.55, *б*, в которой для удобства пассивные элементы обозначены операторными проводимостями *Y*. Определим передаточную функцию данной базовой схемы. В соответствии с первым законом Кирхгофа для всех токов в точке 1 имеем

$$I_1 = I_2 + I_3 + I_4. \qquad (4.145)$$

Учитывая, что для идеального ОУ $U_0 = 0$ и $I_0 = 0$, выразим токи через напряжения (они для упрощения обозначены как вещественные величины):

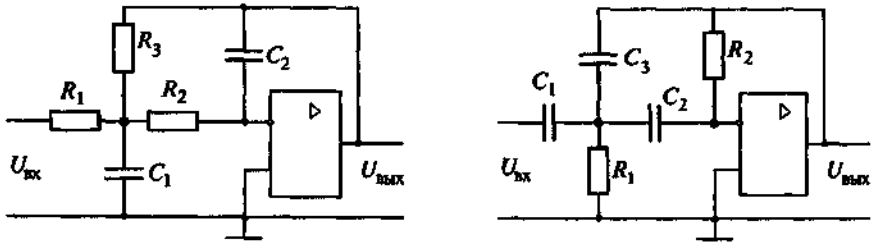
$$I_1 = (U_{вх} - U_1)Y_1; \quad I_2 = U_1Y_2; \quad I_3 = (U_1 - U_0)Y_3 = U_1Y_3;$$
$$I_4 = (U_1 - U_{вых})Y_4; \quad I_5 = (U_0 - U_{вых})Y_5 = -U_{вых}Y_5.$$

Подставив эти соотношения в (4.145), запишем

$$U_{вх}Y_1 = U_1(Y_1 + Y_2 + Y_3 + Y_4) - U_{вых}Y_4. \qquad (4.146)$$

Приравняв токи $I_3 = I_5$ (ведь $I_0 = 0$), получим

$$U_1 = -U_{вых}Y_5/Y_3.$$




*а* *б*

**Рис. 4.56.** Активные фильтры 2-го порядка:
*а* — ФНЧ; *б* — ФВЧ

Подставив это значение $U_1$ в (4.146), находим передаточную функцию

$$K(p)=\frac{U_{вых}}{U_{вх}}=-\frac{Y_1Y_3}{(Y_1+Y_2+Y_3+Y_4)Y_5+Y_3Y_4}. \qquad (4.147)$$

На основе базовой схемы активных фильтров на рис. 4.55, *б* можно получить различные виды электрических фильтров 2-го порядка, например фильтры нижних и верхних частот.

**Активный фильтр нижних частот 2-го порядка** (рис. 4.56, *а*) имеет следующие проводимости элементов:

$$Y_1 = 1/R_1;\quad Y_2 = j\omega C_1;\quad Y_3 = 1/R_2;\quad Y_4 = 1/R_3;\quad Y_5 = j\omega C_2.$$

Подставив значения проводимостей в (4.147), получим

$$K(p)=\frac{U_{вых}}{U_{вх}}=-\frac{Y_1Y_3}{(Y_1+Y_2+Y_3+Y_4)Y_5+Y_3Y_4}. \qquad (4.148)$$

Можно показать, что частота среза такого фильтра

$$\omega_c=\frac{1}{\sqrt{R_2R_3C_1C_2}}. \qquad (4.149)$$

**Пример 4.14.** Синтезировать активный фильтр нижних частот Баттерворта 2-го порядка для круговой частоты среза $\omega_c = 10^3\ с^{-1}$.

Р е ш е н и е. Передаточную функцию фильтра получим из формулы (4.131), заменив $x$ на $p$ и приняв $n = 2$,

$$K(p)=\frac{1}{1+p^4}.$$

Вычислив корни уравнения в знаменателе передаточной функции $p^4 + 1 = 0$, найдем ее полюсы:

$$p_1 = e^{j\pi/4};\quad p_2 = e^{j3\pi/4};\quad p_3 = e^{j5\pi/4};\quad p_4 = e^{j7\pi/4}.$$

Из полученных полюсов синтезируемому ФНЧ Баттерворта 2-го порядка отвечают только те, которые расположены в левой полуплоскости комплексной переменной. Их «зеркальные отражения» в правой полуплоскости связаны с функцией $K(-p)$ и во внимание не принимаются. Значит, передаточная функция фильтра будет определяться только двумя полюсами:

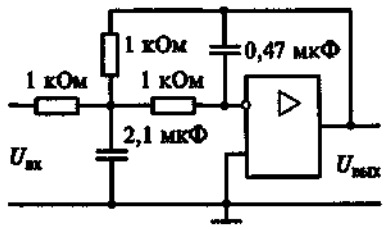



**Рис. 4.57.** Схема синтезированного ФНЧ

$$p_2 = e^{j3\pi/4} = \cos(3\pi/4) + j\sin(3\pi/4) = (-1+j)/\sqrt{2};$$

$$p_3 = e^{j5\pi/4} = \cos(5\pi/4) + j\sin(5\pi/4) = (-1-j)/\sqrt{2}.$$

Подставив данные полюса в (4.139), получим передаточную функцию

$$K(p) = \frac{1}{(p-p_2)(p-p_3)} = \frac{1}{p^2+\sqrt{2}p+1}. \qquad (4.150)$$

Следует учесть, что во все уравнения данного примера входит нормированная частота $x = \omega/\omega_c$, поэтому полюсы для текущей частоты ω будут равны:

$$p_{2н} = \omega_c(-1+j)/\sqrt{2}; \qquad p_{3н} = \omega_c(-1-j)/\sqrt{2}.$$

Теперь передаточная функция (4.150) имеет вид:

$$K(p_н) = \frac{1}{p_н^2/\omega_c^2 + \sqrt{2}p_н/\omega_c + 1}. \qquad (4.151)$$

Так как искомых величин много, зададимся приемлемыми значениями резисторов $R_1 = R_2 = R_3 = R = 1$ кОм. Тогда, приравняв коэффициенты при параметрах $p_н^2$ и $p_н$ в формулах (4.148) и (4.151), запишем:

$$1/\omega_c^2 = R^2C_1C_2, \quad \sqrt{2}/\omega_c = 3RC_2.$$

Подставив известные величины, получим искомые значения:

$$C_1 = 2{,}1 \text{ мкФ}, \qquad C_2 = 0{,}47 \text{ мкФ}.$$

Схема синтезированного по исходному заданию фильтра показана на рис. 4.57.

**Активный фильтр верхних частот 2-го порядка** (рис. 4.56, *б*) получают перестановкой местами резисторов и конденсаторов в схеме ФНЧ на рис. 4.56, *а*. Его передаточная функция запишется следующим образом:

$$K(p) = -\frac{p^2R_1R_2C_1C_2}{p^2R_1R_2C_2C_3 + pR_1(C_1+C_2+C_3)+1}. \qquad (4.152)$$

Частота среза при этом

$$\omega_c = \frac{1}{\sqrt{R_1R_2C_2C_3}}. \qquad (4.153)$$

Аналогично строят и другие типы активных фильтров с разными АЧХ.

## Контрольные вопросы

1. Из каких основных элементов состоят электрические цепи?
2. На какие классы делятся радиотехнические цепи?
3. Какие свойства присущи линейным радиотехническим цепям?
4. В чем заключается отличительное свойство динамических линейных систем?
5. Как определяют частотный коэффициент передачи линейной цепи?
6. В чем заключается спектральный метод анализа линейных цепей?
7. В каких задачах удобен спектральный метод анализа?
8. В чем состоит сущность операторного метода анализа прохождения сигналов через линейные цепи?
9. Запишите прямое и обратное преобразования Лапласа.
10. Что отражают импульсная и переходная характеристики цепи?
11. Как связаны импульсная и переходная характеристики линейной цепи?
12. Как связаны между собой импульсная характеристика и частотный коэффициент передачи линейного четырехполюсника?
13. В чем заключается суть анализа линейных цепей методом интеграла Дюамеля?
14. Какие цепи относятся к многополюсным и как проводится их анализ?
15. Какие цепи относятся к дифференцирующим?
16. Для каких целей применяют интегрирующие цепи?
17. Назовите основные параметры колебательных контуров.
18. Какими параметрами характеризуются связные контуры?
19. По каким признакам классифицируют усилители?
20. Приведите схему усилительного каскада на биполярном транзисторе с ОЭ.
21. Почему и как ООС по постоянному току термостабилизирует режим покоя работы транзистора в схеме ОЭ?
22. С помощью схемы замещения определите основные параметры каскада с ОЭ.
23. В каких случаях в усилителях возникают нелинейные искажения?
24. Приведите схемы усилительных каскадов ОК и ОБ и определите их параметры.
25. Приведите схемы усилительных каскадов ОИ и ОС.
26. Как строят бестрансформаторные усилители мощности?
27. Для каких целей применяются обратные связи в усилителях?
28. В каких случаях в усилителях возникает отрицательная, а в каких положительная обратная связь?
29. Что такое устойчивость линейной динамической системы?
30. В какой области комплексной плоскости должны располагаться полюсы передаточной функции устойчивой линейной системы?
31. На каком принципе основан алгебраический критерий устойчивости?
32. В чем заключается сущность критерия Найквиста?
33. Приведите и определите параметры фундаментальных схем включения ОУ.
34. Составьте схему транзисторного импульсного усилителя.
35. Приведите схему избирательного усилителя на ОУ.
36. На какие основные виды разделяются электрические фильтры?
37. Приведите АЧХ основных типов фильтров.
38. В чем заключается разница между свойствами и передаточными характеристиками фильтров Баттерворта и Чебышева?
39. На чем основан принцип структурного синтеза фильтров?
40. На чем основаны принципы построения активных фильтров?
41. Приведите основные базовые схемы активных фильтров.

# Глава 5. НЕЛИНЕЙНЫЕ И ПАРАМЕТРИЧЕСКИЕ ЦЕПИ

Большинство процессов (нелинейное усиление сигналов, модуляция, демодуляция, ограничение, генерация, умножение, деление и перенос частоты и т. д.), связанных с преобразованием спектра сигналов, осуществляют с помощью нелинейных и параметрических цепей. В нелинейных цепях параметры элементов зависят от входных воздействий, и процессы, протекающие в них, описывают нелинейными дифференциальными уравнениями. При этом к ним неприменим принцип суперпозиции. Эти цепи отличаются большим разнообразием и поэтому не существует общих методов их анализа. В книге анализ нелинейных цепей ограничен рассмотрением только их определенного класса. Это радиотехнические цепи, анализ которых проводится в основном с помощью вольт-амперных характеристик нелинейных элементов.

Промежуточное положение между линейными и нелинейными цепями занимают параметрические цепи, которые являются линейными и к которым применим принцип суперпозиции. Однако в спектре выходного сигнала таких цепей могут появиться новые частоты. Параметрические цепи описывают линейными дифференциальными уравнениями с переменными (т. е. зависящими от времени) коэффициентами. Теория этих уравнений по сравнению с теорией линейных уравнений с постоянными коэффициентами более сложна. Некоторые параметрические цепи работают в существенно нелинейном режиме. Это позволяет методологически объединить параметрические цепи с нелинейными цепями, тем более что результат обработки сигнала связан с преобразованием его спектра.

## 5.1. Общие сведения об аппроксимации характеристик нелинейных элементов

Нелинейными звеньями являются полупроводниковые и электровакуумные приборы, а также элементы с ферромагнитными материалами. В сочетании с инерционными электрическими цепями они образуют нелинейные динамические цепи и устройства. Как правило, инерционными электрическими цепями являются частотно-избирательные цепи. Нелинейные динамические системы описывают нелинейными дифференциальными уравнениями, в этих системах нелинейность обязательно присутствует. Нелинейную цепь можно определить не только по входящим в нее элементам, но и по внешним признакам, к числу которых при гармоническом входном сигнале $u_{вх}(t)$ относят:

• отличие от синусоидальной формы выходного сигнала $i_{вых}(t)$;
• появление в спектре выходного колебания гармоник входного сигнала;
• нелинейность передаточной амплитудной характеристики;
• зависимость фазы усиленного сигнала от амплитуды.

Известны и используют следующие методы анализа нелинейных цепей при прохождении через них детерминированных сигналов:

• линеаризация характеристик *нелинейного элемента* (НЭ) при фильтрации высших гармоник сигнала на выходе цепи;
• аналитические, как правило, приближенные способы решения системы нелинейных уравнений, описывающих работу устройства;
• спектральный, оценивающий нелинейные свойства цепи по спектру выходного сигнала;
• численные способы решения системы нелинейных уравнений с помощью компьютера;
• с помощью фазовой плоскости.

Наиболее часто используют метод анализа нелинейных цепей, основанный на линеаризации характеристик НЭ при фильтрации высших гармоник сигнала на выходе цепи.

*Линеаризация* (от лат. *linearis* — линейный) — метод приближенного представления замкнутых нелинейных систем, при котором исследование нелинейной системы заменяют анализом линейной системы, в некотором смысле эквивалентной исходной. Методы линеаризации имеют ограниченный характер, т. е. эквивалентность исходной нелинейной системы и ее линейного приближения сохраняется лишь при определенном «режиме» работы системы, а если система переходит из одного режима работы на другой, то следует изменить и ее линеаризированную модель. Вместе с тем, применяя линеаризацию, можно выяснить многие качественные и количественные свойства нелинейной системы.

В качестве примера нелинейных цепей, точнее элементов, можно привести полупроводниковый выпрямительный диод, оставляющий от синусоидального сигнала только однополярные (положительные или отрицательные) полусинусоиды, или трансформатор, насыщение сердечника которого магнитным полем приводит к «затуплению» вершин синусоиды (а с точки зрения частотного спектра, это сопровождается появлением гармоник основной частоты, а иногда и частот меньшей в кратное число раз основной частоты — *субгармоник*).

При использовании метода линеаризации анализ прохождения сигнала через нелинейную цепь сравнительно просто осуществить, если нелинейный элемент отвечает условиям *безынерционности*. Физически безынерционность НЭ означает мгновенное изменение отклика на его выходе вслед за изменением входного воздействия. Если говорить строго, то безынерционных (резистивных, или омических, т. е. поглощающих энергию сигнала) НЭ практически не существует. Все НЭ — диоды, транзисторы, микросхемы, электровакуумные приборы и т. д. — обладают инерционными свойствами. Вместе с тем, современные полупроводниковые приборы достаточно совершенны по своим частотным параметрам, и их удается идеализировать с точки зрения безынерционности.

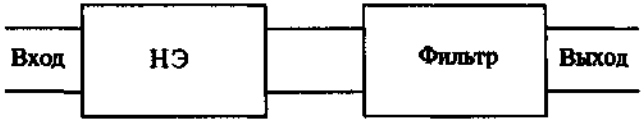



**Рис. 5.1.** Структурная схема нелинейного устройства

Большинство нелинейных радиотехнических цепей и устройств с безынерционном НЭ определяются структурной схемой, представленной на рис. 5.1. Из этой схемы следует, что входной сигнал непосредственно воздействует на НЭ, к выходу которого подключен электрический фильтр (линейная цепь). В этих случаях процесс в нелинейной цепи можно охарактеризовать двумя независимыми друг от друга операциями. В результате первой операции в безынерционном НЭ происходит такое преобразование формы входного сигнала, при котором в его спектре появляются новые гармонические составляющие. Вторую операцию осуществляет линейный фильтр, выделяя нужные спектральные составляющие преобразованного входного сигнала. Меняя определенным образом параметры входных сигналов и используя различные НЭ и электрические фильтры, можно осуществлять требуемую трансформацию спектра и выделять нужные составляющие. К такой удобной теоретической модели сводят многие схемы нелинейных усилителей, умножителей, модуляторов, детекторов, автогенераторов, импульсных схем, делителей и преобразователей частоты.

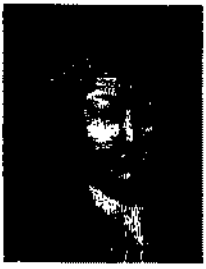
Как правило, нелинейные цепи характеризуются сложной зависимостью между входным сигналом $u_{вх}(t)$ и выходной реакцией $u_{вых}(t)$, которую в общем виде можно записать как

$$u_{вых}(t) = f[u_{вх}(t)].$$

В нелинейных цепях с безынерционными НЭ наиболее удобно в качестве воздействия рассматривать входное напряжение $u_{вх}(t)$, а в качестве отклика — выходной ток $i_{вых}(t)$, связь между которыми определяется следующей нелинейной функциональной зависимостью

$$i_{вых}(t) = f[u_{вх}(t)]. \qquad (5.1)$$

Данное соотношение аналитически может представлять собой обычную вольт-амперную характеристику НЭ. Такой характеристикой обладают и нелинейный двухполюсник (например, полупроводниковый диод) и нелинейный четырехполюсник (транзистор, операционный усилитель, цифровая микросхема), работающие в нелинейном режиме при различных амплитудах входного сигнала. Вольт-амперные характеристики (их получают экспериментально) большинства НЭ имеют сложный вид, поэтому представление их аналитическими выражениями является достаточно трудной задачей. Как правило, не имеет большого смысла проектирование систем анализа и обработки сигналов по высокоточным формулам, если снижение погрешности расчетов и соответствующее усложнение систем не дает ощутимого эффекта в повышении точности обработки данных. Во всех этих условиях возникает *задача аппроксимации* — представление



Б. Тейлор

исходных сложных функций $f(u)$ простыми и удобными для практического использования относительно простыми функциями $i(u)$ (или их набором) таким образом, чтобы отклонение $i(u)$ от $f(u)$ в области ее задания было наименьшим по определенному критерию приближения. Функции $i(u)$ называют *функциями аппроксимации*. Напомним (см. гл. 4), что нахождение аналитической функции по экспериментальной вольт-амперной характеристике нелинейного элемента называют *аппроксимацией*.

Достаточно точным и простым способом аппроксимации может служить представление нелинейной характеристики в виде таблицы. Способ особенно эффективен и удобен для анализа процессов в нелинейных цепях с помощью компьютера; значения функции и аргумента образуют в запоминающем устройстве двумерный массив чисел.

Если анализ цепи необходимо проводить не численными, а аналитическими методами, то желательно подобрать такую несложную аппроксимирующую функцию, которая с требуемой точностью отражала бы все важнейшие особенности реальной характеристики. В радиотехнике и теории передачи информации используют несколько способов аппроксимации характеристик НЭ — степенну́ю, показательную, кусочно-линейную (линейно-ломаную) и т. д. Наиболее распространены аппроксимация степенны́м полиномом и кусочно-линейная аппроксимация сложных функций.

### Аппроксимация степенны́м полиномом

Это вид аппроксимации особенно эффективен при малых амплитудах (как правило, доли вольт) входных сигналов в тех случаях, когда характеристика НЭ имеет вид гладкой кривой, т. е. функция, аналитически описывающая кривую, и ее производные непрерывны и не имеют скачков. Очень часто при аппроксимации кривых в качестве степенно́го полинома используют ряд Тейлора (Брук Тейлор — *Brook Taylor*, английский математик; 1685–1731)

$$i(u) = a_0 + a_1(u - U_0) + a_2(u - U_0)^2 + \ldots + a_n(u - U_0)^n, \qquad (5.2)$$

где $a_0$, $a_1$, $a_2$, ..., $a_n$ — постоянные коэффициенты; $U_0$ — значение напряжения $u$, относительно которого ведется разложение в ряд и называемое *рабочей точкой*. Отметим, что здесь и далее временно́й аргумент $t$ в функциях тока и напряжения для упрощения опущен.

Коэффициенты ряда (полинома) Тейлора определяют по формуле

$$a_n = \frac{1}{n!}\frac{d^n i}{du^n}\bigg|_{u=U_0}. \qquad (5.3)$$

Оптимальное число членов ряда берется в зависимости от требуемой точности аппроксимации. Чем больше выбрано членов ряда, тем аппроксимация точнее. Аппроксимацию характеристик обычно удается достаточно точно осущест-

вить полиномом не выше второй или третьей степени. Для отыскания неизвестных коэффициентов ряда (5.2) необходимо задать диапазон $U_1$, $U_2$ нескольких возможных значений напряжения $u(t)$ и положение рабочей точки $U_0$ в этом диапазоне. Если требуется определить $n$ коэффициентов ряда, то на заданной характеристике выбирают $n + 1$ точку с своими координатами $(I_n, U_n)$.

Для упрощения расчетов одну точку совмещают с рабочей точкой $U_0$, имеющей координаты $(i = I_0, u = U_0)$; еще две точки выбираются на границах диапазона $u = U_1$ и $u = U_2$. Остальные точки располагают произвольно, но с учетом важности аппроксимируемого участка ВАХ. Подставляя координаты выбранных точек в формулу (5.2), составляют систему из $n + 1$ уравнений, которая решается относительно неизвестных коэффициентов $a_n$ ряда Тейлора.

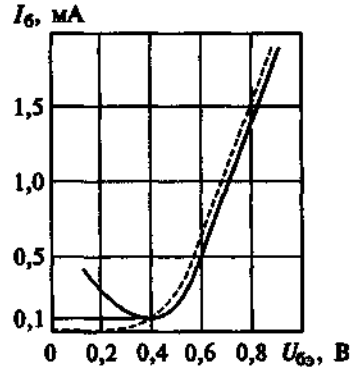



**Рис. 5.2.** Аппроксимация характеристики степенным полиномом

**Пример 5.1.** На рис. 5.2 штриховой линией представлена реальная входная характеристика $I_б = f(U_{бэ})$ биполярного транзистора КТ601А. Аппроксимировать данную характеристику в диапазоне напряжений 0,5...0,8 В полиномом Тейлора второй степени $i_б = a_0 + a_1(u_{бэ} - U_0) + a_2(u_{бэ} - U_0)^2$ относительно рабочей точки с постоянным значением напряжения $U_0 = 0{,}6$ В.

Р е ш е н и е. Для упрощения расчетов в качестве точек аппроксимации нелинейной характеристики транзистора зададимся значениями напряжений на границах диапазона напряжений и в рабочей точке, т. е. 0,4; 0,6 и 0,8 В. Поскольку выбранным точкам аппроксимации характеристики соответствуют токи 0,1; 0,5 и 1,5 мА, то для заданного полинома второй степени получим систему уравнений:

$$0{,}1 = a_0 + a_1(0{,}4 - 0{,}6) + a_2(0{,}4 - 0{,}6)^2 = a_0 - 0{,}2a_1 + 0{,}04a_2;$$

$$0{,}5 = a_0 + a_1(0{,}6 - 0{,}6) + a_2(0{,}6 - 0{,}6)^2 = a_0;$$

$$1{,}5 = a_0 + a_1(0{,}8 - 0{,}6) + a_2(0{,}8 - 0{,}6)^2 = a_0 + 0{,}2a_1 + 0{,}04a_2.$$

Решение полученной системы уравнений дает такие значения коэффициентов полинома: $a_0 = 0{,}5$ мА; $a_1 = 3{,}5$ мА/В; $a_2 = 7{,}5$ мА/В$^2$. Подставив их в формулу (5.2), находим аппроксимирующую функцию (ее график показан на рис. 5.2 сплошной линией)

$$i_б = 0{,}5 + 3{,}5(u_{бэ} - 0{,}6) + 7{,}5(u_{бэ} - 0{,}6)^2.$$

Аппроксимация функций рядом Тейлора имеет много недостатков. Ее применяют в основном, для непрерывных и гладких функций в локальных интервалах задания. Для разрывных и периодически повторяющихся функций использовать его практически невозможно, равно как и для непрерывных недифференцируемых функций. Операция дифференцирования также может быть не очень простой и точной, а получаемые ряды могут сходиться очень медленно.

### Кусочно-линейная аппроксимация

В большинстве практических случаев, когда на НЭ цепи воздействует сигнал значительной амплитуды, реальную вольт-амперную характеристику НЭ можно аппроксимировать кусочно-линейной линией, состоящей из нескольких отрезков с различными углами наклона к оси абсцисс. Аппроксимация связана с двумя параметрами НЭ — напряжением начала характеристики $E_н$ и ее крутизной $S$. Дифференциальная крутизна характеристики в рабочей точке определяется отношением приращения тока к приращению напряжения, и при малых их значениях имеем

$$S = \frac{\Delta i}{\Delta u} = \frac{di}{du}. \qquad (5.4)$$

Уравнение отрезка прямой при кусочно-линейной аппроксимации:

$$i = \begin{cases} 0, & u < E_н; \\ S(u - E_н), & u \geq E_н. \end{cases} \qquad (5.5)$$

Часто вольт-амперную характеристику НЭ, на который подан большой сигнал, удается с приемлемой точностью аппроксимировать двумя отрезками.

**Пример 5.2.** Экспериментальная входная характеристика $I_б = f(U_{бэ})$ транзистора КТ601А показана на рис. 5.3 штриховой линией. Выполнить кусочно-линейную аппроксимацию данной характеристики в окрестности рабочей точки $U_0 = 0{,}6$ В.

Р е ш е н и е. В соответствии с заданной вольт-амперной характеристикой биполярного транзистора находим, что значение тока в рабочей точке $I_0 = 0{,}5$ мА. Крутизну характеристики в рабочей точке вычислим приближенно по формуле (5.4).

Задав линейное приращение напряжения $\Delta u_{бэ} = 0{,}8 - 0{,}6 = 0{,}2$ В, находим приращение тока $\Delta i_б = 1{,}5 - 0{,}5 = 1$ мА. Тогда $S = \Delta i_б / \Delta u_{бэ} = 1/0{,}2 = 5$ мА/В.

В результате аппроксимации характеристики искомый ток базы транзистора в окрестности рабочей точки с координатами $I_0 = 0{,}5$ мА, $U_0 = 0{,}6$ В определится как $i_б = 0{,}5 + 5(u_{бэ} - 0{,}6) = 5(u_{бэ} - 0{,}5)$. Из этой формулы следует, что при $u_{бэ} < 0{,}5$ В ток базы должен принимать отрицательные значения, что не отражает заданная характеристика. Значит, полученная функция будет аппроксимировать заданную зависимость только при амплитуде напряжения $u_{бэ} \geq 0{,}5$ В. Если же напряжение $u_{бэ} < 0{,}5$ В, то можно принять $i_б = 0$. Итак, аппроксимирующая функция (сплошная линия на рис. 5.3), отражающая характеристику транзистора:

$$i = \begin{cases} 0, & u_{бэ} < 0{,}5; \\ 5(u_{бэ} - 0{,}5), & u_{бэ} \geq 0{,}5. \end{cases}$$

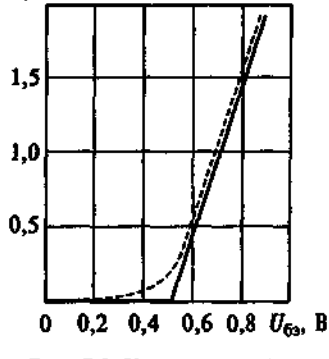



**Рис. 5.3.** Кусочно-линейная аппроксимация характеристики

Повышение точности аппроксимации нелинейных характеристик достигают увеличением числа отрезков прямых линий.

## 5.2. Отклик нелинейной цепи на гармонический входной сигнал

Как уже отмечалось, существенно упростить анализ процессов в нелинейной цепи удается при ее теоретическом представлении последовательно соединенными безынерционным НЭ и линейной цепью — фильтром. Проанализируем физические процессы, протекающие в нелинейной цепи (рис. 5.4, *а*), при воздействии на вход безынерционного НЭ *z* простейшего, гармонического сигнала $u_1(t) = U_m \cos\omega t$ и постоянного напряжения смещения $U_0$.

Используя характеристику безынерционного НЭ и проведя несложные графические построения, найдем аналитическую запись формы тока на выходе цепи в зависимости от фазового угла $\nu = \omega t$ (рис. 5.4, *б*, *в*). Вследствие нелинейности характеристики НЭ форма тока становится несинусоидальной. Причину искажения формы гармонического колебания нетрудно пояснить. Поскольку ток и напряжение связаны линейной зависимостью $\Delta i = S\Delta u$, а крутизна ВАХ на разных участках неодинаковая (имеет нелинейный характер), то равным приращениям напряжения $\Delta u$ отвечают неравные приращения тока $\Delta i$.

Так как функция тока, протекающего через НЭ, периодична (см. рис. 5.4, *в*), то ее можно представить (для удобства сделаем это во времени) тригонометрическим рядом Фурье:

$$i(t) = I_0 + \sum_{n=1}^{\infty} I_n \cos n\omega t, \tag{5.6}$$

где $I_0$, $I_n$ — амплитуды постоянной и гармонических составляющих.

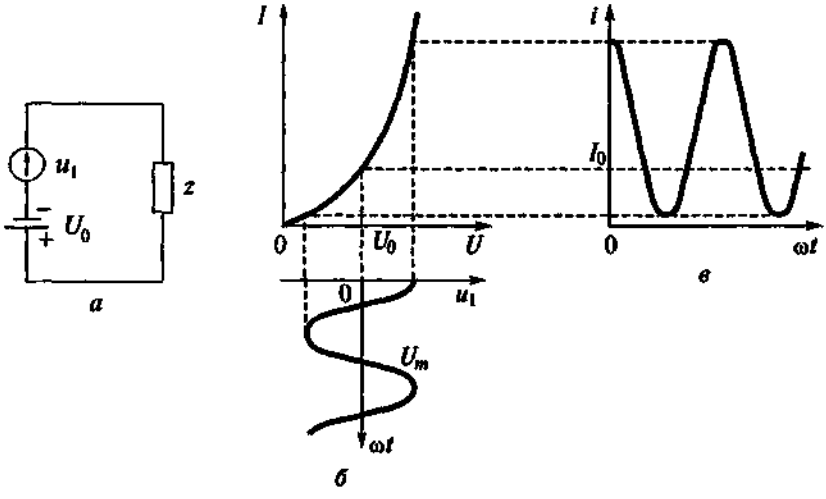



**Рис. 5.4.** Цепь с нелинейным элементом:
*а* — схема; *б*, *в* — графики процессов

**Спектр тока в цепи с НЭ при степенно́й аппроксимации его характеристики**

Пусть суммарное напряжение источников смещения и входного гармонического сигнала

$$u(t) = U_0 + U_m\cos\omega t \tag{5.7}$$

приложено к НЭ, вольт-амперная характеристика которого в окрестности рабочей точки $U_0$ аппроксимирована полиномом Тейлора вида

$$i(u) = a_0 + a_1(u - U_0) + a_2(u - U_0)^2 + a_3(u - U_0)^3 + \ldots \tag{5.8}$$

Подставив формулу (5.7) в выражение (5.8), получим

$$i(u) = a_0 + a_1U_m\cos\omega t + a_2U_m^2\cos^2\omega t + a_3U_m^3\cos^3\omega t + \ldots$$

Используя известные формулы разложения степеней косинусов:

$$\cos^2 x = \frac{1}{2}(1+\cos 2x);\quad \cos^3 x = \frac{1}{4}(3\cos x + \cos 3x);$$

$$\cos^4 x = \frac{1}{8}(3 + 4\cos 2x + \cos 4x)\ \text{и т. д.},$$

запишем общее выражение для тока нелинейной цепи, сгруппировав отдельно в суммы его постоянные составляющие и все члены с косинусами одинаковых аргументов

$$i(t) = \left(a_0 + \frac{1}{2}a_2U_m^2 + \frac{3}{8}a_4U_m^4 + \ldots\right) + \left(a_1U_m + \frac{3}{4}a_3U_m^3 + \frac{5}{8}a_5U_m^5 + \ldots\right)\cos\omega t +$$

$$+ \left(\frac{1}{2}a_2U_m^2 + \frac{1}{8}a_4U_m^4 + \ldots\right)\cos 2\omega t + \left(\frac{1}{4}a_3U_m^3 + \frac{5}{16}a_5U_m^5 + \ldots\right)\cos 3\omega t + \ldots. \tag{5.9}$$

Представленное в более компактной форме соотношение (5.9) примет вид

$$i(t) = I_0 + I_1\cos\omega t + I_2\cos 2\omega t + I_3\cos 3\omega t + \ldots . \tag{5.10}$$

Здесь постоянная составляющая и амплитуды гармоник тока:

$$I_0 = a_0 + \frac{1}{2}a_2U_m^2 + \frac{3}{8}a_4U_m^4 + \ldots;\qquad I_1 = a_1U_m + \frac{3}{4}a_3U_m^3 + \frac{5}{8}a_5U_m^5 + \ldots;$$

$$I_2 = \frac{1}{2}a_2U_m^2 + \frac{1}{8}a_4U_m^4 + \ldots;\qquad I_3 = \frac{1}{4}a_3U_m^3 + \frac{5}{16}a_5U_m^5 + \ldots \tag{5.11}$$

Анализ состава формул (5.11) показывает, что при степенно́й аппроксимации характеристики гармонический состав тока в цепи с НЭ существенно зависит от степени полинома. При этом постоянная составляющая и амплитуды четных гармоник определяются четными, а амплитуды нечетных гармоник — нечетными коэффициентами степенно́го полинома.

### Спектр тока в цепи с НЭ при кусочно-линейной аппроксимации

Пусть суммарное гармоническое и постоянное напряжение вида (5.7) подается на вход электрической цепи с НЭ, характеристика которого аппроксимирована кусочно-линейной линией и описывается формулой (5.5). В этом случае временна́я диаграмма тока, протекающего через НЭ цепи, имеет форму косинусоидальных импульсов с отсечкой их нижней части (рис. 5.5). Значение $\theta$ (в радианах или градусах), при котором ток изменяется от максимального $I_m$ до нуля, называют *углом отсечки*. При этом изменение фазы, соответствующее длительности полного импульса тока на выходе нелинейной цепи, равно $2\theta$. Из рис. 5.5 нетрудно определить, что при фазовом угле $\omega t = \theta$ ток в цепи $i(\omega t) = i(\theta) = 0$ и напряжение начала характеристики согласно (5.5) $E_н = U_0 + U_m\cos\theta$, откуда

$$\cos\theta = (E_н - U_0)/U_m. \tag{5.12}$$

Подставив в (5.5) суммарное напряжение источников сигнала и смещения из (5.7) и напряжение начала характеристики $E_н$, получим запись формы тока

$$i(\omega t) = SU_m(\cos\omega t - \cos\theta),\ -\theta \le \omega t \le \theta. \tag{5.13}$$

Полученную четную функцию $i(\omega t)$ периодической последовательности импульсов тока (5.13) можно разложить в тригонометрический ряд Фурье (2.16), в котором период повторения составляет $2\pi$, длительность импульса — $2\theta$, а текущей переменной является мгновенный фазовый угол $\nu = \omega t$. В этих импульсах тока постоянная составляющая запишется следующим образом:

$$I_0 = \frac{SU_m}{2\pi}\int_{-\theta}^{\theta}(\cos\omega t - \cos\theta)d\omega t = \frac{SU_m}{\pi}(\sin\theta - \theta\cos\theta). \tag{5.14}$$

Амплитуда первой гармоники

$$I_1 = \frac{SU_m}{2\pi}\int_{-\theta}^{\theta}(\cos\omega t - \cos\theta)\cos\omega t d\omega t = \frac{SU_m}{\pi}(\theta - \sin\theta\cos\theta). \tag{5.15}$$

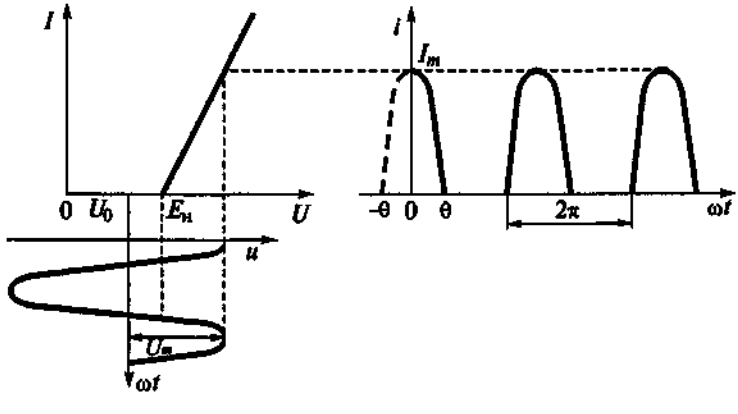



**Рис. 5.5.** Форма тока при кусочно-линейной аппроксимации характеристики НЭ

Аналогично определяют амплитуды составляющих $I_n$ и для $n = 2, 3, \ldots$ . При этом обобщенная формула для вычисления этих гармоник имеет вид

$$I_n = \frac{2SU_m}{\pi} \frac{\sin n\theta \cos\theta - n\cos n\theta \sin\theta}{n(n^2-1)}. \qquad (5.16)$$

В радиотехнике полученные результаты принято записывать в форме

$$I_0 = SU_m\gamma_0; \quad I_1 = SU_m\gamma_1; \quad \ldots; \quad I_n = SU_m\gamma_n. \qquad (5.17)$$

Здесь $\gamma_0$, $\gamma_1$, ..., $\gamma_n$ — *функции Берга* (Берг Аксель Иванович, русский ученый, инженер-адмирал, академик АН СССР; 1893–1979), или *коэффициенты гармоник*, отражающие значения присутствующих гармоник в спектре выходного тока НЭ, которые аналитически записывают следующим образом:

$$\gamma_0 = \frac{1}{\pi}(\sin\theta - \theta\cos\theta),$$

$$\gamma_1 = \frac{1}{\pi}(\theta - \sin\theta\cos\theta), \qquad (5.18)$$

$$\gamma_n = \frac{2}{\pi} \frac{\sin n\theta\cos\theta - n\cos n\theta\sin\theta}{n(n^2-1)}, \; n = 2,3,\ldots.$$

**Пример 5.3.** Характеристика НЭ представлена с помощью метода кусочно-линейной аппроксимации двумя отрезками прямых. Параметры аппроксимации: $E_н$ = 0,6 В, $S$ = 0,25 мА/В. На нелинейный элемент воздействует суммарное (постоянное и переменное) напряжение $u(t) = 0{,}2 + 0{,}8\cos\omega t$ В. Определить постоянную составляющую и амплитуду первой гармоники тока, протекающих через НЭ.

Р е ш е н и е. Воспользовавшись формулой (5.12), находим, что для данных параметров аппроксимации $\cos\theta = (0{,}6 - 0{,}2)/0{,}8 = 0{,}5$. Отсюда угол отсечки тока, протекающего через нелинейный элемент, $\theta = 60°$. Два первых коэффициента гармоник, соответствующие этому углу, будут: $\gamma_0 = 0{,}11$; $\gamma_1 = 0{,}2$. Подставив последовательно эти значения в соотношение (5.17), вычисляем соответственно амплитуды постоянной составляющей и первой гармоники, протекающих через НЭ цепи:

$$I_0 = 2 \text{ мА}, \; I_1 = 4 \text{ мА}.$$

Коэффициенты гармоник задают таблицами и графиками. На рис. 5.6 показаны графики первых трех из этих функций в зависимости от угла отсечки.

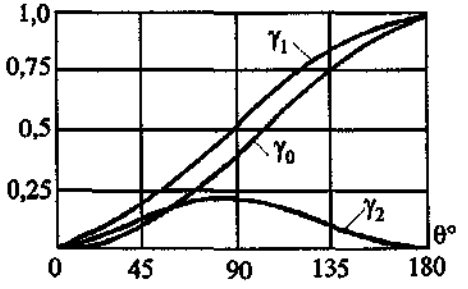



**Рис. 5.6.** Коэффициенты гармоник



А.И. Берг

## 5.3. Нелинейные усилители мощности и умножители частоты

Широкое применение в различных радиоэлектронных устройствах (особенно в передатчиках) находят резонансные усилители мощности и умножители частоты. Рассмотренный в 4.6 транзисторный резонансный усилитель работает в линейном режиме, и поэтому имеет КПД менее 50 %. В передающих же устройствах, особенно мощных, требуется высокий КПД. Эффективное средство повышения энергетических показателей резонансного усилителя мощности — использование заведомо нелинейного режима работы его активного элемента (режим с отсечкой тока). Естественно, необходимым условием работы подобных усилительных и умножительных схем и устройств является сохранение формы усиливаемого сигнала с возможно меньшими нелинейными искажениями.

### Нелинейный резонансный усилитель мощности

Проанализируем и оценим параметры упрощенной электрической схемы нелинейного резонансного усилителя мощности на биполярном транзисторе (рис. 5.7, *а*), к входу которого последовательно подключены источники гармонического напряжения $u_{вх}(t) = U_{mвх}\cos\omega_p t$ и постоянного напряжения смещения $U_0$, а резонансный контур нагрузки настроен на частоту усиливаемого сигнала $\omega_p$. Напряжение смещения $U_0$ на базе транзистора следует выбирать таким образом, чтобы в отсутствие переменного входного сигнала выходной ток транзистора был равен нулю.

Положим, что переменный коллекторный ток биполярного транзистора имеет форму косинусоидальных импульсов с отсечкой их нижней части. Временные диаграммы импульсов коллекторного тока $i_к(\omega t) = i_к(t)$, тока первой гармоники $i_1(t) = i_1(\omega t) = I_1\cos\omega_p t$ и выходного напряжения $u_{вых}(\omega t) = u_{вых}(t)$ показаны соответственно на рис. 5.7, *б*. Известно, что спектральный состав косинусоидальных импульсов коллекторного тока обычно содержит множество гармонических составляющих кратных частот, однако наибольшее значение имеет амплитуда первой гармоники. Это объясняется тем, что на резонансной частоте активное сопротивление параллельного контура максимально и на нем выделяется усиливаемое напряжение с частотой входного сигнала $\omega_p$. Очевидно, что сопротивление параллельного колебательного контура на частотах $2\omega_p$, $3\omega_p$, ... столь мало, что эти высшие гармонические составляющие практически не дают вклада в формирование выходного сигнала $u_{вых}(t)$.

Используя формулу (5.17), запишем соотношение для амплитуды выходного напряжения

$$U_{mвых} = I_1 R_0 = S U_{mвх} \gamma_1 R_0, \qquad (5.19)$$

где $R_0$ — резонансное сопротивление параллельного контура; $\gamma_1$ — функция Берга для первой гармоники.

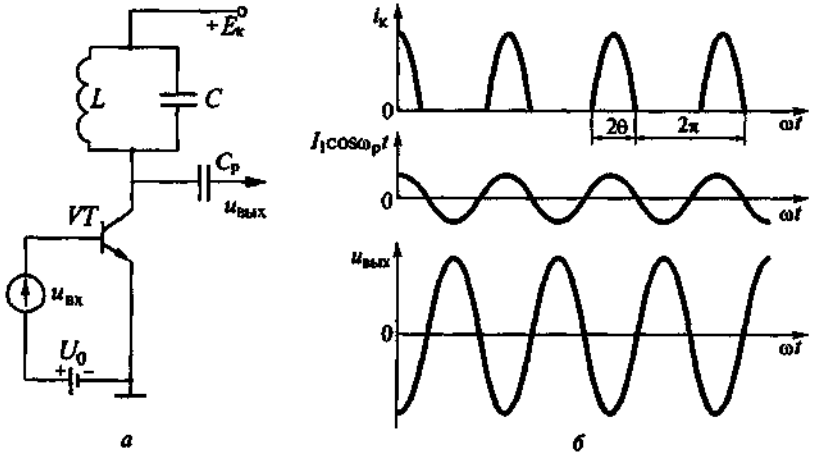



**Рис. 5.7.** Транзисторный резонансный усилитель:
*а* — схема; *б* — временные диаграммы токов и напряжений

### Умножители частоты

Необходимость в умножителях возникает при разработке высокостабильных источников гармонических колебаний повышенной частоты, когда непосредственное генерирование сигналов такого диапазона затруднительно. Умножитель частоты — устройство, повышающее частоту входного сигнала в $n$ раз, где $n$ — *коэффициент умножения* (целое число).

Наличие в спектре коллекторного тока гармонических составляющих с частотами, кратными входной частоте, позволяет использовать нелинейный резонансный усилитель в качестве умножителя частоты. Для этого достаточно в схеме резонансного усилителя (рис. 5.7, *а*) настроить параллельный колебательный контур на требуемую частоту. Известно, что при больших значениях умножения $n$ коэффициенты гармоник $\gamma_n$ довольно малы, поэтому важно выбрать такой угол отсечки $\theta$, при котором соответствующие коэффициенты гармоник максимальны.

В радиотехнике практически доказано, что оптимальный угол отсечки, дающий наибольшую амплитуду выходного напряжения в умножителях частоты, примерно равен $180°/n$.

Принципы действия умножителя частоты и нелинейного резонансного усилителя мощности в основном одинаковы, а их различия заключаются лишь в выборе угла отсечки тока. По аналогии с выражением (5.19) определим амплитуду выходного напряжения умножителя частоты при кусочно-линейной аппроксимации характеристики транзистора

$$U_{mn} = I_n R_{0n} = S U_{mx} \gamma_n R_{0n}, \qquad (5.20)$$

где $R_{0n}$ — резонансное сопротивление контура на $n$-й гармонике; $\gamma_n$ — функция Берга для $n$-й гармоники.

## 5.4. Модуляция колебаний

Модуляция является нелинейной операцией, ее осуществляют нелинейными устройствами — *модуляторами*. Упрощенно многочисленные типы модуляторов делят на амплитудные, частотные, фазовые, импульсные и цифровые.

### Амплитудные модуляторы

Формальный анализ аналитического выражения для однотонального АМ-сигнала $u_{АМ}(t) = U_н(1 + M\cos\omega t)\cos\omega_0 t$ показывает, что при его формировании необходимо перемножить модулирующий сигнал $e(t) = E_0\cos\Omega t$ и несущее колебание $u_н(t) = U_н\cos\omega_0 t$ (если же говорить строго, то требуется умножить сумму из двух элементарных слагаемых $1 + M\cos\Omega t$ на несущее колебание, однако это не имеет существенного значения для изучения подобных процессов). Но до настоящего времени в радиотехнике не разработано эффективных методов непосредственного перемножения двух или нескольких аналоговых сигналов. Поэтому при амплитудной модуляции применяют косвенные методы перемножения с помощью нелинейных или параметрических цепей.

**Амплитудные модуляторы на основе схем резонансных усилителей мощности.** При разработке амплитудных модуляторов часто используют эффект преобразования суммы модулирующего и несущего колебаний, подаваемых на безынерционный НЭ. Простейший амплитудный модулятор создают на основе нелинейного резонансного усилителя (рис. 5.8, *а*).

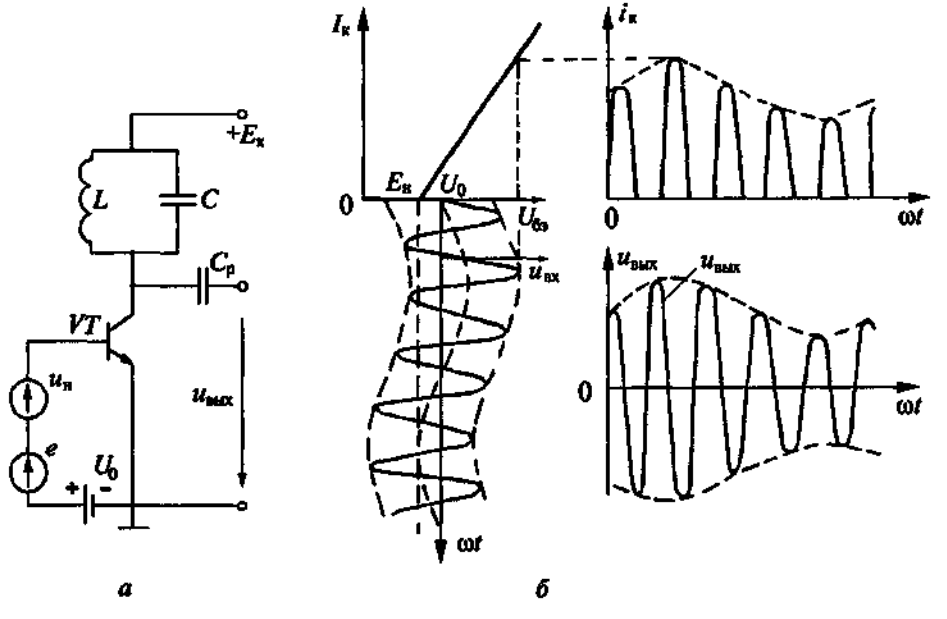



**Рис. 5.8.** Амплитудный модулятор:
*а* — упрощенная схема; *б* — диаграммы токов и напряжений

Для создания модулятора на входе транзистора *VT* включают последовательно источники напряжения смещения $U_0$, модулирующего сигнала $e(t)$, и генератор несущего колебания $u_н(t)$ и настривают колебательный контур на несущую частоту $\omega_0$.

Рассмотрим принцип получения однотонального АМ-сигнала на основе так называемого *базового модулятора* (рис. 5.8, *а*). В этом случае к входу модулятора необходимо приложить суммарное напряжение вида

$$u_{вх}(t) = U_0 + E_0\cos\Omega t + U_н\cos\omega_0 t. \qquad (5.21)$$

С помощью временны́х диаграмм токов и напряжений, представленных на рис. 5.8, *б*, и поясняется принцип действия модулятора. Положим, что *сквозная характеристика* транзистора — зависимость тока коллектора $I_к$ от напряжения база–эмиттер $U_{бэ}$ — аппроксимирована двумя отрезками прямых линий. Вследствие перемещения рабочей точки относительно напряжения смещения $U_0$ по закону модулирующего сигнала $e(t)$ происходит изменение угла отсечки тока в кривой несущего колебания. В результате импульсы коллекторного тока $i_к$ транзистора, отражающие изменение амплитуды несущего колебания, оказываются промодулированными по амплитуде. В спектре импульсов коллекторного тока транзистора модулятора содержится множество гармонических составляющих с частотами $\omega_0$ и $\Omega$, а также с кратными и *комбинационными* (суммарными и разностными составляющими гармоник $\omega_0$ и $\Omega$) частотами. Колебательный контур должен иметь полосу пропускания $\Delta\omega_{АМ} = 2\Omega$, и тогда он выделит из спектра импульсов коллекторного тока только гармоники с тремя частотами $\omega_0 - \Omega$, $\omega_0$ и $\omega_0 + \Omega$.

Для оценки качества работы модулятора, с точки зрения вносимых искажений, используют *статическую модуляционную характеристику* — зависимость амплитуды первой гармоники тока $I_{к1}$ транзистора от постоянного напряжения смещения на базе $U_{бэ}$ (рис. 5.9). Отметим, что для исключения нелинейных искажений необходимо использовать только линейный участок модуляционной характеристики в диапазоне токов $I_{к\,min} - I_{к\,max}$.

При многотональной амплитудной модуляции (модуляции реальным сигналом) приведенные выше рассуждения справедливы, и при выборе линейного участка модуляционной характеристики нелинейные искажения будут минимальны. Однако при модуляции многотональным сигналом могут возникнуть частотные искажения, обусловленные следующим — чем дальше отстоит боковая составляющая от несущей, тем меньше она усиливается вследствие резонансного характера АЧХ контура. Для снижения частотных искажений сигналов полосу пропускания резонансной нагрузки в схеме модулятора следует увеличивать, а с точки зрения фильтрации паразитных гармоник — уменьшать. На практике принимают компромиссное решение — выбирают полосу пропускания нагрузки, равную удвоенному значению высшей частоты модуляции.

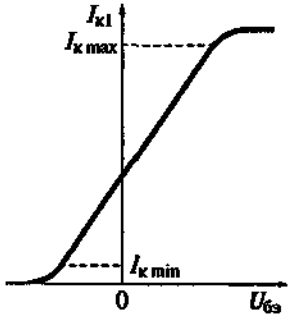



Рис. 5.9. Модуляционная характеристика

**Пример 5.4.** На базу транзистора амплитудного модулятора поступают гармонический модулирующий сигнал с амплитудой $E_0 = 0{,}05$ В, несущее колебание с амплитудой $U_н = U_m = 0{,}1$ В и постоянное напряжение смещения $U_0 = 0{,}6$ В. Сквозная характеристика транзистора аппроксимирована двумя отрезками прямых линий (см. рис. 5.8, *б*) с напряжением $E_к = 0{,}6$ В. Определить коэффициент амплитудной модуляции.

Р е ш е н и е. Используя исходные данные, находим, что рабочая точка перемещается по характеристике от максимального напряжения $U_{max} = U_0 + E_0 = 0{,}65$ В до минимального $U_{min} = U_0 - E_0 = 0{,}55$ В.

Подставляя эти величины вместо $U_0$ в (5.12) и проводя несложные тригонометрические преобразования, вычислим предельные значения угла отсечки:

$$\theta_{max} = \arccos\frac{E_к - U_{max}}{U_m} = 2{,}1 \text{ рад}, \quad \theta_{min} = \arccos\frac{E_к - U_{max}}{U_m} = 1{,}05 \text{ рад}.$$

Амплитуда первой гармоники коллекторного тока пропорциональна функции Берга $\gamma_1(\theta,?)$ максимальное и минимальное значения которой, согласно (5.18), равны: $\gamma_1(\theta_{max}) = 0{,}805$, $\gamma_1(\theta_{min}) = 0{,}196$.

Воспользовавшись выражением (2.78), определим коэффициент модуляции

$$M = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} = \frac{SU_m(\gamma_{max} - \gamma_{min})}{SU_m(\gamma_{max} + \gamma_{min})} = 0{,}6.$$

В последнее время в амплитудных модуляторах широко применяются аналоговые интегральные микросхемы, совмещающие в своем составе элементы, выполняющие ряд специфических функций.

**Амплитудные модуляторы на интегральных перемножителях напряжений.** С помощью интегрального перемножителя напряжений реализуют передаточную функцию $u_{вых} = k_a u_1 u_2$, где $k_a$ — масштабный коэффициент, а $u_1$ и $u_2$ — перемножаемые аналоговые напряжения. Обратимся к упрощенной структурной схеме интегрального перемножителя (рис. 5.10), в котором используют *квадратурный* метод перемножения двух непрерывных напряжений.

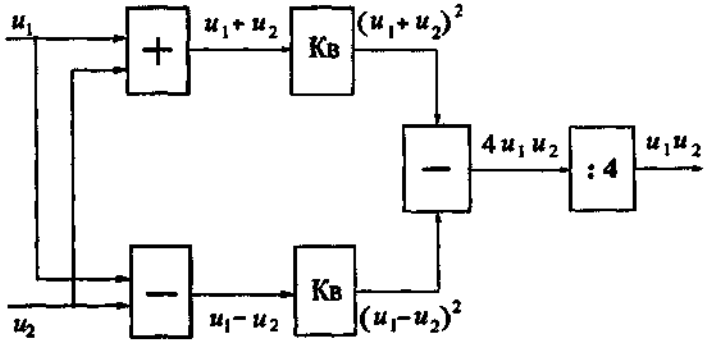



**Рис. 5.10.** Структурная схема интегрального перемножителя напряжений

В схеме обозначено: «+» — сумматор; «–» — вычитающее устройство; «Кв» — устройство возведения в квадрат; «:4» — делитель напряжения на 4 (этот элемент необязателен). При перемножении напряжений производятся:

- суммирование $u_1 + u_2$;
- вычитание $u_1 - u_2$;
- возведение в квадрат $(u_1 + u_2)^2$, $(u_1 - u_2)^2$;
- вычитание квадратов $(u_1 + u_2)^2 - (u_1 - u_2)^2 = 4u_1u_2$;
- деление напряжения на четыре $4u_1u_2/4 = u_1u_2$.

В основу аналогового перемножителя заложены идентичные, со стабильными параметрами НЭ, имеющие квадратичные характеристики. Наиболее просто это достигается в интегральных перемножителях напряжений. На входах перемножителя напряжений используют дифференциальные усилители, поэтому его входные сопротивления достаточно велики.

Интегральные перемножители аналоговых напряжений наиболее широко применяют в передающих устройствах в качестве *балансных амплитудных модуляторов* для получения *АМ-сигналов с подавленной несущей*. Подобные модуляторы осуществляют прямое перемножение модулирующего сигнала и несущего колебания. Действительно, если на входы перемножителя подаются модулирующий сигнал $e(t) = E_0\cos\Omega t$ и несущее колебание $u_н(t) = U_н\cos\omega_0 t$, то нетрудно заметить, что выходное напряжение

$$u_{БМ}(t) = k_a E_0\cos\Omega t U_н\cos\omega_0 t = 0{,}5k_a E_0 U_н[\cos(\omega_0 - \Omega)t + \cos(\omega_0 + \Omega)t] \qquad (5.22)$$

состоит из нижней и верхней боковых высокочастотных составляющих АМ-сигнала, т. е. является балансно-модулированным колебанием, в котором отсутствует несущая компонента.

### Угловая модуляция

На практике радиосигналы с угловой модуляцией получают либо непосредственной перестройкой частоты задающего генератора (*частотная модуляция*), либо изменением фазы несущего колебания (*фазовая модуляция*).

**Частотные модуляторы.** Наиболее просто частотную модуляцию несущего колебания можно осуществить путем электронной (как правило, мгновенной) перестройки резонансной частоты колебательного контура автогенератора. В практических радиоэлектронных схемах это выполняется с помощью нелинейного полупроводникового элемента — *варикапа* (рис. 5.11). Из теории полупроводниковых приборов известно, что барьерная емкость $C$ $p$-$n$-перехода варикапа существенно зависит от приложенного напряжения (рис. 5.11, *а*) и определяется *вольт-фарадной характеристикой* $C(u)$.

Как уже отмечалось в 2.4, для реализации частотной модуляции необходимо по закону модулирующего сигнала изменять частоту несущего колебания. На рис. 5.11, *б* в схеме частотного модулятора штриховой линией обозначен *автогенератор* на ОУ (автогенераторы более подробно рассматриваются в гл. 6), вырабатывающий в отсутствие модулирующего сигнала гармоническое несущее

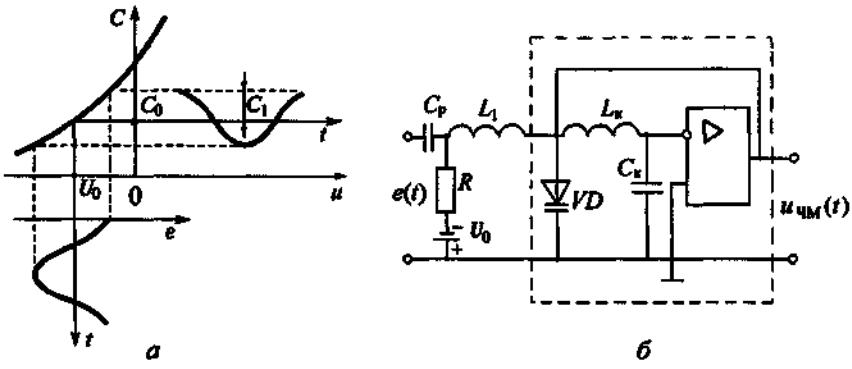



**Рис. 5.11.** Частотный модулятор с варикапом:
*а* — вольт-фарадная характеристика варикапа; *б* — схема модулятора

колебание $u_н(t) = U_н\cos\omega_0 t$. В этой схеме частотного модулятора индуктивность $L_к$, емкость $C_к$ и варикап $VD$ образуют колебательный контур, резонансная частота которого равна несущей частоте. Перестройка частоты генерируемых колебаний достигается в модуляторе путем изменения емкости варикапа.

При отключенном модулирующем сигнале емкость варикапа определяется постоянным напряжением смещения $U_0$ и равна $C_0$. Если же на входе автогенератора действует гармонический сигнал $e(t) = E_0\cos\Omega t$, то емкость варикапа $C(t)$ будет изменяться во времени относительно $C_0$ почти по гармоническому закону (см. рис. 5.11, *а*). По такому же закону начнет перестраиваться и резонансная частота колебательного контура (ведь $\omega_р = 1/\sqrt{LC}$) и, соответственно, частота выходного сигнала автогенератора, если ее девиация невелика.

Во избежание шунтирования контура малыми сопротивлениями источников $U_0$ и $e(t)$ их подключение осуществляется через дроссель $L_1$, представляющий большое сопротивление для несущего колебания и малое сопротивление для модулирующего сигнала и постоянного напряжения смещения. В схеме модулятора постоянная емкость $C_р$ — разделительный конденсатор.

**Цифровой частотный модулятор.** Напомним, что цифровую частотную модуляцию (*ЦЧМ*, так называемую *манипуляцию с минимальным сдвигом — ММС*; иначе *цифровую частотную модуляцию с непрерывной фазой — ЦЧМНФ*) можно реализовать различными способами, например управлением частотой генератора гармонических сигналов по закону, аналогичному (2.199):

$$u_{\text{ЦЧМНФ}}(t) = u_{\text{ММС}}(t) = U_н\cos\left[\omega_0 t + k_{\text{ЧМ}}\sum_{n=0}^{\infty} e_n^{(k)} v(t - n\Delta t) + \varphi_0\right], \qquad (5.23)$$

где $k_{\text{ЧМ}}$ — девиация частоты; $e_n^{(k)}$ — кодовые символы; $v(t - n\Delta t)$ — форма импульсного сигнала; $\varphi_0$ — начальная фаза несущего колебания.

Структурная схема модулятора с ММС показана на рис. 5.12. Собственно модулятор построен по квадратурной схеме, содержащей *задающий генератор* (*несущей*) *частоты* $f_0$, *фазовращатель* на 90°, два *умножителя* и *сумматор*.

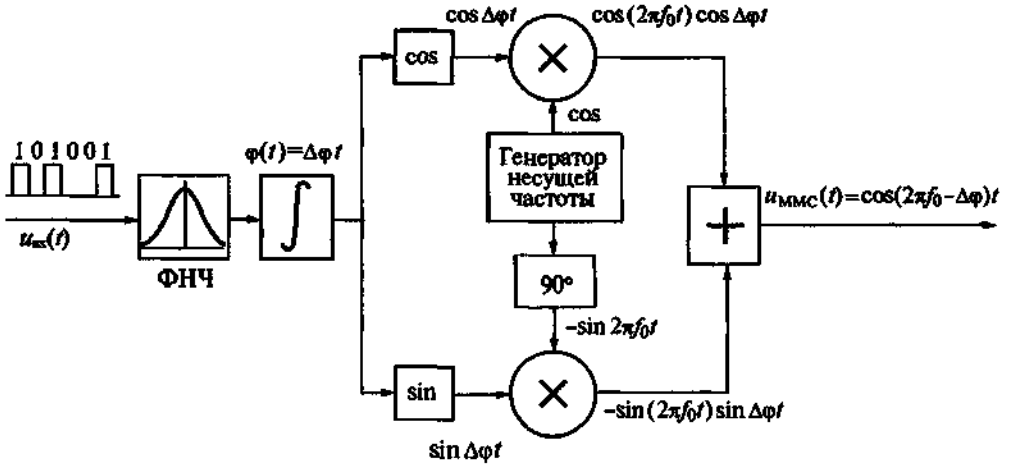



**Рис. 5.12.** Квадратурная схема формирования двоичного сигнала ММС

Для сглаживания закона изменения фазы перед модулятором включен ФНЧ с гауссовской АЧХ и интегратор. Применение ФНЧ Гаусса позволяет при дискретном изменении частоты получать *«сглаженные переходы»* в текущей фазе несущего колебания. В системах связи вид модуляции с использованием ФНЧ Гаусса получил название *гауссовской манипуляции с минимальным* частотным *сдвигом* (*ГММС*; *Gaussian Minimum Shift Keying — GMSK*). Диаграммы, поясняющие структуру *GMSK*-сигнала и принцип его формирования, показаны на рис. 5.13. Сигнальные биты $u_{мс}(t) = u_c$ представлены либо импульсом положительной (1), либо отрицательной полярности (0). Входной сигнал, преобразованный в ФНЧ и интеграторе в фазовый сдвиг $\varphi(t) = \Delta\varphi t$ (рис. 5.13, *б*), разбивают в блоках cos и sin на две квадратурные составляющие $\cos\varphi(t)$ и $\sin\varphi(t)$.

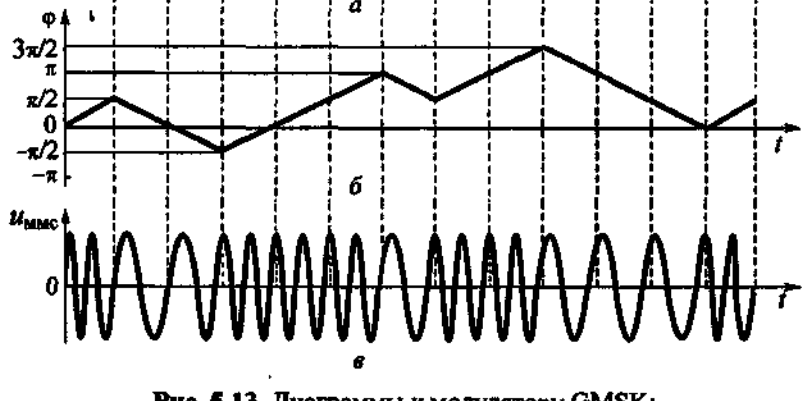



**Рис. 5.13.** Диаграммы к модулятору GMSK:
*а* — сигнальные биты; *б* — закон изменения фазы; *в* — закон изменения несущего

Эти составляющие поступают на блоки умножителей (×), на первый из которых от задающего генератора подают колебание несущей частоты $\cos(2\pi f_0 t)$, а на другой — ортогональное (квадратурное) колебание несущей частоты — $\sin(2\pi f_0 t)$. Сигналы с умножителей поступают в сумматор, на выходе которого формируется *GMSK*-сигнал $u_{\text{ММС}}(t) = u_{\text{ЧМНФ}}(t)$ с девиацией частоты, когерентной (от лат. *cohaerens* — находящийся в связи) скорости манипуляции (рис. 5.13, *в*).

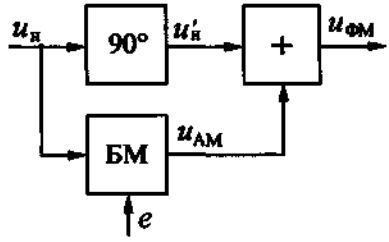



**Рис. 5.14.** Модулятор Армстронга

Цифровая модуляция GMSK обладает эффективными свойствами: достаточно узкой рабочей полосой частот; постоянной амплитудой радиосигнала, позволяющей использовать усилители мощности передатчиков в почти ключевом режиме (КПД до 90%); хорошей помехоустойчивостью каналов радиосвязи.

**Фазовые модуляторы. Метод Армстронга.** Для фазовой модуляции при небольших индексах Э. Армстронг в 1932 г. предложил складывать под углом 90° несущее и балансно-модулированное колебания (рис. 5.14). Фазовращатель (обозначен как 90°) сдвигает фазу несущего на 90°. На выходе *балансного модулятора* (БМ) получают АМ-колебание с подавленной несущей. На вход сумматора «+» поступают два колебания: напряжение несущей $u'_н(t) = U_н \sin\omega_0 t$ и АМ-сигнал $u_{АМ}(t) = MU_н e(t)\cos\omega_0 t$. Путем несложных математических выкладок можно показать, что выходной сигнал

$$u_{ФМ}(t) = u'_н(t) + u_{АМ}(t) = U_н\sqrt{1 + Me^2(t)}\,\cos\{\omega_0 t + \operatorname{arctg}[Me(t)]\} \qquad (5.24)$$

представляет гармоническое колебание с фазой, изменяющейся по закону модулирующего сигнала $e(t)$.

Метод Армстронга имеет здесь скорее познавательное значение. Практическая его реализация обычно затруднена, поскольку из-за малого индекса модуляции он работает на низкой частоте и требует большого числа умножителей.

**Пример 5.5.** Для аналоговой радиовещательной сети требуется передать ЧМ-сигнал со спектром звуковых частот от $f_н$ = 20 Гц до $f_в$ = 20 кГц при девиации частоты $f_д$ = 50 кГц. Определить число каскадов умножения частоты.

Р е ш е н и е. Для нижней граничной частоты спектра передаваемого сигнала индекс частотной модуляции $m = f_д/f_н = 5 \cdot 10^4/20 = 2{,}5 \cdot 10^3$. В модуляторе Армстронга считают допустимым небольшой индекс модуляции, например $m$ = 0,25. Тогда для нижней граничной частоты спектра ФМ-сигнала необходимо увеличить индекс модуляции в $2{,}5 \cdot 10^3/0{,}25 = 10\,000$ раз. Поскольку один умножитель повышает частоту не более чем в 4 – 5 раз, то для умножения в 10 000 раз требуется около шести каскадов (например, с коэффициентами умножения $5 \cdot 5 \cdot 5 \cdot 5 \cdot 4 \cdot 4$).

Рассмотренная в примере 5.5 система, в которой используются несколько умножителей частоты, громоздка. Поэтому в передатчиках с фазовой модуляцией применяют модулятор с варикапом в контуре автогенератора и схемой ФАПЧ.

## 5.5. Детектирование колебаний

*Детектированием* (демодуляцией) называют процесс преобразования модулированного высокочастотного сигнала в колебание, форма которого воспроизводит низкочастотный модулирующий сигнал. Детекторы (демодуляторы) выполняют функцию, обратную функции, осуществляемой модуляторами, и делятся на амплитудные, частотные, фазовые, импульсные, цифровые и т. д. Задача детектора — максимально безошибочно распознать принятый сигнал, насколько это возможно при ухудшении качества сигнала в процессе передачи. Существует две причины роста вероятности ошибки. Первая — последствия фильтрации в передатчике, канале и приемнике. Вторая причина роста вероятности ошибки — электрические помехи, порождаемые различными источниками, такими как галактика и атмосфера, импульсные помехи, комбинационные помехи, а также интерференция с сигналами от других источников.

Амплитудные и частотные детекторы являются нелинейными устройствами, в которых происходит взаимодействие полезного сигнала и помех. В частности, сильный сигнал подавляет в детекторе слабую помеху. Возможна и другая ситуация — сильная помеха может практически полностью подавить слабый сигнал. Эти явления существенным образом сказываются на отношении сигнал/шум. Существуют и такие помехи, устранить которые нельзя: это тепловой шум, который аддитивно накладывается на сигнал. Теоретически тепловой шум может быть бесконечно большим, но на практике очень большие амплитуды шума крайне редки. Поскольку тепловой шум присутствует во всех системах связи и для многих систем является доминирующим источником помех, характеристики теплового шума часто используются для моделирования шума при детектировании, а также проектировании приемников.

### Амплитудные детекторы

Рассмотрим процесс детектирования простейшего, однотонального АМ-сигнала. На вход детектора АМ-сигнала (*АМ-детектора*) подают высокочастотное модулированное колебание

$$u_{вх}(t) = U_{вх}(1 + M\cos\Omega t)\cos\omega_0 t = U_{вх}(t)\cos\omega_0 t, \qquad (5.25)$$

где $U_{вх}(t) = U_{вх}(1 + M\cos\Omega t)$.

Выходное же напряжение АМ-детектора должно быть низкочастотным $u_{вых}(t) = U_{вых}\cos\Omega t$, пропорциональным передаваемому сигналу. Эффективность работы АМ-детектора оценивают *коэффициентом передачи* (*коэффициентом детектирования*), аналитически представляющим отношение амплитуды выходного напряжения к амплитуде огибающей входного АМ-сигнала:

$$k_д = U_{вых}/(MU_{вх}). \qquad (5.26)$$

В зависимости от амплитуды АМ-сигнала и степени нелинейности характеристики детекторного элемента возможны два режима детектирования: *линей-*

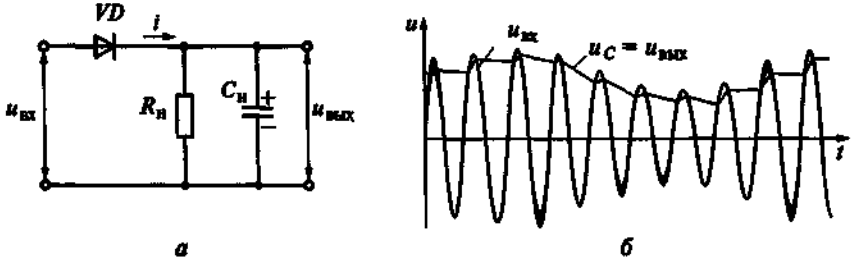



**Рис. 5.15.** Последовательный диодный детектор:
*а* — схема; *б* — диаграммы напряжений

*ный* (режим больших амплитуд с кусочно-линейной аппроксимацией характеристики) и *квадратичный* (работа при малых амплитудах на участке характеристики, описываемой полиномом второй степени).

**Линейный диодный детектор.** При линейном режиме работы детектора (рис. 5.15) амплитуды сигналов на входе и выходе связаны прямо пропорциональной зависимостью. На рис. 5.15, *а* представлена схема так называемого *последовательного диодного детектора*, у которого диод *VD* включен последовательно с низкочастотным $R_нC_н$-фильтром. Чтобы цепь реальной нагрузки любого детектора эффективно отфильтровывала полезный модулирующий сигнал и подавляла паразитные высокочастотные составляющие, необходимо выполнение двух неравенств:

$$1/(\Omega C_н) >> R_н; \qquad 1/(\omega_0 C_н) << R_н. \tag{5.27}$$

Еще одно непременное условие хорошей работы детектора — сопротивление нагрузки $R_н$ должно быть значительно больше сопротивления диода в его прямой проводимости, что всегда выполняется на практике.

Пусть на вход диодного детектора поступает однотональный АМ-сигнал $u_{вх}(t) = U_{вх}(t)\sin\omega_0 t$ (рис. 5.15, *б*). Ток через диод протекает в моменты времени, когда амплитуда входного напряжения $u_{вх}$ превышает напряжение на конденсаторе $C_н$ (а значит, и на выходе детектора $u_{вых}$). Конденсатор $C_н$ заряжается через малое сопротивление открытого диода намного быстрее, чем разряжается на высокоомное сопротивление нагрузки $R_н$. Поэтому диод большую часть периода входного колебания закрыт и амплитуда выходного напряжения близка к амплитуде входного. Для упрощения анализа и расчетов схемы положим, что на вход детектора подается достаточно большое гармоническое напряжение, при котором ВАХ диода можно аппроксимировать отрезками двух прямых линий (рис. 5.16, *а*). Как следует из рис. 5.16, *б*, амплитуды входного и выходного напряжений связаны простым соотношением: $U_{вых} = U_0 = U_{вх}\cos\theta$. В этом случае

$$k_д = U_{вых}/U_{вх} = \cos\theta. \tag{5.28}$$

Постоянная составляющая тока амплитудного детектора в соответствии с (5.17): $I_0 = SU_{вх}\gamma_0$. Поэтому среднее значение выходного напряжения

$$U_{вых} = U_0 = I_0 R_н = SU_{вх}\gamma_0 R_н. \tag{5.29}$$

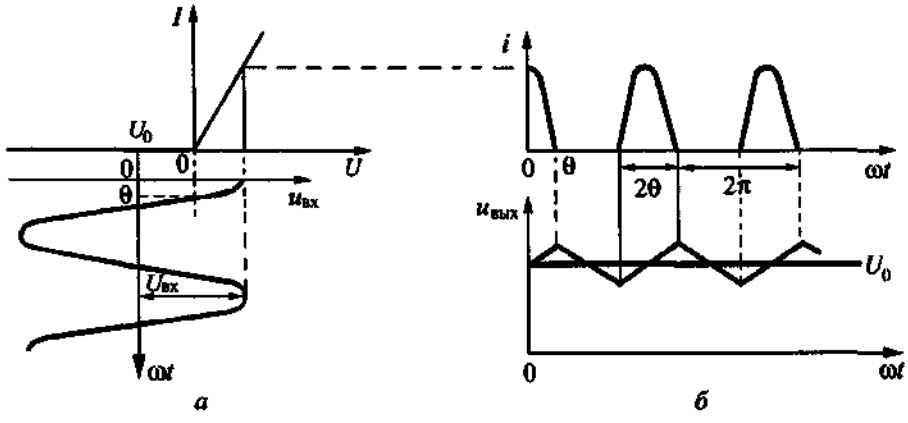



**Рис. 5.16.** Диаграммы к линейному детектору:
*а* — аппроксимация ВАХ; *б* — диаграммы выпрямленных тока и напряжения

Подставив в это соотношение $\cos\theta$ из (5.28) и $\gamma_0$ из (5.18), получим трансцендентное уравнение (напомним, что в них неизвестное входит в аргумент):

$$\cos\theta = S\gamma_0 R_н = \frac{SR_н}{\pi}(\sin\theta - \theta\cos\theta). \qquad (5.30)$$

Поделив обе части этого уравнения на $\cos\theta$, запишем

$$\operatorname{tg}\theta - \theta = \pi/(SR_н). \qquad (5.31)$$

Из этой формулы следует, что угол отсечки не зависит от амплитуды входного сигнала и определяется только произведением $SR_н$, причем чем оно больше, тем меньше угол отсечки. Как правило, $SR_н >> 1$, поэтому угол отсечки $\theta$ близок к нулю. Из математики известно, что при малых параметрах (в данном случае углах) $\theta$ имеет место равенство $\operatorname{tg}\theta = \theta + \theta^3/3$. Приняв во внимание это равенство, из соотношения (5.31) нетрудно вывести следующую формулу для расчета коэффициента детектирования

$$k_д = \cos\theta = \cos\sqrt[3]{3\pi/(SR_н)}. \qquad (5.32)$$

**Входное сопротивление последовательного детектора.** В схемах детекторов высокочастотное модулированное колебание, подаваемое на АМ-детектор, снимают с резонансного усилителя промежуточной частоты приемника. При этом контур УПЧ шунтируется низким входным сопротивлением детектора, что приводит к уменьшению его добротности и расширению полосы пропускания усилителя (ухудшению избирательности). Определим входное сопротивление последовательного детектора АМ-сигнала. Для этого положим, что угол отсечки очень мал, и поэтому $\cos\theta = 1$ и $U_{вх} = U_0$. Мощность немодулированного гармонического сигнала, подводимую к детектору, определяют как $P_{вх} = 0{,}5U_{вх}^2/R$, где $R_{вх}$ — его искомое входное сопротивление. Мощность, выделяемая на нагрузке детектора, $P_н = U_0^2/R_н$.

Поскольку сопротивление диода в прямом направлении близко к нулю, то входная мощность выделяется на нагрузке ($P_н = P_{вх}$), и поэтому $U_0^2/R_н = 0{,}5\,U_{вх}^2/R_{вх}$. Отсюда входное сопротивление диодного детектора

$$R_{вх} = R_н/2. \tag{5.33}$$

Принцип действия последовательного линейного детектора можно распространить на любые типы амплитудных детекторов, имеющих НЭ с односторонней (вентильной) проводимостью.

**Пример 5.6.** АМ-сигнал с тестовой однотональной модуляцией, имеющий вид $u_{АМ}(t) = 5(1 + 0{,}8\cos\Omega t)\cos\omega_0 t$, подан на вход линейного диодного детектора. Активное сопротивление нагрузки детектора $R_н = 20$ кОм, а крутизна характеристики диода $S = 15$ мА/В. Определить амплитуду выходного сигнала детектора.

Р е ш е н и е. Поскольку значение параметра $SR_н = 300$ достаточно велико, то в соответствии с формулой (5.32), коэффициент детектирования

$$k_д = U_{вых}/U_{вх} = \cos\sqrt{3 \cdot 3{,}14/300} = 0{,}95.$$

Из записи входного АМ-сигнала находим его амплитуду $U_{вх} = 5(1 + 0{,}8) = 9$ В. Тогда амплитуда выходного сигнала $U_{вых} = k_д U_{вх} = 0{,}95 \cdot 9 = 8{,}55$ В.

**Квадратичный детектор.** Принципы действия квадратичного и линейного детекторов существенно отличаются, и это легко показать. Известно, что при малых амплитудах АМ-сигнала характеристику НЭ наиболее точно аппроксимируют полиномом Тейлора второй степени

$$i(t) = a_0 + a_1 u_{вх}(t) + a_2 u_{вх}^2(t). \tag{5.34}$$

Подставляя в это выражение значение напряжения $u_{вх}(t)$ из формулы (5.25), после несложных тригонометрических преобразований получим

$$i(t) = a_0 + a_1 U_{вх}(t)\cos\omega_0 t + 0{,}5a_2 U_{вх}^2(t) + 0{,}5a_2 U_{вх}^2(t)\cos 2\omega_0 t. \tag{5.35}$$

На выходе АМ-детектора переменные составляющие тока $i(t)$ с высокими частотами $\omega_0$ и $2\omega_0$ отфильтровывает $R_н C_н$-цепь. Передаваемая же информация содержится в низкочастотной составляющей продетектированного колебания, которая, после несложных математических выкладок, имеет следующий вид

$$i_н(t) = 0{,}5a_2 U_{вх}^2(1 + 0{,}5M^2 + 2M\cos\Omega t + 0{,}5M^2\cos 2\Omega t). \tag{5.36}$$

Как следует из формулы (5.36), полезный эффект детектирования (передаваемое сообщение) пропорционален квадрату входной амплитуды АМ-сигнала $U_{вх}^2$, поэтому такое детектирование и называют квадратичным.

Постоянная составляющая тока легко отфильтровывается (например, разделительным конденсатором), и выходной ток детектора определяется как

$$i_{вых}(t) = a_2 U_{вх}^2 M\cos\Omega t + 0{,}25a_2 M^2 U_{вх}^2 \cos 2\Omega t. \tag{5.37}$$

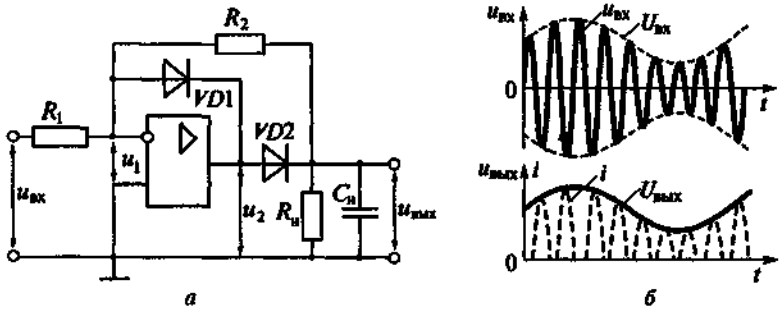



**Рис. 5.17.** Детектор на ОУ:
*а* — схема; *б* — диаграммы напряжений на входе и выходе

Полезным в (5.37) является только первое слагаемое, второе же определяет нелинейные искажения передаваемого сигнала, вносимые детектором. Снизить нелинейные искажения можно путем уменьшения глубины модуляции.

**Интегральные амплитудные детекторы.** Диодные (как и транзисторные) амплитудные детекторы при малых входных напряжениях вносят в выходной полезный сигнал значительные нелинейные искажения. Поэтому в последние годы в радиоприемниках в основном применяют детекторы, построенные на интегральных микросхемах — ОУ, которые осуществляют одновременно и усиление выходного сигнала и ряд других функций (рис. 5.17, *а*). Поскольку детектор выполнен по схеме инвертирующего усилителя (возможно и неинвертирующее включение ОУ), то при поступлении положительных полуволн гармонического сигнала (рис. 5.17, *б*) напряжение $u_2$ (аргумент $t$ для упрощения опущен) на выходе ОУ будет отрицательным. При этом диод *VD*1 открыт, а диод *VD*2 закрыт. Выход ОУ через малое прямое сопротивление диода *VD*1 оказывается подключенным к его входу, что создает глубокую отрицательную ОС по напряжению. В результате напряжение на выходе ОУ будет равно напряжению на его входе и близко к нулю, т. е. $u_2 \approx u_1 \approx 0$. Выходное напряжение детектора тоже будет равно нулю. Во время поступления отрицательной полуволны входного гармонического сигнала напряжение $u_2$ на выходе ОУ будет положительным, поэтому диод *VD*1 будет закрыт, а диод *VD*2 открыт и через него протекает ток $i$. При этом напряжение на выходе ОУ и выходное напряжение детектора (рис. 5.17, *б*)

$$u_{вых} = u_2 = -u_{вх}R_2/R_1 \qquad (5.38)$$

и соответствуют закону изменения входного сигнала.

### Детектирование сигналов с угловой модуляцией

При детектировании радиосигналов с угловой модуляцией их предварительно преобразуют в колебания с неглубокой амплитудной модуляцией и затем детектируют амплитудным детектором. Такое преобразование необходимо потому, что нелинейные элементы реагируют на изменения только амплитуды, а не частоты и фазы колебаний.

**Частотный детектор.** Для выделения передаваемого сигнала из частотно-модулированного колебания применяют частотные детекторы. Преобразование частотной или фазовой модуляции в амплитудную осуществляют с помощью линейных цепей, в частности резонансного контура, амплитуда напряжения на котором зависит от частоты входных колебаний. Положим, что контур настроен на частоту $\omega_р$ и на него подают однотональный ЧМ-сигнал с постоянной амплитудой и меняющейся по гармоническому закону частотой $\omega(t) = \omega_0 + \omega_д \cos\Omega t$ (здесь $\omega_0$ — частота несущего; $\omega_д$ — девиация частоты). Поскольку модуль полного сопротивления контура зависит от частоты, то амплитуда напряжения на нем будет изменяться во времени при отклонениях частоты ЧМ-сигнала от несущей $\omega_0$. Это положение поясняет рис. 5.18, где соответственно показаны: зависимость амплитуды напряжения на контуре от частоты $U_к(\omega)$; изменения во времени частоты $\omega(t)$ ЧМ-сигнала и амплитуды $U_к(t)$ колебания на контуре.

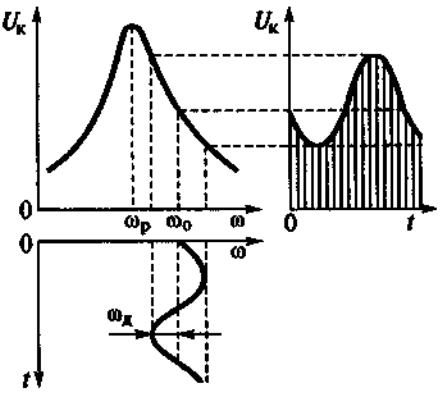



**Рис. 5.18.** Преобразование ЧМ-сигнала в АМ-ЧМ-колебание

Итак, амплитуда ЧМ-колебания на выходе резонансного контура изменяется во времени пропорционально гармоническому модулирующему сигналу, т. е. ЧМ-сигнал преобразуется в напряжение, модулированное еще и по амплитуде. Затем такое, достаточно сложное по структуре АМ-ЧМ-колебание, детектируется амплитудным детектором. Недостаток этого метода детектирования — ограниченный линейный участок на скате резонансной кривой контура.

**Частотный дискриминатор.** На рис. 5.19 представлен ЧМ-детектор (частотный дискриминатор) и временны́е диаграммы к нему. Детектор используют в многих приемниках ЧМ-сигналов, а также в устройствах измерительной техники и генераторах гармонических колебаний для автоматической подстройки их частоты. ЧМ-детектор содержит резонансный усилитель-ограничитель на ОУ, к выходу которого подключены индуктивно связанные контуры (емкости $C$ и индуктивности $L$), настроенные на несущую частоту $\omega_0$ и преобразующие ЧМ-сигнал в АМ-ЧМ-колебание. Преобразованное таким образом сложное колебание детектируется одновременно двумя идентичными АМ-детекторами, включенными последовательно. Конденсатор $C_р$ в схеме — разделительный, $C_б$ — блокировочный, а дроссель $L_д$ преграждает путь протеканию току высокой частоты через нагрузку.

Принцип действия этого частотного детектора заключается в линейном преобразовании отклонения частоты в изменение фазового сдвига высокочастотного колебания и соответственно изменения направления вектора амплитуды с последующим амплитудным детектированием. Для пояснения его работы обратимся к векторным диаграммам (рис. 5.19, *б*, *г*).

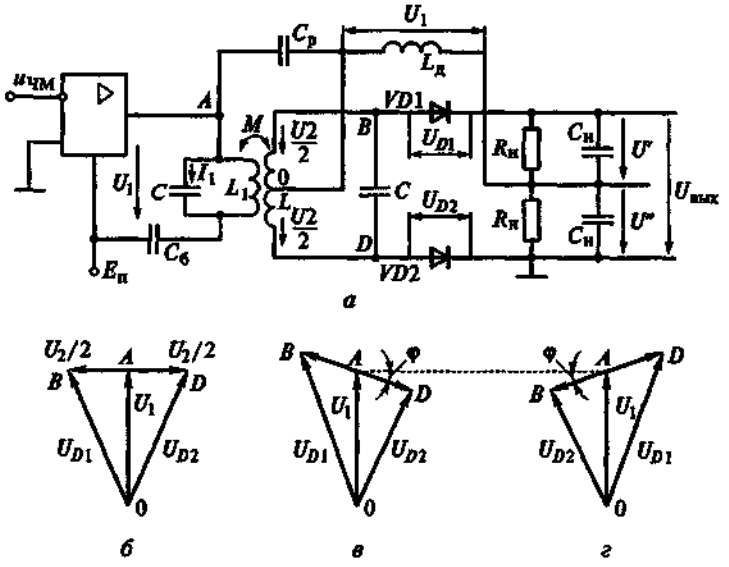



**Рис. 5.19.** Частотный детектор и векторные диаграммы к нему:
*а* — упрощенная схема; *б* — диаграмма в отсутствие модуляции;
*в* — диаграмма при частоте выше резонансной; *г* — диаграмма при частоте ниже резонансной

На диаграммах векторы напряжений отражают разности потенциалов между точками 0, *A*, *B* и *D*, показанными на рис. 5.19, *а*. Поскольку контуры усилителя и амплитудных детекторов имеют индуктивную связь, то одно из напряжений $U_2/2$ на входах диодов синфазно, а другое — противофазно выходному напряжению операционного усилителя $U_1$. Амплитудные детекторы реагируют на амплитуды высокочастотного входного колебания, поступающие на аноды диодов *VD*1 и *VD*2 (напряжения $U_{D1}$ и $U_{D2}$ относительно конденсатора $C_п$). Напряжение на аноде диода *VD*1 равно сумме двух напряжений: $U_{D1} = U_1 + U_2/2$; соответственно напряжение на аноде диода *VD*2: $U_{D2} = U_1 - U_2/2$.

В отсутствие модуляции, когда частота ЧМ-сигнала $\omega_0$ совпадает с резонансными частотами $\omega_р$ колебательных контуров, напряжение $U_1$ в контуре усилителя опережает по фазе ток $I_1$ на 90°. Напряжения $U_2/2$ и $-U_2/2$ на контурах амплитудных детекторов также опережают по фазе ток $I_1$ входного контура детектора на 90°. Соответствующая этому случаю векторная диаграмма показана на рис. 5.19, *б*. Напряжения $U_{D1}$ и $U_{D2}$ на диодах амплитудных детекторов при этом равны, и выходное напряжение $U_{вых} = 0$. При отклонении мгновенной частоты принимаемого ЧМ-сигнала $\omega$ от резонансной $\omega_р$ на некоторое значение $\Delta\omega$ напряжение $U_1$ на резонансном контуре усилителя получит по отношению к току $I_1$ дополнительный фазовый сдвиг $\pm\varphi$, знак и значение которого определяется ФЧХ параллельного контура (см. рис. 4.13, *б*). Из схемы видно, что дроссель $L_д$ по высокой частоте подключен параллельно входному контуру *CL* и поэтому на нем действует такое же напряжение $U_1$.

Векторные диаграммы напряжений на контурах на частотах ЧМ-сигнала выше и ниже резонансной представлены соответственно на рис. 5.19, *в*, *г*.

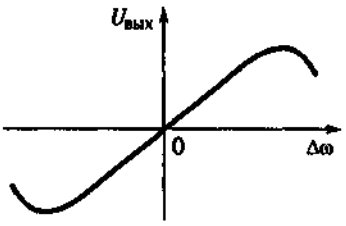



**Рис. 5.20.** Детекторная характеристика

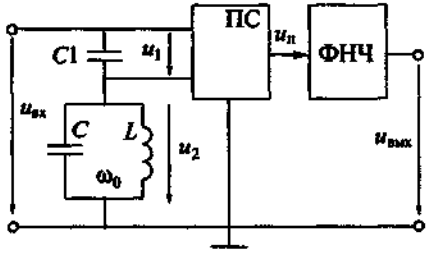



**Рис. 5.21.** ЧМ-детектор на интегральном перемножителе

Очевидно, что в этих случаях напряжения $U_{D1}$ и $U_{D2}$ не будут равны, вследствие чего на выходе ЧМ-детектора появится выходное напряжение $U_{вых}$, отражающее передаваемое сообщение. Зависимость выходного напряжения $U_{вых}$ от величины $\Delta\omega = \omega - \omega_0$ называется *детекторной характеристикой*, типичный вид которой представлен на рис. 5.20.

**ЧМ-детектор на интегральном перемножителе.** Современные детекторы ЧМ-сигналов выполняют на интегральных перемножителях (рис. 5.21). В схеме на входе аналогового перемножителя (ПС) включают параллельный резонансный контур, настроенный на несущую частоту, а на выходе — ФНЧ, настроенный на полезный сигнал. Пусть на вход ЧМ-детектора на интегральном перемножителе поступает гармоническое напряжение $u_{вх}(t) = U_{вх}\cos\omega t$. Аналоговый перемножитель обладает большими входными сопротивлениями, поэтому его входы практически не потребляют токи, и весь входной ток $i_{вх}(t) = I_{вх}\cos\omega t$ протекает через емкость $C1$ и параллельный $LC$-контур. Поскольку напряжение на конденсаторе $C1$ отстает от тока по фазе на $\pi/2$, то напряжение на входе перемножителя

$$u_1(t) = U_C\cos(\omega t - \pi/2), \tag{5.39}$$

где $U_C$ — амплитудное значение напряжения на конденсаторе $C1$.

Напряжение на контуре

$$u_2(t) = U_к\cos(\omega t + \varphi), \tag{5.40}$$

где $U_к$ — амплитудное значение напряжения, а фазовый сдвиг

$$\varphi = -\operatorname{arctg}(2Q\Delta\omega/\omega_0). \tag{5.41}$$

Здесь $\Delta\omega = \omega - \omega_0$ — абсолютная расстройка; $Q$ — добротность контура.

С помощью (5.40) и (5.41) определим напряжение на выходе интегрального перемножителя. Используя формулу произведения косинусов, запишем

$$u_п(t) = k_a u_1(t)u_2(t) = k_a U_C U_к\cos(\omega t - \pi/2)\cos(\omega t + \varphi) =$$

$$= 0{,}5k_a U_C U_к[\cos(\varphi + \pi/2) + \cos(2\omega t + \varphi - \pi/2)]. \tag{5.42}$$

Включение в схему ФНЧ позволяет избавиться от составляющей с удвоенной частотой, поэтому выходное напряжение детектора

$$U_{вых} = 0{,}5k_a U_C U_к\cos(\varphi + \pi/2) = -0{,}5k_a U_C U_к\sin\varphi. \tag{5.43}$$

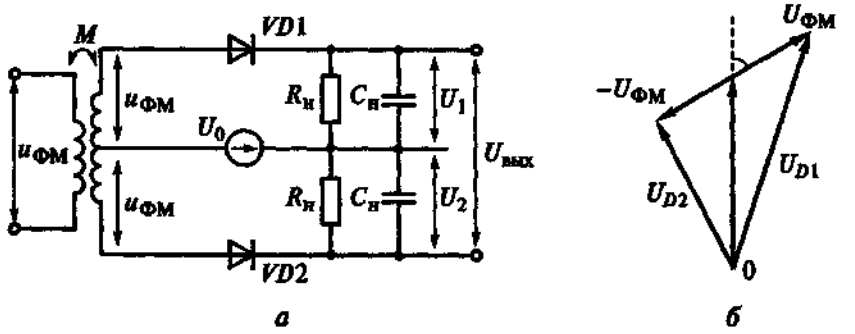



**Рис. 5.22.** Фазовый детектор:
*а* — схема; *б* — векторная диаграмма

Как правило, в резонансном контуре параметры $2Q\Delta\omega/\omega_0 << 1$, и поэтому справедливо $\sin\varphi \approx \varphi = 2Q\Delta\omega/\omega_0$. С учетом этого равенства получим

$$U_{вых} = -k_a U_C U_к Q\Delta\omega/\omega_0. \qquad (5.44)$$

Итак, согласно (5.44), схема детектора, представленная на рис. 5.21, осуществляет линейное детектирование ЧМ-сигналов, при котором выходное напряжение пропорционально отклонению частоты сигнала от несущей $\omega_0$.

**Фазовый детектор.** Фазовый детектор преобразует ФМ-колебание в низкочастотное напряжение, изменяющееся по закону модулирующего сигнала. Напряжение на выходе ФМ-детектора определяется разностью фаз двух сравниваемых сигналов (рис. 5.22, *а*): ФМ-колебания $u_{ФМ}(t) = U_{ФМ}\cos(\omega_0 t + \varphi)$ и опорного напряжения $u_0(t) = U_0\cos\omega_0 t$, создаваемого *генератором опорного напряжения*. Амплитуды напряжений на входах диодов можно определить с помощью векторной диаграммы (рис. 5.22, *б*). Из нее следует, что

$$U_{D1} = \sqrt{U_{ФМ}^2 + U_0^2 + 2U_{ФМ}U_0\cos\varphi}\,;$$
$$U_{D2} = \sqrt{U_{ФМ}^2 + U_0^2 - 2U_{ФМ}U_0\cos\varphi}\,. \qquad (5.45)$$

Полученные сигналы затем преобразуются АМ-детекторами с коэффициентом передачи $k_д$, и на нагрузках схемы возникают два напряжения

$$U_1 = k_д U_{D1} \text{ и } U_2 = k_д U_{D2}.$$

Результирующее напряжение на выходе фазового детектора

$$U_{вых} = U_1 - U_2 = k_д\left(\sqrt{U_{ФМ}^2 + U_0^2 + 2U_{ФМ}U_0\cos\varphi} - \sqrt{U_{ФМ}^2 + U_0^2 - 2U_{ФМ}U_0\cos\varphi}\right). \qquad (5.46)$$

Амплитуду опорного напряжения в фазовых детекторах выбирают намного больше амплитуды ФМ-колебания, т. е. $U_0 >> U_{ФМ}$ (обычно в 3...5 раз и более). Это делается для «вывода» полезного сигнала из напряжения шумов, поскольку они часто соизмеримы по амплитуде и сигнал может быть искажен.

В этом случае выражение (5.44) можно представить упрощенно

$$U_{\text{вых}} \approx 2k_{\text{д}}U_{\text{ФМ}}\cos\varphi. \qquad (5.47)$$

Из (5.47) следует, что низкочастотное напряжение на выходе фазового детектора изменяется практически в полном соответствии с фазой ФМ-сигнала.

В последние годы в радиоприемниках предпочтение отдают цифровым фазовым детекторам, обладающим высокой помехозащищенностью.

**Фазовые детекторы на аналоговом перемножителе.** Если подать на соответствующие входы перемножителя ФМ-сигнал $u_1(t) = U_{\text{ФМ}}\cos(\omega t + \varphi)$ и опорное колебание $u_2(t) = U_0\cos\omega t$, то напряжение на его выходе

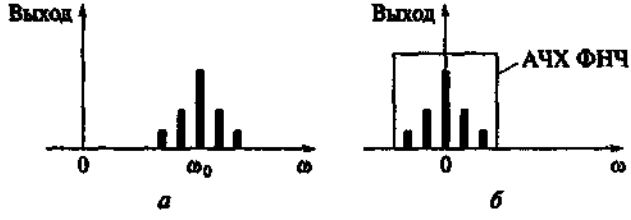
$$U_{\text{вых}} = \frac{U_{\text{ФМ}}U_0}{2}\cos\varphi - \frac{U_{\text{ФМ}}U_0}{2}\cos(2\omega t + \varphi).$$

При подключении к выходу перемножителя ФНЧ, который не пропускает сигнал удвоенной частоты, на его выходе будет присутствовать только сигнал вида (5.47), пропорциональный косинусу угла сдвига фаз входных сигналов.

В последние годы в радиоприемных устройствах предпочтение отдают цифровым фазовым детекторам, обладающим высокой помехозащищенностью.

**Синхронное детектирование.** Рассмотрим особый вид фазового детектирования — синхронное (*lock-in*), получаемое при частоте гетеродина $\omega_{\text{г}}$, равной частоте полезного сигнала $\omega_{\text{г}} = \omega_0$, при этом разностная частота $\omega_0 - \omega_{\text{г}} = 0$ и спектр выходного сигнала смещается на нулевую частоту, т. е. совпадает со спектром модулирующей функции при амплитудной модуляции (рис. 5.23). Так выполняют преобразования, обратные амплитудной модуляции. Получающаяся при преобразовании двойная частота $\omega_0 + \omega_{\text{г}} = 2\omega_0$ не пропускается ФНЧ, включенным на выходе синхронного детектора. При слабых детектируемых сигналах $U_{\text{ФМ}} << U_0$, детекторная характеристика линейна и детектирование по существу является *линейной* операцией.

Основным преимуществом синхронного детектирования является повышенная избирательность радиоприема слабых сигналов на фоне шума (устраняется взаимодействие сигнала с помехой в нелинейном устройстве, которым является обычный амплитудный детектор). Отметим, что реализация метода синхронного детектирования связана со значительными трудностями, так как обеспечение синхронизма частоты гетеродина с частотой принимаемого сигнала является сложной задачей, особенно при приеме слабых сигналов на фоне помех.




*а* *б*

**Рис. 5.23.** Спектры сигналов синхронного детектора:
*а* — на входе; *б* — на выходе

## 5.6. Преобразования сигналов в параметрических цепях

**Особенности линейно-параметрических цепей.** Радиотехнические цепи, один или несколько параметров которых изменяются во времени по заданному закону, называют *параметрическими* (*линейными цепями с переменными параметрами*). Предполагается, что изменение какого-либо параметра осуществляют электронным методом с помощью управляющего сигнала. В линейно-параметрической цепи параметры элементов не зависят от уровня сигнала, но могут независимо изменяться во времени. Реально параметрический элемент получают из нелинейного элемента, на вход которого подают сумму двух независимых сигналов. Один из них несет информацию и имеет малую амплитуду, так что в области его изменений параметры цепи практически постоянны. Вторым является управляющий сигнал большой амплитуды, который изменяет положение рабочей точки нелинейного элемента, а следовательно, его параметр.

В радиотехнике широко применяют параметрические сопротивления $R(t)$, параметрические индуктивности $L(t)$ и параметрические емкости $C(t)$.

Для параметрического сопротивления $R(t)$ управляемым параметром является дифференциальная крутизна

$$S(t) = \frac{di}{du}.$$

Малость амплитуды входного сигнала $u(t)$, сохраняющей линейность цепи, позволяет ввести передаточную функцию и импульсную характеристику, которые дополнительно зависят от времени $t$, при котором фиксируется параметр

$$\mathbf{K}(\omega, \tau),\ h(t, \tau),$$

где $\tau$ — сдвиг времени относительно момента времени $t$.

Примером *параметрического сопротивления* может служить канал МДП-транзистора, на затвор которого подано управляющее (гетеродинное) переменное напряжение $u_г(t)$. В этом случае крутизна его стоко-затворной характеристики изменяется во времени и связана с управляющим напряжением зависимостью $S(t) = S[u_г(t)]$. Если к МДП-транзистору подключить еще и напряжение модулированного сигнала $u(t)$, то его ток определится выражением

$$i_с(t) = i(t) = S(t)u(t) = S[u_г(t)]u(t). \qquad (5.48)$$

Как к классу линейных, к параметрическим цепям применим принцип суперпозиции: если приложенное напряжение является суммой двух переменных

$$u(t) = u_1(t) + u_2(t), \qquad (5.49)$$

то, подставив (5.49) в (5.48), найдем ток цепи как сумму двух составляющих

$$i(t) = S(t)u_1(t) + S(t)u_2(t) = i_1(t) + i_2(t). \qquad (5.50)$$

Формула (5.50) показывает, что отклик параметрической цепи на сумму двух сигналов равен сумме ее откликов на каждый сигнал в отдельности.

**Преобразование сигналов в цепи с параметрическим сопротивлением**

Наиболее широко параметрические сопротивления применяют для преобразования частоты сигналов. Отметим, что термин «преобразование частоты» не совсем корректен, поскольку частота сама по себе неизменна. Очевидно, это понятие возникло из-за неточного перевода английского слова «*heterodyning* — гетеродинирование». Итак, *гетеродинирование* — процесс нелинейного или параметрического смешивания двух сигналов разных частот для получения колебаний третьей частоты, в результате которого происходит смещение спектра исходного сигнала. *Преобразование частоты* — линейный перенос (смещение, трансформация, гетеродинирование, транспонирование) спектра модулированного сигнала (а также любого радиосигнала) из области несущей частоты в область промежуточной частоты (или с одной несущей частоты на другую, в том числе и на более высокую) без изменения вида или характера модуляции.

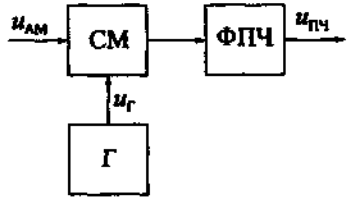



**Рис. 5.24.** Структурная схема преобразователя частоты

Дело в том, что принятый приемником высокочастотный (часто употребляют термин «полосовой») сигнал вначале преобразуется в низкочастотный, после чего наступает этап детектирования. Для линейных систем математические методы детектирования не зависят от смещения частоты, что отмечает теорема эквивалентности. Фактически *теорему эквивалентности* можно записать следующим образом: выполнение полосовой линейной обработки сигнала с последующим переносом частоты сигнала (превращением полосового сигнала в низкочастотный) дает те же результаты, что и перенос частоты сигнала с последующей низкочастотной линейной обработкой сигнала. Как следствие теоремы эквивалентности, любая линейная модель обработки сигналов может использоваться на низкочастотных сигналах (что предпочтительнее с точки зрения простоты) с теми же результатами, что и на полосовых сигналах. Это означает, что производительность большинства цифровых систем связи часто можно описать и проанализировать, считая канал передачи низкочастотным.

*Преобразователь частоты* (рис. 5.24) состоит из смесителя (СМ) — параметрического элемента (например, МДП-транзистора, варикапа и т. д.), гетеродина (Г) — вспомогательного генератора гармонических колебаний с частотой $\omega_г$, служащего для параметрического управления смесителем, и фильтра промежуточной частоты (ФПЧ) — полосового фильтра на ПАВ

Принцип действия преобразователя частоты рассмотрим на примере переноса спектра однотонального АМ-сигнала. Допустим, что под воздействием гетеродинного напряжения

$$u_г(t) = U_г \cos\omega_г t \tag{5.51}$$

крутизна характеристики МДП-транзистора изменяется приближенно по закону

$$S(t) = S_0 + S_1 \cos\omega_г t, \tag{5.52}$$

где $S_0$ и $S_1$ — соответственно среднее значение и первая гармоническая составляющая крутизны характеристики.

При поступлении на преобразующий МДП-транзистор смесителя приемника АМ-сигнала $u_{AM}(t) = U_н(1 + M\cos\Omega t)\cos\omega_0 t$ переменная составляющая выходного тока в соответствии с (5.48) и (5.50) будет определяться выражением:

$$i_c(t) = S(t)u_{AM}(t) = (S_0 + S_1\cos\omega_г t)U_н(1 + M\cos\Omega t)\cos\omega_0 t =$$

$$= U_н(1 + M\cos\Omega t)[S_0\cos\omega_0 t + 0{,}5S_1\cos(\omega_г - \omega_0)t + 0{,}5S_1\cos(\omega_г + \omega_0)t]. \quad (5.53)$$

Пусть в качестве промежуточной частоты параметрического преобразователя выбрана частота

$$\omega_{пч} = |\omega_г - \omega_0|. \quad (5.54)$$

Тогда, выделив ее с помощью контура УПЧ из спектра тока (5.53), получим преобразованный АМ-сигнал с тем же законом модуляции, но существенно меньшей несущей частотой

$$i_{пч}(t) = 0{,}5S_1U_н(1 + M\cos\Omega t)\cos\omega_{пч}t. \quad (5.55)$$

Заметим, что наличие только двух боковых составляющих спектра тока в (5.53) определяется выбором предельно простой кусочно-линейной аппроксимации крутизны характеристики транзистора. В реальных схемах смесителей в спектре тока содержатся также составляющие комбинационных частот

$$\omega_{пч} = |m\omega_г \pm n\omega_0|, \quad (5.56)$$

где $m$ и $n$ — любые целые положительные числа.

Временны́е и спектральные диаграммы сигналов с амплитудной модуляцией на входе и выходе преобразователя частоты показаны на рис. 5.25.

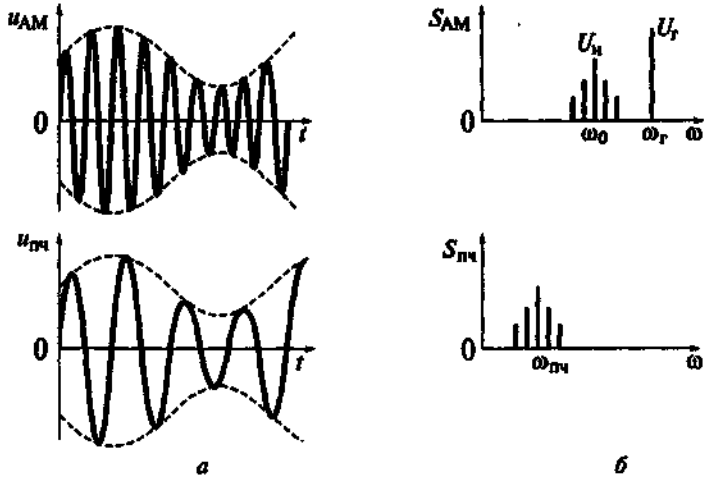



**Рис. 5.25.** Диаграммы сигналов на входе и выходе преобразователя частоты: *а* — временны́е; *б* — спектральные

**Преобразование частоты в интегральных перемножителях.** Современные преобразователи частоты с параметрическими резистивными цепями построены на принципиально новой основе. В них в качестве смесителей используются аналоговые перемножители. Если на входы аналогового перемножителя подать два гармонических колебания: модулированный сигнал

$$u_c(t) = U_c(t)\cos\omega_0 t \tag{5.57}$$

и напряжение гетеродина (5.51), то в выходном спектре будут составляющие

$$u_{вых}(t) = k_a u_c(t) u_г(t) = 0{,}5 k_a U_c(t) U_г[\cos(\omega_г - \omega_0)t + \cos(\omega_г + \omega_0)t]. \tag{5.58}$$

Спектральная составляющая с разностной частотой $\omega_{пч} = |\omega_г - \omega_0|$ выделяется узкополосным фильтром УПЧ и используется в качестве промежуточной частоты преобразованного сигнала.

**Преобразование частоты в параметрической цепи с варикапом**

Если на варикап подать только гетеродинное напряжение (5.51), то его емкость приближенно будет изменяться во времени по закону (см. рис. 5.11):

$$C(t) = C_0 + C_1\cos\omega_г t, \tag{5.59}$$

где $C_0$ и $C_1$ — среднее значение и первая гармоника емкости варикапа.

Положим, что на варикап воздействует два сигнала — гетеродинное и (для упрощения расчетов) немодулированное гармоническое напряжение (5.57) с постоянной амплитудой $U_C$. В этом случае заряд на емкости варикапа будет определяться следующим выражением

$$q(t) = C(t)u_C(t) = (C_0 + C_1\cos\omega_г t)U_C\cos\omega_0 t =$$

$$= C_0 U_C\cos\omega_0 t + 0{,}5C_1 U_C\cos(\omega_г - \omega_0)t + 0{,}5C_1 U_C\cos(\omega_г + \omega_0)t, \tag{5.60}$$

а ток, протекающий через него,

$$i(t) = dq/dt = -\omega_0 C_0 U_C\sin\omega_0 t - 0{,}5(\omega_г - \omega_0)C_1 U_C\sin(\omega_г - \omega_0)t -$$

$$- 0{,}5(\omega_г + \omega_0)C_1 U_C\sin(\omega_г + \omega_0)t. \tag{5.61}$$

Включив последовательно с варикапом колебательный контур, настроенный на промежуточную частоту $\omega_{пч} = |\omega_г - \omega_0|$, выделяют желаемый сигнал.

С реактивным элементом типа варикапа (для СВЧ-диапазона — *варактор*) можно создать также параметрический генератор, усилитель мощности, умножитель частоты. Такая возможность основана на преобразовании энергии в параметрической емкости. Из курса физики известно, что энергия, накопленная в конденсаторе, связана с его емкостью $C$ и зарядом на ней $q$ формулой

$$Э = 0{,}5q^2/C. \tag{5.62}$$

Пусть заряд остается постоянным, а емкость конденсатора уменьшается. Поскольку энергия обратно пропорциональна значению емкости, то при уменьшении последней энергия растет.

Количественное соотношение этой связи получим, дифференцируя соотношение (5.62) по параметру $C$

$$\frac{d\mathcal{Э}}{dC} = -\frac{q^2}{2C^2} = -\frac{\mathcal{Э}}{C}. \tag{5.63}$$

Это выражение также справедливо и для малых приращений емкости $\Delta C$ и энергии $\Delta\mathcal{Э}$, поэтому можно записать

$$\Delta\mathcal{Э} = -\Delta C\frac{\mathcal{Э}}{C}. \tag{5.64}$$

Знак минус здесь показывает, что при постоянном заряде уменьшение емкости конденсатора ($\Delta C < 0$) вызывает увеличение запасаемой в нем энергии ($\Delta\mathcal{Э} > 0$). Увеличение энергии происходит за счет внешних затрат на выполнение работы против сил электрического поля при уменьшении емкости (например, путем изменения напряжения смещения на варикапе, или разнесении обкладок конденсатора в электрическом поле).

При одновременном воздействии на параметрическую емкость (или параметрическую индуктивность) нескольких источников сигналов с разными частотами между источниками будет происходить *перераспределение* (*обмен*) *их энергий*. Обычно энергия внешнего источника, называемого *генератором накачки*, через параметрический элемент передается в цепь полезного сигнала.

Для анализа энергетических соотношений в многоконтурных цепях с параметрической емкостью обратимся к обобщенной схеме (рис. 5.26). В ней параллельно параметрической емкости $C$ включены три цепи, две из которых содержат источники $e_1(t)$ и $e_2(t)$, создающие гармонические колебания с частотами $\omega_1$ и $\omega_2$. Источники колебаний $e_1(t)$ и $e_2(t)$ соединены через узкополосные фильтры $\Phi_1$ и $\Phi_2$, пропускающие сигналы с частотами $\omega_1$ и $\omega_2$. Третья цепь содержит сопротивление нагрузки $R_н$ и узкополосный фильтр $\Phi_3$ — *холостой контур*, настроенный на комбинационную частоту

$$\omega_3 = m\omega_1 + n\omega_2, \tag{5.65}$$

где $m$ и $n$ — целые числа.

Положим для упрощения, что в схеме с параметрической емкостью применены фильтры без потерь (теоретически параметрическая емкость тоже не потребляет мощности). Если источники колебаний $e_1(t)$ и $e_2(t)$ отдают мощности $P_1$ и $P_2$, то сопротивление нагрузки $R_н$ потребляет мощность $P_н$. Для замкнутой системы в соответствии с законом сохранения энергии получаем условие баланса мощностей

$$P_1 + P_2 + P_н = 0. \tag{5.66}$$

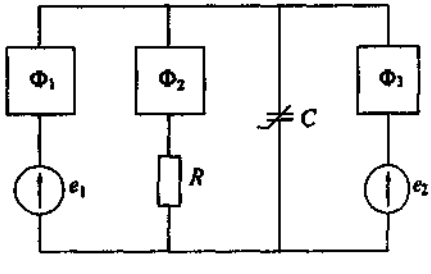



**Рис. 5.26.** Обобщенная схема с параметрической емкостью

Умножим и разделим каждое слагаемое на соответствующую частоту

$$\frac{\omega_1 P_1}{\omega_1} + \frac{\omega_2 P_2}{\omega_2} + \frac{(m\omega_1 + n\omega_2)P_н}{m\omega_1 + n\omega_2} = 0. \qquad (5.67)$$

Запишем это уравнение в другом виде

$$\omega_1\left(\frac{P_1}{\omega_1} + \frac{mP_н}{m\omega_1 + n\omega_2}\right) + \omega_2\left(\frac{P_2}{\omega_2} + \frac{nP_н}{m\omega_1 + n\omega_2}\right) = 0. \qquad (5.68)$$

Полученное равенство должно быть тождественным при любых частотах. Это возможно, если

$$\frac{P_1}{\omega_1} + \frac{mP_н}{m\omega_1 + n\omega_2} = 0; \qquad \frac{P_2}{\omega_2} + \frac{nP_н}{m\omega_1 + n\omega_2} = 0. \qquad (5.69)$$

Эти фундаментальные соотношения в радиотехнике называют *уравнениями Мэнли–Роу*. Они позволяют просто и наглядно выяснить закономерность преобразования мощностей сигнала и накачки в многоконтурных параметрических цепях. Рассмотрим наиболее распространенный случай, имеющий место в параметрических цепях устройств радиоэлектроники.

**Параметрическое усиление с преобразованием частоты**

Положим, что в обобщенной параметрической схеме (см. рис. 5.26,) $e_1 = e_1(t)$ — источник усиливаемого сигнала, а $e_2 = e_2(t)$ — генератор накачки. Установив в формулах (5.69) значения коэффициентов $m = n = 1$, получим

$$\frac{P_1}{\omega_1} + \frac{P_н}{\omega_1 + \omega_2} = 0; \qquad \frac{P_2}{\omega_2} + \frac{P_н}{\omega_1 + \omega_2} = 0. \qquad (5.70)$$

Будем считать (как это принято в физике) мощность, выделяемую на активной нагрузке, положительной, а мощность, отдаваемую генератором, отрицательной. Тогда из (5.70) следует, что при $P_н > 0$ мощности $P_1 < 0$ и $P_2 < 0$. Значит, при настройке холостого контура на частоту $\omega_3 = \omega_1 + \omega_2$ источник усиливаемого сигнала и генератор накачки будут отдавать мощность в нагрузку, и схема на рис. 5.26 превращается в параметрический усилитель входного сигнала. При этом усиление осуществляется за счет генератора накачки и сопровождается повышением частоты сигнала. Отметим, что параметрический усилитель обладает малым уровнем шумов, поскольку механизм усиления в нем не связан с транспортировкой зарядов. Также в параметрических усилителях отсутствуют дробовые шумы, являющиеся основными в обычных усилителях. Известно также, что параметрический усилитель может самовозбуждаться.

С помощью параметрической цепи можно осуществить и усиление мощности входного сигнала с понижением его частоты. Можно показать, что это будет иметь место при настройке холостого контура на разностную частоту $\omega_3 = \omega_2 - \omega_1$ и выборе $m = 1$, $n = -1$ в (5.69). Обычно частота накачки равна удвоенной частоте сигнала, что удобно с практической точки зрения.

## 5.7. Логарифмирование сигналов

Нелинейные элементы позволяют существенно расширить функциональные возможности устройств, выполненных на ОУ. При этом ОУ продолжает, как правило, работать в линейной области усиления.

Если вместо резистора $R_{ос}$ в цепи обратной связи инвертирующего усилителя на ОУ ввести НЭ (например, диод *VD*), то в этом случае схема, показанная на рис. 5.27, может выполнять операцию логарифмирования входного сигнала. Коротко поясним суть процесса.

Как известно из курса физики, сопротивление полупроводникового диода (условно обозначим его через $R_D$), как нелинейного элемента, меняется в зависимости от величины тока $i_D$, протекающего через него, — при малых токах оно достаточно велико, а при больших — существенно меньше. В результате такого физического явления рассматриваемое устройство с инвертирующим включением ОУ при малых входных токах обладает большим коэффициентом усиления, а при больших входных токах — малым. С помощью несложных математических выкладок покажем, что в рассматриваемой схеме (рис. 5.27) выходное напряжение пропорционально логарифму входного напряжения.

Из теории полупроводниковых приборов известно, что вольт-амперная характеристика при прямом смещении диода описывается экспонентой

$$i_D = I_0 \exp[u_D/\varphi_T], \tag{5.71}$$

где $I_0$ — обратный ток диода; $u_D$ — напряжение на диоде; $\varphi_T$ — температурный потенциал.

При анализе работы устройства учтем, что в схеме с идеальным ОУ входной ток равен току цепи обратной связи (току диода), т. е. $i_{вх} = u_{вх}/R = i_D$. Тогда из выражения (5.71) находим, что напряжение на полупроводниковом диоде, а следовательно, и выходное напряжение равно:

$$u_{вых} = -u_D = m\varphi_T \ln[u_{вх}/(I_0 R)]. \tag{5.72}$$

Это соотношение показывает, что выходное напряжение в рассматриваемой схеме является логарифмической функцией от входного напряжения. Следует особо подчеркнуть, что данное устройство является однополярной схемой и конкретно в нем для прямого смещения диода входное напряжение $u_{вх}$ должно быть всегда положительным.

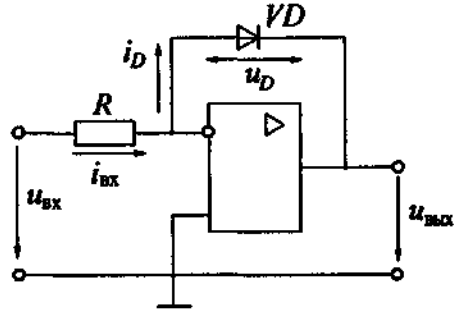



**Рис. 5.27.** Логарифмический усилитель

## Контрольные вопросы

1. Какие радиотехнические цепи относятся к нелинейным?
2. Для чего используют аппроксимацию характеристик НЭ?
3. Какие виды аппроксимации характеристик НЭ используются в радиоэлектронике и теории связи?
4. Когда удобнее применять степенную или кусочно-линейную аппроксимацию?
5. Как определяется спектральный состав тока в цепи с безынерционным НЭ при гармоническом входном сигнале и различных видах аппроксимации нелинейных характеристик?
6. Почему, несмотря на импульсный характер тока в нелинейном резонансном усилителе, его выходное напряжение имеет форму входного сигнала?
7. Каков физический принцип действия умножителя частоты?
8. На чем основан принцип работы простейшего амплитудного модулятора?
9. На чем основано действие частотного модулятора с варикапом?
10. Как действует модулятор Армстронга?
11. Опишите схему линейного АМ-детектора.
12. В чем состоит основное отличие линейного и квадратичного детектирования АМ-сигналов?
13. На чем основан принцип детектирования ЧМ-сигнала?
14. Для каких целей в фазовом детекторе используют опорное напряжение?
15. Чем характерен спектр тока в цепи с параметрическим сопротивлением?
16. На каком принципе основана работа варактора в схемах многоконтурных электрических цепей?
17. В чем заключается принцип действия логарифмического усилителя на ОУ?

# Глава 6. ГЕНЕРИРОВАНИЕ ЭЛЕКТРОМАГНИТНЫХ КОЛЕБАНИЙ

Важное значение в радиотехнике имеют колебательные системы, генерирующие электромагнитные колебания. *Колебательной системой*, или устройством с самовозбуждением, называют динамическую систему, преобразующую энергию источника постоянного тока в энергию незатухающих колебаний, причем основные характеристики колебаний (амплитуда, частота, форма колебаний и т. д.) определяются, главным образом, параметрами самой системы. Процесс получения переменных сигналов требуемой формы и частоты называют *генерированием электрических колебаний.* С точки зрения математических моделей колебательные системы разделяют на линейные и нелинейные, автономные и неавтономные. Особый класс представляют автоколебательные системы, или автогенераторы.

В радиопередатчиках систем связи автогенераторы применяют в основном в качестве каскадов, создающих электромагнитные колебания несущей частоты. Основное требование — это высокая стабильность генерируемой частоты. В СВЧ-диапазоне автогенераторы часто используют в качестве выходных каскадов передатчиков. Требования к таким автогенераторам аналогичны требованиям к усилителям мощности — обеспечение мощных колебаний при высоких КПД, выходной мощности и стабильности частоты.

Основное внимание в этой главе уделено маломощным высокостабильным задающим генераторам; вместе с тем изложенные сведения будут полезны и при изучении мощных генераторов выходных каскадов передатчиков.

## 6.1. Общие сведения. Генераторы гармонических колебаний

В радиоэлектронике и технике связи разработано и используют большое количество различных схем автогенераторов. Классифицируют автогенераторы по нескольким специфическим признакам.

1. В зависимости от диапазона генерируемых частот автогенераторы делят на три большие группы: низкочастотные, высокочастотные и сверхвысокочастотные. Отличительным признаком может служить даже не само значение частоты генерируемых колебаний, а тип используемых электрических цепей. В низкочастотных и высокочастотных генераторах таковыми являются цепи с сосредоточенными параметрами, в сверхвысокочастотных — с распределенными параметрами, т. е. фидерные и полосковые линии и волноводы.

2. По использованию в составе радиотехнического устройства существует такое разделение автогенераторов:

• опорные или эталонные с повышенной стабильностью частоты, синхронизирующие работу всех звеньев и каскадов устройства;

• диапазонные, перестраиваемые по частоте, в том числе и в составе синтезаторов частот.

3. По взаимодействию с другими узлами аппаратуры различают автогенераторы функционирующие в:

• автономном режиме;

• режиме синхронизации частоты внешним сигналом;

• составе схемы автоматической подстройки частоты.

Для преобразования энергии источника постоянного тока в энергию колебаний требуется активный элемент. Активными (усилительными) элементами в генераторах могут быть электронные лампы, биполярные и полевые транзисторы, интегральные микросхемы (операционные усилители), а также генераторные диоды — туннельные, лавинно-пролетные, диоды Ганна и др.

4. По типу полупроводникового усилительного прибора и схеме различают два основных типа автогенераторов:

• на основе усилительного элемента (транзистора, электровакуумного прибора или интегральной микросхемы) с использованием резонатора и цепи положительной обратной связи;

• с применением генераторного СВЧ-диода (диода Ганна, туннельного или лавинно-пролетного), т. е. двухполюсника с отрицательным дифференциальным сопротивлением в эквивалентной схеме.

Положительная ОС в автогенераторе может быть внешней и внутренней.

Принцип действия автогенератора на основе усилительного элемента связан с обеспечением его режима работы приблизительно такого же, как и в усилителе мощности. При этом на вход усилителя подают колебания не от внешнего источника, а из собственной колебательной системы (контура или резонатора) через цепь внешней ПОС. Введение обратной связи, с одной стороны, позволяет в ряде случаев существенно улучшить характеристики цепей, с другой, при определенных условиях эти цепи становятся неустойчивыми и в них возникают автоколебания. На этом принципе построены автоколебательные системы, прежде всего автогенераторы гармонических колебаний.

Диодные автогенераторы обеспечивают стационарные колебания за счет внутренних физических процессов в генераторных диодах (появления на ВАХ участка отрицательного дифференциального сопротивления), обратная связь здесь внутренняя и осуществляется автоматически без применения специальных дополнительных элементов.

Автогенератор является каскадом, обязательно входящим в радиопередающее и радиоприемное устройства. Автогенератор существенно отличается от других каскадов радиопередатчиков тем, что частота и амплитуда колебаний здесь определяются не внешним источником, а параметрами собственной колебательной системы и активного элемента. Работу автогенератора характеризуют

следующие основные параметры: диапазон частот или значение фиксированной частоты, мощность автоколебаний в нагрузке, нестабильность частоты — долговременная и кратковременная.

Прежде чем приступить к анализу автогенераторов, сформулируем основные понятия. Итак, устройства, предназначенные для создания электрических колебаний, называют *генераторами*. С точки зрения режима работы их разделяют на автогенераторы и генераторы с внешним возбуждением. Поэтому-то автогенераторы, в отличие от генераторов с внешним возбуждением (усилителей мощности, в которых амплитуда и частота колебаний навязываются извне), часто называют генераторами с самовозбуждением.

*Автогенератор* (часто, просто *генератор*) — устройство, преобразующее энергию постоянного тока в энергию электрических колебаний требуемой частоты и формы. Автогенератор можно считать источником электромагнитных колебаний, возбуждающихся самопроизвольно без внешнего воздействия.

*Генератор с внешним возбуждением* (в импульсной технике — *ждущий генератор*) переходит в режим генерации, формирования, или усиления колебаний только при поступлении на его вход сигналов возбуждения (запуска).

В зависимости от формы генерируемых напряжений различают *генераторы гармонических и релаксационных (импульсных) колебаний*. Можно специально отметить и автогенераторы линейных пилообразных напряжений, которые относят к релаксационным. Особую группу представляют автогенераторы *случайных* колебаний (сигналов или шумов) — шумовые генераторы, рассматриваемые в учебнике схематично.

Генератор гармонических колебаний должен обязательно содержать *узкополосную колебательную систему*. Принцип действия релаксационных генераторов основан на зарядно-разрядных или накопительно-поглощающих явлениях, протекающих в широкополосных энергоемких цепях ПОС.

Генератор гармонических колебаний (к ним относятся и СВЧ-генераторы) вырабатывает колебание, спектр которого содержит практически одну гармонику. Основным элементом генератора гармонических колебаний является резонатор, главное свойство которого — колебательный характер переходного процесса в нем. Простейший резонатор — колебательный контур. Если в колебательный контур ключом ввести энергию, то при достаточно высокой его добротности ($Q >> 1$) возникают колебания тока, затухающие со временем. Уменьшение амплитуды колебаний объясняется потерями мощности в контуре. Итак, для создания автогенератора гармонических колебаний необходимо использовать резонатор с достаточно высокой добротностью и компенсировать потери. Для выполнения последнего условия необходимо периодически вводить в резонатор электромагнитную энергию синхронно с возбуждаемыми колебаниями. В качестве резонаторов в диапазоне высоких частот, кроме $LC$-контуров, применяют кварцевые пластины; на СВЧ — отрезки линий с распределенными параметрами, диэлектрические шайбы, ферритовые сферы и др.

Выходные колебания релаксационного генератора содержат спектр гармонических составляющих, часто имеющих соизмеримые амплитуды.

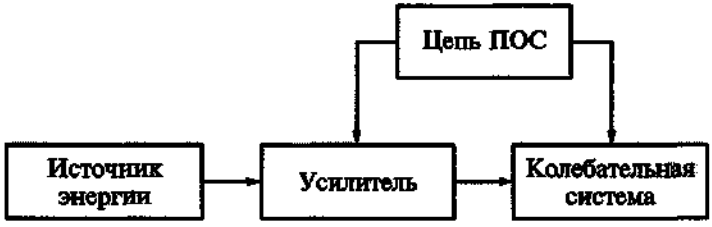



**Рис. 6.1**. Обобщенная структурная схема автогенератора

Независимо от назначения, принципа действия и схемотехнического выполнения автогенератор любых перечисленных колебаний (кроме параметрических схем и схем на генераторных диодах) состоит из *нелинейного усилителя, резонатора, цепи положительной обратной связи* (ОС, точнее ПОС) и *источника энергии* (*источника питания постоянного тока*).

Обобщенная структурная схема автогенератора приведена на рис. 6.1. В этой схеме с помощью цепи ПОС часть мощности выходного сигнала из колебательной системы (колебательного контура) поступает на вход усилителя и после усиления в нем вновь возвращается в контур. При этом необходимо выполнение двух основных условий. Во-первых, количество дополнительной энергии, поступающей в контур, должно быть равно количеству энергии, теряемой в нем за счет его активного сопротивления потерь элементов. Во-вторых, дополнительные колебания, поступающие по цепи ПОС на вход усилителя, должны совпадать по фазе с основными колебаниями.

### Самовозбуждение автогенератора

Прежде чем переходить к рассмотрению теории колебательных систем, отметим, что выдающийся вклад в теорию автоколебаний внесли русские ученые, академики Александр Александрович Андронов (1901–1952), Леонид Исаакович Мандельштам (1879–1944), Ю.Б. Кобзарев и многие др. А.А. Андронов первый (1928 г.) создал эффективный математический аппарат для рассмотрения задач теории нелинейных колебаний, с помощью которого он разработал основы строгой теории *автоколебаний* — колебаний. Период этих колебаний определяется параметрами самой системы.

При анализе и расчете автогенераторов электрических колебаний необходимо обязательно решить две основные задачи:

- выяснить, при каких условиях устройство с обратной связью становится неустойчивым, т. е. самовозбуждается;
- определить амплитуду и частоту автоколебаний генератора в стационарном режиме.



А.А. Андронов



Л.И. Мандельштам

Первую из этих задач решить относительно несложно, поскольку при малых амплитудах автоколебаний на начальном этапе процесса нелинейный усилительный элемент может быть эквивалентно заменен линейной схемой замещения. Сложнее решить вторую задачу, заключающуюся в исследовании автоколебательной системы с обратной связью при условии, что нелинейными эффектами пренебречь нельзя.

Рассмотрим условия самовозбуждения автогенератора гармонических колебаний (см. рис. 6.1). Механизм возникновения колебаний можно упрощенно трактовать следующим образом. При запуске автогенератора в его колебательной системе самопроизвольно возникают слабые свободные колебания, обусловленные включением источников питания, замыканием цепей, скачками токов и напряжений в усилительном приборе, разряда емкостей и т. д. Благодаря цепи ПОС часть энергии колебаний, возникающих на выходе усилителя, поступает на его вход. При соответствующем выборе параметров такая колебательная система становится неустойчивой.

Если мощность, отдаваемая усилительным элементом, превышает мощность, потребляемую колебательной системой (резонатором) и нагрузкой, т. е. выполняется условие самовозбуждения, то амплитуда любых малых колебаний начинают неограниченно возрастать. Однако по мере роста амплитуды существенную роль начинают играть нелинейные свойства усилительного элемента, в результате рост отдаваемой мощности замедляется и при некоторой амплитуде колебаний отдаваемая мощность оказывается равной потребляемой мощности. Это ведет к тому, что амплитуда автоколебаний достигает некоторого установившегося значения и в дальнейшем остается практически постоянной. Ввиду наличия узкополосной (обязательно высокодобротной) колебательной системы описанные процессы происходят практически на одной частоте $\omega_0$ и резко затухают на других частотах. Говорят, что при этом автогенератор работает в *стационарном режиме*.

Наиболее распространенными в радиотехнике устройствами получения гармонических колебаний являются схемы *LC*-генераторов с колебательными контурами для получения высокочастотных колебаний и *RC*-генераторов — низкочастотных колебаний.

Для теоретических описания и расчетов автогенератор гармонических колебаний (как, впрочем, и колебаний любой формы и частоты) можно представить эквивалентной структурной схемой (рис. 6.2), содержащей нелинейный усилительный каскад с комплексным коэффициентом усиления $\mathbf{K} = K(j\omega)$ и обобщенной цепи положительной ОС с комплексным коэффициентом передачи по напряжению $\beta = \beta(j\omega)$. В представленной схеме автогенератора отмечены комплексные амплитуды следующих напряжений: входного $\mathbf{U}_{вх} = U_{вх}(j\omega)$; выходного $\mathbf{U}_{вых} = U_{вых}(j\omega)$ и обратной связи $\mathbf{U}_{ос} = U_{ос}(j\omega)$.

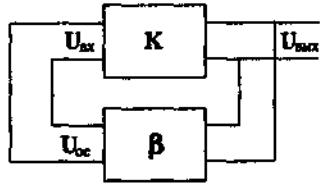



**Рис. 6.2.** Эквивалентная структурная схема автогенератора

Напряжение ПОС на любой частоте генерации ω запишем в виде

$$\mathbf{U_{ос} = U_{вх} = \beta U_{вых}}. \quad (6.1)$$

Тогда выходное напряжение автогенератора $\mathbf{U_{вых} = KU_{вх}}$, или с учетом (6.1)

$$\mathbf{U_{вых} = K\beta U_{вых}}. \quad (6.2)$$

Как следует из соотношения (6.2), автогенератор будет работать в стационарном режиме при условии, что

$$\mathbf{K\beta = 1}. \quad (6.3)$$

Если же $\mathbf{K\beta > 1}$, то амплитуда выходных колебаний будет непрерывно нарастать, что определяет необходимое условие самовозбуждения генератора.

Представим формулу (6.3) следующим образом:

$$K(\omega)e^{j\varphi_к(\omega)}\beta(\omega)e^{j\varphi_\beta(\omega)} = K\beta e^{j(\varphi_к+\varphi_\beta)} = 1. \quad (6.4)$$

Здесь $K(\omega) = K$ и $\beta(\omega) = \beta$ — действительные значения коэффициента усиления собственно усилителя (без цепи ПОС) и коэффициента передачи цепи положительной ОС; $\varphi_к(\omega) = \varphi_к$ и $\varphi_\beta(\omega) = \varphi_\beta$ — фазовые сдвиги, вносимые соответственно усилителем и цепью ПОС на текущей частоте ω.

В теории автогенераторов формулу (6.4) представляют двумя равенствами:

$$K\beta = K_{ос} = 1; \quad (6.5)$$

$$\varphi_к + \varphi_\beta = 2\pi n, \quad (6.6)$$

где $K_{ос}$ — коэффициент усиления усилителя с ОС; $n = 0, 1, 2, 3, ...$

Соотношение (6.5) определяет *условие баланса амплитуд* в автогенераторе. Из него следует, что в стационарном режиме на генерируемой частоте коэффициент усиления усилителя с обратной связью $K_{ос}$ равен единице.

Равенство (6.6) характеризует *условие баланса фаз*. Оно показывает, что в стационарном режиме суммарные фазовые сдвиги сигнала на частоте генерации, создаваемые усилителем и цепью ПОС, должны быть равны (или кратны) $2\pi$. Следует отметить, что только условие баланса фаз позволяет определить частоту генерируемых колебаний.

В схемах генераторов, работающих в стационарном режиме, условия (6.5) и (6.6) выполняются на одной частоте $\omega_р$, которая является *резонансной* для узкополосной колебательной системы. При работе генератора негармонических колебаний условия (6.5) и (6.6) должны выполняться для полосы частот.

В генераторах гармонических колебаний в качестве узкополосных колебательных систем используют *LC*-контуры и частотно-зависимые (*фазирующие*) *RC*-цепи. Генераторы гармонических колебаний с резонансными контурами называют *LC-генераторами*, а с фазирующими *RC*-цепями — *RC-генераторами*. *LC*-генераторы способны вырабатывать колебания достаточно высокой частоты (более 100 кГц), а *RC*-генераторы используют для создания гармонических колебаний низкочастотного диапазона (от долей герц до десятков килогерц).

## 6.2. *LC*-генераторы

Условия самовозбуждения, механизм возникновения стабильных гармонических колебаний и зависимость их частоты от параметров схемы рассмотрим на примере *LC*-генератора на операционном усилителе — ОУ (рис. 6.3). Усилитель *LC*-генератора охвачен двумя цепями обратной связи, обеспечивающих условия балансов амплитуд и фаз. Баланс амплитуд устанавливает цепь отрицательной обратной связи, включающей резисторы $R_1$ и $R_2$. С ее помощью задается необходимый коэффициент усиления собственно усилителя $|K| = R_2/R_1$. Баланс фаз обеспечивается в автогенераторе цепью положительной обратной связи, состоящей из последовательно включенных резистора $R$ и параллельного колебательного *LC*-контура.

Пусть при включении питания ОУ в колебательном контуре автогенератора возникли слабые свободные электрические колебания. Благодаря наличию цепи положительной ОС часть энергии свободных колебаний передается на неинвертирующий вход усилителя, усиливаются им и поступают на выход схемы автогенератора. Если выходные колебания обладают энергией, превосходящей потери в нагрузке, амплитуда напряжения на колебательном контуре будет расти до тех пор, пока нелинейность передаточной характеристики ОУ не станет ограничивать этот рост.

Итак, для поддержания автоколебаний на выходе автогенератора энергия, поставляемая в колебательный контур усиленными усилителем колебаниями, должна быть не меньше потерь энергии свободных колебаний в резонансном контуре. Это условие и есть выполнение баланса амплитуд.

Нетрудно понять, что это условие является необходимым для поддержания режима незатухающих колебаний в автогенераторе, поскольку в случае его невыполнения свободные колебания в контуре быстро затухнут. Однако выполнения этого условия недостаточно для поддержания режима незатухающих колебаний в анализируемой схеме автогенератора; для этого еще очень важны фазовые соотношения между свободными колебаниями в резонансном контуре и колебаниями, поступающими в контур из выходной цепи ОУ.

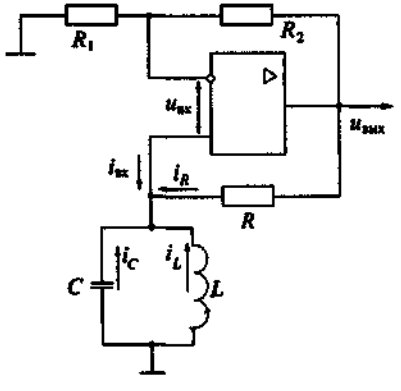



**Рис. 6.3.** Схема *LC*-генератора на ОУ

Фаза усиленных в усилителе на ОУ колебаний должна совпадать с фазой свободных колебаний в резонансном контуре, так как в ином случае усиленные усилителем колебания приведут к уменьшению, а не к увеличению амплитуды свободных колебаний в контуре. Это условие синфазности усиленных ОУ и свободных колебаний и есть баланс фаз.

Коэффициент передачи цепи ПОС определяется формулой

$$\beta = \frac{R_0}{R_0 + R}, \tag{6.7}$$

где $R_0$ — резонансное сопротивление параллельного контура.

На основании первого закона Кирхгофа запишем уравнение токов для неинвертирующего входа ОУ (аргумент $t$ в функциях тока и напряжения здесь и далее опущен для упрощения)

$$i_{вх} + i_R + i_C + i_L = 0.$$

Поскольку в идеальном ОУ $i_{вх} = 0$, то

$$i_R + i_C + i_L = 0. \tag{6.8}$$

Выразим эти токи через соответствующие им напряжения

$$\frac{u_{вх} - u_{вых}}{R} + C\frac{du_{вх}}{dt} + \frac{1}{L}\int u_{вх}dt = 0. \tag{6.9}$$

Введем известное соотношение $K = u_{вых}/u_{вх}$. Тогда, после дифференцирования по времени и несложных преобразований, уравнение (6.9) примет вид

$$\frac{1-K}{R}\frac{du_{вых}}{dt} + C\frac{d^2u_{вых}}{dt^2} + \frac{1}{L}u_{вых} = 0.$$

Поделив все члены этой формулы на $C$, запишем

$$\frac{d^2u_{вых}}{dt^2} + \frac{1-K}{RC}\frac{du_{вых}}{dt} + \omega_p^2 u_{вых} = 0, \tag{6.10}$$

где $\omega_p = 1/\sqrt{LC}$ — *резонансная частота контура.*

Обозначив *эквивалентный коэффициент затухания*

$$\alpha = \frac{1-K}{2RC}, \tag{6.11}$$

получим общепринятую в радиотехнике форму записи дифференциального уравнения, описывающего колебательный процесс в рассматриваемом автогенераторе:

$$\frac{d^2u_{вых}}{dt^2} + 2\alpha\frac{du_{вых}}{dt} + \omega_p^2 u_{вых} = 0. \tag{6.12}$$

Уравнение (6.12) является *нелинейным*, так как коэффициент усиления усилителя $K$, а следовательно, и параметр $\alpha$, зависят от значения входного напряжения. Как показывает математический анализ, точное решение уравнения (6.12) очень сложно и громоздко, поэтому в теории автогенераторов используют приближенные методы. Для подобной схемы автогенератора наиболее простой (с достаточно грубым приближением) способ решения состоит в линеаризации уравнения (6.12), и его можно применить для определения условий самовозбуж-

дения генератора. Решение будет точным лишь при малых амплитудах входного напряжения, когда рабочий участок амплитудной характеристики усилителя можно считать линейным. Полученное решение линеаризированного дифференциального уравнения (6.12) описывает гармоническое колебание с экспоненциально изменяющейся амплитудой:

$$u_{вых}(t) = U(0)e^{-\alpha t}\sin\omega_{св}t, \qquad (6.13)$$

где $U(0)$ — постоянная, определяемая начальными условиями, а

$$\omega_{св} = \sqrt{\omega_p^2 - \alpha^2} \qquad (6.14)$$

— частота свободных колебаний в контуре.

**Пример 6.1.** Вычислить, насколько частота выходных колебаний генератора на ОУ отклоняется от резонансной частоты колебательного контура (см. рис. 6.3), если изменение фазы $\varphi_к = 25°$. Контур имеет следующие параметры: $L = 160$ мкГ; $C = 160$ пФ; $Q = 50$. Выходное сопротивление ОУ считать равным нулю.

**Р е ш е н и е.** Фазовая характеристика параллельного колебательного контура вблизи резонансной частоты определяет следующая формула: $\varphi_к = -\operatorname{arctg}(2Q\Delta f/f_p)$, где $\Delta f = f - f_p$ — относительная расстройка. Подставив сюда все заданные значения параметров, находим отклонение частоты:

$$\Delta f = f_p(\operatorname{tg}\varphi_к)/(2Q) = 4\,500\ \text{Гц} = 4{,}5\ \text{кГц}.$$

Характер возникновения и изменения амплитуды выходных колебаний зависит от значения и знака параметра $\alpha$, и соответственно, от коэффициента усиления $K$. При работе автогенератора возможны три специфических случая (рис. 6.4).

1. $\alpha = 0$ ($K = 1$). Генерируется выходное гармоническое колебание с постоянной амплитудой и частотой (рис. 6.4, *а*).

2. $\alpha < 0$ ($K > 1$). Возникают выходные колебания, амплитуда которых нарастает по экспоненциальному закону (рис. 6.4, *б*).

3. $\alpha > 0$ ($K < 1$). Амплитуда выходных колебаний генератора затухает по экспоненциальному закону (рис. 6.4, *в*).

Несложный анализ выведенных положений показывает, что самовозбуждение генератора, как и следовало ожидать, возможно при коэффициенте усиления $K > 1$. Амплитуда выходного колебания в этом случае будет нарастать до перехода усилителя в нелинейный режим усиления. Из-за нелинейности амплитудной характеристики

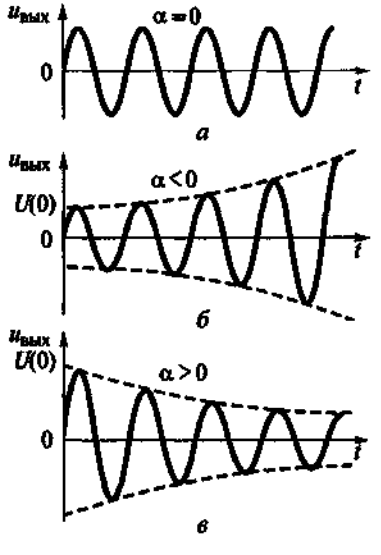



**Рис. 6.4.** Характер изменения колебаний в *LC*-генераторе: *а* — $\alpha = 0$; *б* — $\alpha < 0$; *в* — $\alpha > 0$

усилителя с цепью ПОС величина $K_{oc}$ будет автоматически уменьшаться до единицы и завершится переводом автогенератора в стационарный режим. Отметим, что реальная форма кривой выходного колебания несколько отличается от синусоидальной. Однако на достаточно высоких частотах несложно реализовать колебательный контур с большой добротностью, поэтому выходное колебание может быть практически синусоидальным с частотой колебаний $\omega_p = \omega_{св}$.

**Пример 6.2.** Задана схема автогенератора на высокоточном ОУ (см. рис. 6.3), работающего в стационарном режиме с циклической частотой выходных колебаний $f_p = 1$ МГц. Определить основные параметры элементов автогенератора, если $Q = 50$; $L = 75$ мкГ; $K = 1{,}1$; $R_1 = 10$ кОм.

Р е ш е н и е. Поскольку в стационарном режиме произведение $K\beta = 1$ и задан коэффициент $K = 1{,}1$, то $\beta = 0{,}91$. Определим параметры колебательного контура: $C = 1/(4\pi^2 f_p^2 L) = 340$ пФ; $\rho = 2\pi f_p L = 470$ Ом; $R_0 = \rho Q = 23$ кОм. Поскольку коэффициент $\beta = R_0/(R_0 + R)$, то при $\beta = 0{,}91$ имеем $R = 2{,}6$ кОм. Сопротивление резистора $R_2$ цепи ООС находим из формулы (4.122): $R_2 = |K|R_1 = 1{,}1 \cdot 10 = 11$ кОм.

### Режимы самовозбуждения автогенераторов

Чтобы выявить особенности *самовозбуждения* генератора и определить *стационарную амплитуду* выходных колебаний, удобно использовать метод совместного анализа амплитудной характеристики усилителя $K$ и прямой линии ОС $\beta = U_{oc}/U_{вых}$, отражающей влияние цепи ПОС (рис. 6.5). Отметим, что амплитудная характеристика собственно усилителя в теории генераторов соответствует *колебательной характеристике*. Суть метода традиционна и заключается в том, что схему генератора (см. рис. 6.3) мысленно (и по существу) разделяют на две цепи — линейную и нелинейную. Линейная цепь представляет петлю ПОС, а нелинейная — собственно усилитель (ОУ и цепь ООС).

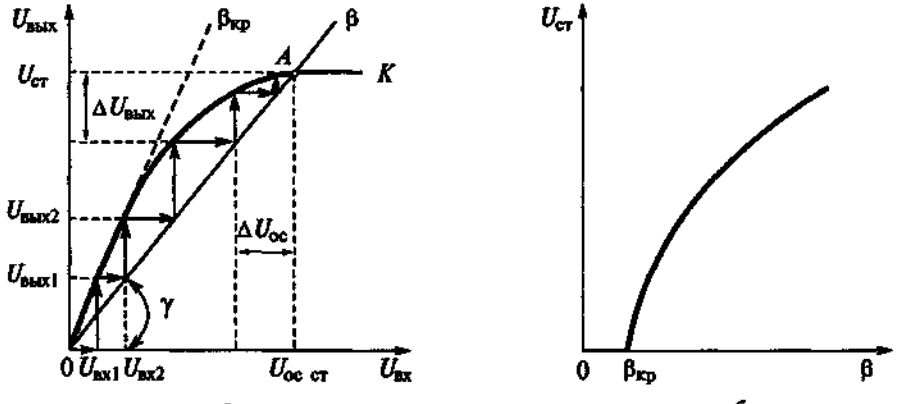



**Рис. 6.5.** Мягкий режим самовозбуждения автогенератора:
*а* — амплитудная характеристика и линия обратной связи;
*б* — зависимость амплитуды $U_{ст}$ от коэффициента передачи $\beta$

**Мягкий режим самовозбуждения.** Обратимся к типичному виду амплитудной характеристики нелинейного усилителя на ОУ (рис. 6.5, *а*). Из графика следует, что при малой амплитуде входного напряжения усилитель работает в линейном режиме и $U_{вых}/U_{вх} = K$. С ростом же амплитуды начинает проявляться нелинейность передаточной характеристики ОУ, и коэффициент усиления $K$ (а значит, и выходное напряжение) будет практически постоянным и даже может уменьшаться. На линейном участке напряжение ОС $U_{ос} = U_{вх}$ линейно связано с выходным напряжением $U_{вых}$ и определяется коэффициентом передачи цепи ПОС $\beta$ (ведь $U_{ос} = \beta U_{вых}$). Это напряжение действует на входе усилителя, поэтому линию ОС (зависимость $U_{вых}$ от $U_{ос}$) проводят на графике в виде прямой линии $\beta$ под углом $\gamma = \operatorname{arctg}(1/\beta)$ к оси абсцисс (см. рис. 6.5, *а*).

Несложный анализ графиков амплитудной характеристики и линии ОС позволяет пояснить возникновение, развитие и установление колебательного процесса в автогенераторе. Графически результат анализа показан стрелками на рис. 6.5, *а*. Положим, что на вход усилителя воздействует небольшое входное напряжение $U_{вх1}$. Тогда после усиления в $K$ раз на выходе генератора появится напряжение $U_{вых1}$. Это напряжение, ослабленное цепью положительной ОС в $\beta$ раз, поступает на вход усилителя в виде напряжения $U_{вх2}$. Оно затем, в свою очередь, усилится до напряжения $U_{вых2}$. Подобный процесс будет протекать до тех пор, пока амплитуда выходного колебания не достигнет стационарного значения, при котором выполняется условие баланса амплитуд (6.5).

Стационарную амплитуду автоколебаний генератора можно определить по координатам точки пересечения амплитудной характеристики усилителя с линией ОС (точка $A$ на рис. 6.5, *а*). Нетрудно показать, что точка $A$ является точкой устойчивого равновесия, и при случайном отклонении амплитуды выходного напряжения от стационарного значения $U_{ст}$ автогенератор всегда возвращается в исходное состояние. Допустим, что амплитуда выходного напряжения $U_{вых}$ уменьшилась относительно $U_{ст}$ на величину $\Delta U_{вых}$. Это вызовет снижение напряжения ОС $U_{ос}$ на значение $\Delta U_{ос}$, что, в соответствии с амплитудной характеристикой, в свою очередь, приведет к увеличению выходного напряжения $U_{вых}$ (см. линии со стрелками на рис. 6.5, *а*). При этом выходное напряжение будет расти до стационарного значения $U_{ст}$, а нестабильность напряжения ОС $\Delta U_{ос}$ уменьшится до нуля и перейдет в точку $U_{ос\,ст}$.

Исследуем влияние значения коэффициента передачи цепи ПОС $\beta$ на режим самовозбуждения автогенератора гармонических колебаний с типом амплитудной характеристики усилителя, показанной на рис. 6.5, *б*. Кстати заметим, что изменение значения коэффициента передачи цепи ПОС $\beta$ в схеме рис. 6.3 можно, согласно формуле (6.7), осуществлять либо регулировкой значения сопротивления резистора $R$, либо изменением *коэффициента включения* колебательного контура (неполным включением контура).

Если плавно увеличивать коэффициент передачи $\beta$ (уменьшать наклон линии $\beta$), то, начиная с некоторого критического значения $\beta_{кр}$, амплитуда стационарного колебания $U_{ст}$ будет нарастать (см. рис. 6.5). Такой режим самовозбуж-

дения генератора называют *мягким*. Для его обеспечения амплитудная характеристика усилителя должна выходить из нуля и иметь достаточно большой угол наклона к оси абсцисс в начале координат. Мягкий режим характерен тем, что подбором коэффициента передачи β можно установить любую, очень небольшую (близкую к уровню шумов), стационарную амплитуду выходных колебаний. В мягком режиме самовозбуждения на выходе генератора возникают колебания при появлении на входе усилителя малых уровней шумовых напряжений.

**Жесткий режим самовозбуждения**. Другая картина процессов наблюдается в процессах в автогенераторах, амплитудная характеристика усилителя которых имеет *S*-образную форму (рис. 6.6, *а*). Такой амплитудной характеристикой обладает усилитель при расположении его рабочей точки на нелинейном участке передаточной характеристики ОУ. Для самовозбуждения автогенераторов требуется очень сильная ПОС, и выходные колебания возникают мгновенно — скачком. Резкое («взрывное») самовозбуждение автогенератора происходит при значении коэффициента передачи цепи ОС $\beta = \beta_1$, когда линия ОС (линия *1* на рис. 6.6, *а*) касается снизу амплитудной характеристики в точке 0. Генерация колебаний срывается скачком при значении коэффициента передачи β, меньшем $\beta_2$, когда линия ОС (линия *2*) становится касательной к выпуклой части амплитудной характеристики. На графиках рис. 6.6, *а* точка *A* отражает стационарный режим работы автогенератора, а точка *C* — режим неустойчивого равновесия. Такое положение объясняют следующим образом: при амплитудах выходных колебаний автогенератора, располагающихся на графиках ниже точки *C*, колебания затухают, а при амплитудах, находящихся выше точки *C*, — будут нарастать и достигнут стационарной амплитуды в точке *A*.

Чтобы вновь возбудить автогенератор, необходимо опять увеличить коэффициент передачи цепи ПОС β до значения $\beta_1$, соответствующего положению линии *1*. Такой режим самовозбуждения автогенератора называют *жестким*. На рис. 6.6, *б* показана зависимость стационарной амплитуды $U_{ст}$ от коэффициента передачи цепи положительной ОС, причем стрелками показано направление изменения β. Жесткий режим самовозбуждения автогенератора неэффективен, считается нежелательным и применяется крайне редко.

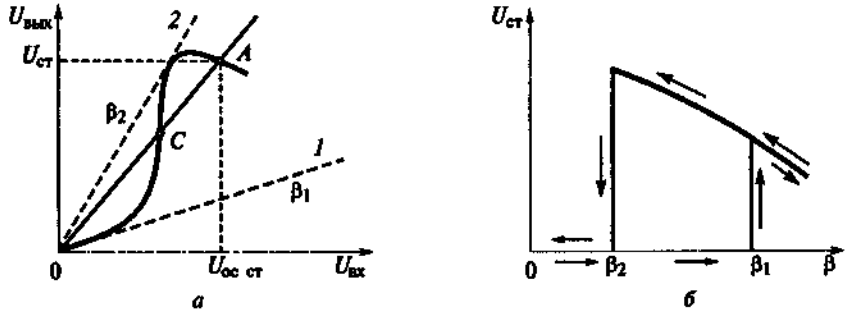



**Рис. 6.6. Жесткий режим самовозбуждения автогенератора:**
*а* — амплитудная характеристика и линия обратной связи;
*б* — зависимость амплитуды $U_{ст}$ от коэффициента передачи β

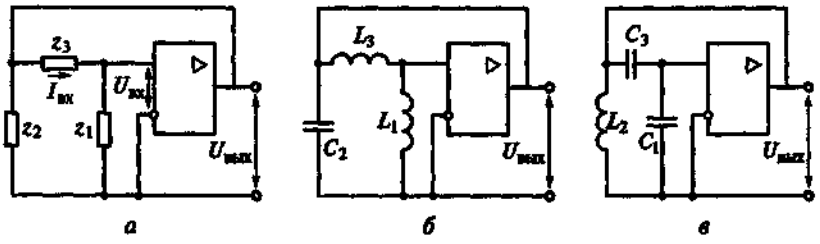



**Рис. 6.7.** Схемы трехточечных автогенераторов:
*а* — структурная; *б* — индуктивная трехточка; *в* — емкостная трехточка

### Трехточечные схемы автогенераторов гармонических колебаний

В большинстве схем *LC*-генераторов напряжение ОС снимают с части колебательного контура, т. е. используют неполное его включение. Поскольку колебательный контур при этом имеет три точки соединений (выводов), то из-за этого схемы *LC*-генераторов (рис. 6.7) получили название *трехточечных*. Упрощенная структурная схема трехточечного автогенератора, созданного на неинвертирующем усилителе, представлена на рис. 6.7, *а*. В схеме реактивные элементы $z_1$, $z_2$ и $z_3$ образуют колебательную систему, с помощью которой создают частотно-зависимую ПОС. В реальных схемах автогенераторов индуктивности и емкости контуров имеют малые омические потери, поэтому при анализе можно учитывать только их реактивные сопротивления $x_1$, $x_2$ и $x_3$.

Положив, что входное сопротивление усилителя много больше модуля сопротивления элемента $z_1$, находим ток $\mathbf{I}_{вх} = \mathbf{U}_{вых}/(z_1 + z_3)$. Тогда напряжение ОС

$$\mathbf{U}_{ос} = \mathbf{U}_{вх} = I_{вх}z_1 = \mathbf{U}_{вых}z_1/(z_1 + z_3) \tag{6.15}$$

и коэффициент передачи петли положительной ОС

$$\beta = \frac{\mathbf{U}_{ос}}{\mathbf{U}_{вых}} = \frac{\mathbf{U}_{вх}}{\mathbf{U}_{вых}} = \frac{z_1}{z_1 + z_3} = \frac{x_1}{x_1 + x_3}. \tag{6.16}$$

Так как ОУ в автогенераторе включен по неинвертирующей схеме, то на резонансной частоте, для которой полное реактивное сопротивление контура

$$x_1 + x_2 + x_3 = 0, \tag{6.17}$$

усилительный каскад не вносит в выходной сигнал фазовый сдвиг и угол $\varphi_к = 0$.

Условие баланса фаз (6.6) выполняется в том случае, когда и цепь ПОС не вносит фазового сдвига ($\varphi_\beta = 0$). Это обеспечивается, если $x_1$ и $x_3$ — реактивные сопротивления одного знака, т. е. либо оба элемента емкости, либо — индуктивности. При этом для выполнения равенства (6.17) сопротивление $x_2$ должно иметь знак, противоположный знаку сопротивлений $x_1$ и $x_3$. Варианты построения схем трехточечных генераторов показаны на рис. 6.7, *б*, *в*.

В схеме *LC*-генератора (см. рис. 6.7, *б*), известной в радиотехнике под названием *индуктивной трехточки*, индуктивная ветвь параллельного контура содержит индуктивности $L_1$ и $L_3$. Напряжение ОС с индуктивности $L_1$ поступает

на неинвертирующий вход ОУ. Поскольку знаки мгновенных значений напряжений на индуктивностях $L_1$ и $L_3$ относительно их общей точки соединения одинаковы, то ОС будет положительной. В емкостной ветви колебательного контура включен конденсатор $C_2$. По аналогии с индуктивной трехточкой можно построить автогенератор с *емкостной трехточкой* (рис. 6.7, *в*). В данной схеме емкостная ветвь колебательного контура содержит конденсаторы $C_1$ и $C_3$. Напряжение ОС с конденсатора $C_1$ поступает на неинвертирующий вход усилителя. При таком включении конденсаторов в колебательный контур, полярности мгновенных значений напряжений на их обкладках относительно общей точки соединения одинаковы, поэтому в схеме ОС — положительная.

Широкое применение в радиотехнике находят генераторы, в которых операционные усилители включены по инвертирующей схеме. При этом достаточно в схеме рис. 6.7, *б*, *в* поменять местами индуктивность $L_1$ и емкость $C_1$.

Настройку *LC*-генератора на требуемую частоту осуществляют обычно изменением небольшой емкости дополнительного конденсатора, включенного параллельно основному конденсатору контура. В современных автогенераторах для изменения частоты колебаний применяют варикапы и подстройку (или даже перестройку) частоты производят электронным способом.

## 6.3. *RC*-генераторы

Характеристики *LC*-генераторов в диапазонах низких частот (килогерцы и ниже) существенно ухудшаются из-за резкого возрастания значений индуктивностей и емкостей колебательных контуров и соответствующих им размеров индуктивностей и конденсаторов. Поэтому в низкочастотных генераторах в качестве колебательных систем и цепей положительной обратной связи используют частотно-избирательные *RC*-цепи.

Практически все современные *RC*-генераторы малой и средней мощностей (10...15 Вт) строятся на ОУ. Напряжение положительной ОС в *RC*-генераторах на ОУ можно подавать как на инвертирующий, так и на неинвертирующий входы. В схемах *RC*-генераторов с неинвертирующим включением ОУ частотно-избирательная цепь ПОС не должна вносить фазового сдвига в выходной сигнал. В *RC*-генераторах с инвертирующим включением ОУ, наоборот, *RC*-цепь положительной обратной связи на частоте генерации должна сдвигать фазу выходных колебаний на угол $\varphi_\beta = \pi$.

Наиболее распространены в радиоэлектронике низкочастотные автогенераторы двух видов — с трехзвенной фазосдвигающей *RC*-цепью и с мостом Вина.

### Автогенератор с фазосдвигающей *RC*-цепью

Такой генератор содержит инвертирующий усилитель и трехзвенную (иногда, лестничную) *RC*-цепь положительной ОС (рис. 6.8, *а*). Из курса теории цепей известно, что данная трехзвенная *RC*-цепь (ее называют *R-параллель*) имеет частотную $\beta(f)$ и фазовую $\varphi_\beta(f)$ характеристики, показанные на рис. 6.8, *б*.

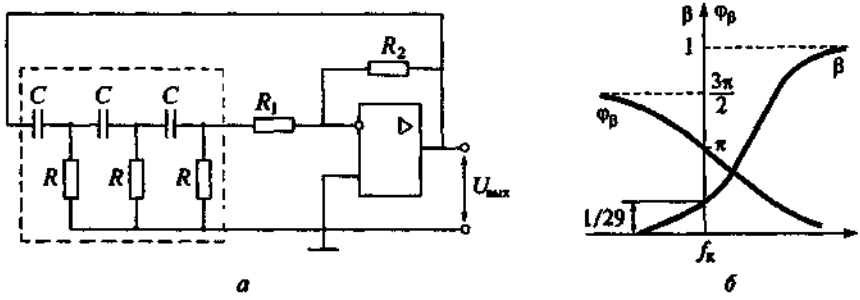



**Рис. 6.8.** Автогенератор с трехзвенной *RC*-цепью:
*а* — схема; *б* — частотная и фазовая характеристики цепи

Анализ частотной и фазовой характеристик показывают, что на *квазирезонансной* (напомним, якобы резонансной) циклической частоте генерации $f_к$ трехзвенная *RC*-цепь положительной ОС имеет вещественное значение коэффициента передачи $\beta = 1/29$ и вносит фазовый сдвиг $\varphi_\beta = \pi$. Поэтому для обеспечения в генераторе баланса амплитуд (6.5) необходимо выбирать коэффициент усиления усилителя $|K| = R_2/R_1 \geq 29$, а баланс фаз (6.6) обеспечивается автоматически инвертирующим включением ОУ. Можно показать, что квазирезонансная частота генерации для схемы с идеальным ОУ (см. рис. 6.8, *а*) определяется формулой

$$f_к = \frac{1}{2\pi\sqrt{6}RC}. \qquad (6.18)$$

Как следует из этой формулы, частота генерации зависит только от параметров цепи ОС *R* и *C* (т. е. внешних элементов усилителя).

**Пример 6.3.** Для частоты автоколебаний $f_к = 250$ Гц и сопротивления резистора $R = 25$ кОм найти емкость конденсатора в генераторе с трехзвенной *RC*-цепью.

**Р е ш е н и е.** Воспользовавшись формулой (6.18), находим, что необходим конденсатор емкостью $C = 2{,}5$ мкФ.

Если в схеме *RC*-генератора (см. рис. 6.8, *а*) поменять местами резисторы и конденсаторы в трехзвенной цепи, то квазирезонансная частота

$$f_к = \frac{\sqrt{6}}{2\pi RC}. \qquad (6.19)$$

При этом необходимо обеспечить коэффициент усиления собственно усилителя $K \geq 18{,}4$, поскольку коэффициент передачи такой цепи $\beta \approx 0{,}056$.

Недостатки *RC*-генератора, построенном на инвертирующем усилителе и трехзвенной *RC*-цепью, — довольно большое количество элементов в петле положительной ОС и, как следствие, трудность перестройки частоты генерации в широком диапазоне. Поэтому чаще применяют автогенераторы с мостом Вина в цепи положительной обратной связи.

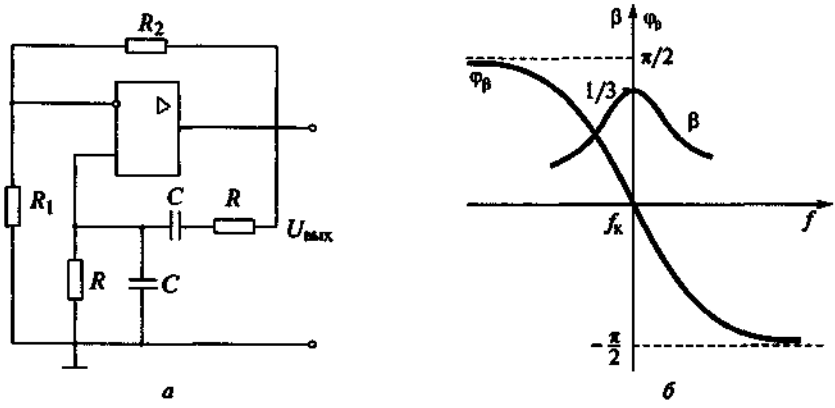



**Рис. 6.9.** *RC*-генератор с мостом Вина:
*а* — схема; *б* — частотная и фазовая характеристики моста

### *RC*-генератор с мостом Вина

Схемы данных автогенераторов с *RC*-цепями имеют более компактную структуру. В них цепь ПОС (мост Вина) включается между выходом и неинвертирующим входом ОУ (рис. 6.9, *а*). *Мост Вина* представляет собой частотно-избирательную последовательно-параллельную *RC*-цепь, состоящую из двух пар емкостей $C$ и сопротивлений $R$. Частотная $\beta(f)$ и фазовая $\varphi_\beta(f)$ характеристики моста Вина (рис. 6.9, *б*) давно известны и изучаются в курсе основ теории цепей. Из графика частотной характеристики следует, что на квазирезонансной частоте коэффициент передачи моста Вина $\beta = 1/3$. Это значит, что самовозбуждение автогенератора обеспечивается при коэффициенте усиления усилителя $|K| = R_2/R_1 \geq 3$.

Фазовый сдвиг в выходном сигнале автогенератора отсутствует ($\varphi_\beta = 0$) на квазирезонансной частоте

$$f_к = \frac{1}{2\pi RC}. \qquad (6.20)$$

Перестройку частоты выходных колебаний автогенератора осуществляют с помощью сдвоенного конденсатора.

**Пример 6.4.** Циклическая частота колебаний в низкочастотном *RC*-генераторе с мостом Вина $f_к = 500$ Гц. Определите параметры схемы, обеспечивающие требуемую частоту.

Р е ш е н и е. Зададим сопротивление резисторов моста Вина $R = 3{,}9$ кОм. Тогда, согласно соотношению (6.20), емкости моста $C = 1/(2\pi f_к R) \approx 0{,}08$ мкФ. Поскольку в автогенераторе с мостом Вина $|K| = R_2/R_1 \geq 3$, то, выбрав сопротивление резистора $R_2 = 30$ кОм, находим $R_1 = 10$ кОм.

## 6.4. Стабилизация частоты в автогенераторах

Наиболее важным требованием, предъявляемым к автогенераторам, является высокая стабильность частоты выходных колебаний. Это связано с тем, что во время работы генератора частота колебаний меняется в некоторых пределах по случайному закону под воздействием различных дестабилизирующих факторов: изменений температуры, влажности и напряжения питания, наличия внешних электромагнитных полей, механических воздействий и пр. Влияние дестабилизирующих факторов проявляется в изменениях значений индуктивностей катушек, емкостей конденсаторов и сопротивлений резисторов, входящих в состав колебательных контуров и частотно-избирательных *RC*-цепей.

В отмеченном смысле качество работы автогенераторов принято оценивать *абсолютной и относительной нестабильностями частоты*. Абсолютная нестабильность представляет собой разность $\Delta f$ между текущим $f$ и номинальным $f_р$ (или $f_к$) значениями частоты колебаний. Относительную нестабильность частоты определяют *коэффициентом нестабильности* $\Delta f/f_р$ или $\Delta f/f_к$.

При расчете коэффициентов нестабильности используют такие формулы:

• *для LC-генераторов*

$$\Delta f/f_р = -0{,}5(\Delta L/L + \Delta C/C); \qquad (6.21)$$

• *для RC-генераторов*

$$\Delta f/f_к = -(\Delta C/C + \Delta R/R). \qquad (6.22)$$

Здесь $\Delta L$, $\Delta C$ и $\Delta R$ — значения изменений индуктивностей катушек, емкостей конденсаторов и сопротивлений резисторов от номинальных значений, вызванные воздействием любого из дестабилизирующих факторов.

Известны такие способы стабилизации частоты колебаний генераторов:

- параметрическая с использованием обычных колебательных систем;
- кварцевая с использованием в качестве резонатора кристалла кварца;
- на основе диэлектрического резонатора (применяемая в СВЧ-диапазоне);
- молекулярная за счет индуцированного возбуждения атомов, находящихся на высоком энергетическом уровне. Молекулярные генераторы, работающие в диапазоне более 10 ГГц, используют как эталоны частоты.

В схемах генераторов гармонических (часто и импульсных) колебаний применяют два способа стабилизации частоты: параметрический и кварцевый.

### Параметрический способ стабилизации частоты

Данный способ стабилизации частоты заключается в ослаблении влияния дестабилизирующих факторов и подборе высокочастотных и прецизионных элементов колебательных контуров автогенераторов. Для исключения влияния температуры на параметры усилительных элементов автогенераторы в отдельных случаях помещают в термостаты. Уменьшение влияния механических воздействий обеспечивают применением печатного монтажа и проводов индуктив-

ностей, вжигаемых в керамику. Параметрическая стабилизация позволяет снизить нестабильность до $10^{-5}$ (уход частоты на 10 Гц при колебаниях в 1 МГц).

### Кварцевая стабилизация частоты

Данный способ стабилизации (кварцевая стабилизация частоты ламповых генераторов разработана в 1918 г. У. Кэди, США) основан на применении в электрических схемах вместо *LC*-контуров или их элементов кварцевого резонатора, что позволяет снизить нестабильность частоты колебаний автогенератора до $10^{-7}$ (отклонение частоты на $\Delta f = 0{,}1$ Гц от генерируемой в $f_0 = 1$ МГц).

Кварцевый резонатор (сокращенно *кварц*) представляет собой помещенную в кварцедержатель тонкую прямоугольную пластинку минерала кварца, грани которой определенным образом ориентированы по отношению к осям кристалла. Известно, что кварц обладает прямым и обратным *пьезоэлектрическим эффектом* (проще, *пьезоэффектом*). Прямой пьезоэффект возникает при механическом сжатии или растяжении кварцевой пластинки и сопровождается появлением на ее противоположных гранях электрических зарядов.

При воздействии на кварцевую пластинку переменного электрического поля в ней возникают упругие механические колебания (*обратный пьезоэффект*), приводящие, в свою очередь, к появлению электрических зарядов на ее гранях. Кварц можно рассматривать как электромеханическую колебательную систему и сравнивают ее свойства с обычным колебательным *LC*-контуром (рис. 6.10). Добротность кварцевого резонатора достигает сотен тысяч, тогда как у колебательного контура она не превышает 300...400. Механическая прочность и слабая зависимость частотных свойств от температуры обусловливают достаточно высокую эталонность частоты кварцевых резонаторов.

При расчетах кварцевый резонатор представляют эквивалентной схемой (рис. 6.10, *а*), в которой элементы $L_{кв}$, $C_{кв}$ и $R_{кв}$ характеризуют соответственно индуктивность, емкость и омические потери кварца. Емкость $C_{ок}$ отражает наличие кварцедержателя. Зависимость реактивного сопротивления кварцевого резонатора от частоты $x(f)$ показана на рис. 6.10, *б*. Она имеет два резонанса: последовательный на частоте $f_{к1}$ и параллельный на частоте $f_{к2}$.

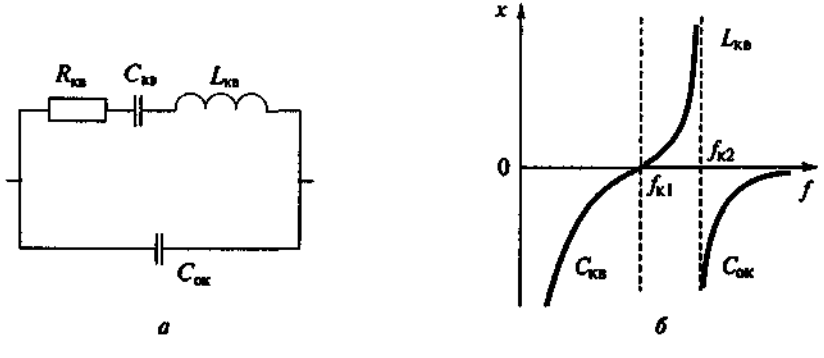



**Рис. 6.10.** Кварцевый резонатор:
*а* — эквивалентная схема; *б* — зависимость реактивного сопротивления от частоты

*Последовательный резонанс* создают $L_{кв}$ и $C_{кв}$, задающие резонансную частоту

$$f_{к1} = \frac{1}{2\pi\sqrt{L_{кв}C_{кв}}}. \qquad (6.23)$$

*Параллельный резонанс* в устройствах с кварцем почти не используется.

**Схемы кварцевых автогенераторов.** В практической радиотехнике кварц часто применяют в *LC*-генераторах в качестве индуктивности (рис. 6.11, *а*), что упрощает конструкцию, а также уменьшает мощность, рассеиваемую в резонаторе. Условия возникновения гармонических колебаний в автогенераторе можно проанализировать, заменив кварцевый резонатор (Кв) его эквивалентной схемой и применив общие уравнения, характеризующие условия самовозбуждения автогенератора.

На рис. 6.11, *б* изображена упрощенная схема *RC*-генератора с мостом Вина, в котором в качестве одного из резисторов включен кварцевый резонатор, работающий в режиме резонанса напряжений. Для того чтобы резонансная частота кварца совпадала с квазирезонансной частотой моста Вина, сопротивление резистора $R$ подбирают равным резонансному активному сопротивлению кварца $R_{кв}$. Цепь отрицательной ОС с терморезистором $R_2$, включенная между выходом и инвертирующим входом ОУ, компенсирует температурные изменения резонансного сопротивления кварца, стабилизируя тем самым амплитуду выходных колебаний автогенератора. Отметим, что подстройку частоты в принципиальной электрической схеме данного автогенератора осуществляют с помощью конденсаторов. Для этого обычно используют полупроводниковую емкость — варикап.

В последние годы большое распространение получили стабилизированные генераторы сигналов (т. н. осцилляторы), в которых в цепь ПОС транзисторного усилителя включен резонатор на ПАВ. Такие генераторы достаточно просты, малогабаритны и работают в диапазоне частот от 20 МГц до нескольких ГГц. В них возможна электронная перестройка частоты, или частотная модуляция.

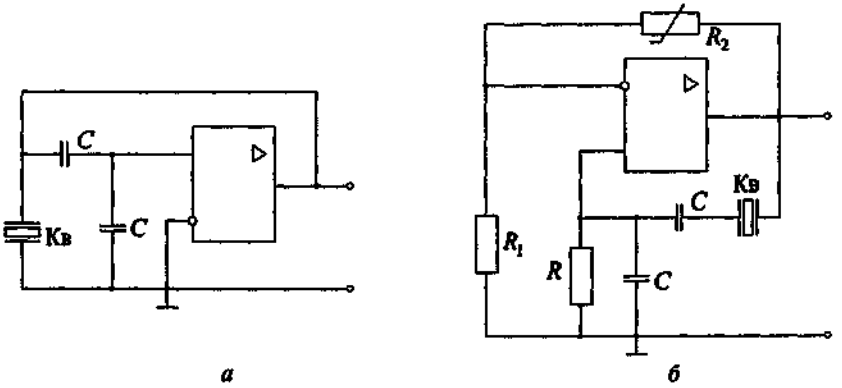



**Рис. 6.11.** Схемы кварцевых автогенераторов:
*а* — кварц — аналог индуктивности;
*б* — кварц включен в мост Вина в качестве сопротивления

## 6.5. Цифровые генераторы

**Цифровые генераторы низких частот** отличаются от аналоговых более эффективными метрологическими характеристиками: высокими точностью установки и стабильностью частоты, малым коэффициентом нелинейных искажений (строго синусоидальной формой), постоянством уровня выходного сигнала. Цифровые генераторы удобнее аналоговых: выше быстродействие, проще установка требуемой частоты, более наглядная индикация, возможна автоматическая перестройка частоты по заданной программе.

Действие цифровых генераторов низких частот основано на формировании числового кода с превращением его в гармоническое колебание. При этом используют метод аппроксимации формы выходного колебания. Простейший вид аппроксимации — ступенчатая. Ее суть заключается в представлении синусоиды напряжением ступенчатой формы, мало отличающейся от синусоидальной кривой (рис. 6.12, *а*). Аппроксимирую синусоиду $u(t) = U_m \sin\omega t$ дискретизируют, и в интервале, разделяющем два соседних момента времени $t_i$ и $t_{i+1}$, ее заменяют постоянным напряжением — ступенькой, высота которой равна значению аппроксимируемого напряжения в момент $t_i$, т. е. $u(t_i) = U_m \sin\omega t_i$. В результате такой замены вместо синусоиды получают ступенчатую линию (рис. 6.12, *а*.)

При заданном периоде $T$ гармонического колебания число ступенек $p$, приходящихся на один период, определяется шагом дискретизации $p = T/\Delta t$. Если же из технических соображений число ступенек задано, то изменение шага дискретизации приводит к изменению периода формируемого напряжения, поскольку период $T = p\Delta t$. Учитывая, что $t_i = i\Delta t$, уравнение ступенчатой кривой будет иметь вид $u(i\Delta t) = U_m \sin(i\omega\Delta t)$, или с учетом $p$ и $\omega = 2\pi/T$

$$u(i\Delta t) = U_m \sin(i2\pi/p).$$

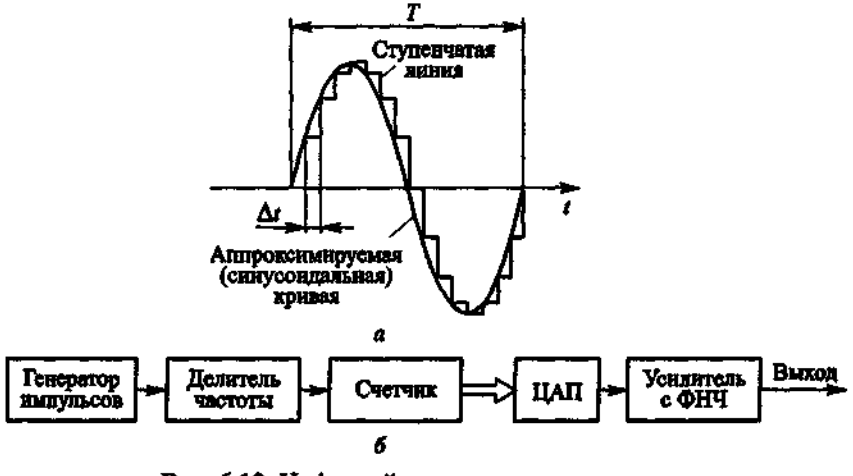



**Рис. 6.12.** Цифровой генератор низких частот:
*а* — ступенчатая аппроксимация синусоиды; *б* — упрощенная структурная схема

Ступенчатая кривая тем точнее приближается по форме к синусоиде, чем больше выбрано число ступеней $p$. Когда это число велико, ступенчатое напряжение можно рассматривать как синусоидальное, немного искаженное высокочастотной помехой. Спектральный анализ напряжения, полученного при ступенчатой аппроксимации, показывает, что его спектр содержит гармонику основной частоты и ряд высших гармоник. При этом оказывается, что ближней к основной гармонике высшей гармоникой будет составляющая с номером $(p-1)$, следующей — с номером $(p+1)$, затем гармоники с номерами $(2p-1)$ и $(2p+1)$ и т. д. Например, при $p = 25$ и частоте напряжения $f$ основной гармоники ближайшими высшими гармониками будут 24-, 26-, 49-, 51-я гармоники, т. е. напряжения частот $24f$, $26f$, $49f$, $51f$. Такие соотношения между основной и высшими гармониками позволяют осуществить качественную фильтрацию и получить синусоидальное напряжение с малыми нелинейными искажениями.

Упрощенная структурная схема цифрового генератора низких частот показана на рис. 6.12, *б*. Генератор импульсов вырабатывает периодическую последовательность коротких импульсов с периодом следования $T_0$. На выходе делителя частоты с регулируемым коэффициентом деления $g$ получают периодическую последовательность импульсов с периодом следования $\Delta t = gT_0$, задающим шаг дискретизации, которые поступают затем в счетчик импульсов емкостью $p$.

Кодовая комбинация, определяемая числом $i$ импульсов, накопленных в счетчике, передается в ЦАП, формирующим сигнал, соответствующий числу $i$, т. е. $u(i\Delta t) = U_m \sin(i2\pi/p)$. Таким образом создается $p$ ступенек аппроксимируемой кривой. При накоплении $p$ импульсов счетчик переполняется и сбрасывается в нуль. С приходом $(p+1)$-го импульса начинается формирование нового периода ступенчатой кривой. Частоту формируемого колебания при фиксированном числе ступенек аппроксимации $p$ регулируют, меняя шаг дискретизации $\Delta t$, что достигается изменением коэффициента деления $g$ делителя частоты.

**Цифровые генераторы *M*-последовательностей**. Схемотехнически цифровую $M$-последовательность формируют $m$-разрядным двоичным регистром сдвига, который охвачен обратной связью через сумматор, осуществляющим *сложение по модулю 2* (способы введения ОС известны и табулированы).

На рис. 6.13 в качестве примера показан четырехразрядный сдвигающий регистр, который формирует $M$-последовательность, содержащую 15 элементов в одном периоде сигнала.

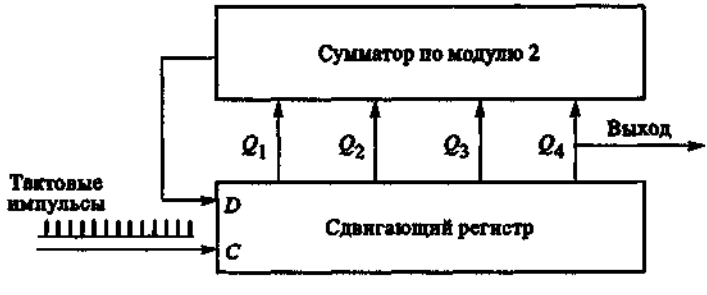



**Рис. 6.13.** Генератор 15-элементной $M$-последовательности

Регистр содержит четыре последовательно соединенных триггерных ячейки, прямые выходы которых обозначены соответственно $Q_1 - Q_4$. Выход регистра $Q_4$ служит выходом генератора. Сдвигающий (тактовый) импульс, подводимый к входу $C$ регистра, подается на все триггеры одновременно. Если триггер $i$-го разряда находится в положении 0 (напомним, что при этом 0 на выходе $Q_i$), то импульс не воздействует на него. Если триггер пребывает в состоянии 1, сдвигающий импульс переключает его в положение 0, в результате чего кодовая 1 записывается в триггер $(i + 1)$-го разряда. Тактовые импульсы на сдвигающий регистр поступают непрерывно с частотой $F_c$ (интервалом $\Delta t$) и каждый новый импульс продвигает единицу дальше по направлению к выходу регистра. Когда на выходе сумматора по модулю 2 появляется единица, она поступает на вход $D$ регистра и записывается в первую триггерную ячейку. Очередной входной тактовый импульс сдвигает единицу в следующую ячейку и т. д.

Положим, что в начальном состоянии в регистре записано число 1 000. Его считаем 1-й комбинацией единиц и нулей. С приходом первого сдвигающего импульса это число изменится — получится 2-я комбинация — 1 100, и т. д. После 15-й комбинации образуется вновь 1-я и число на выходе регистра — 1 000. В табл. 6.1 приведены 15 комбинаций (в скобках указана цифра, получающаяся на выходе сумматора по модулю 2). Итак, на выходе генератора, т. е. выходе $Q_4$, за один цикл (период $T$) образуется последовательность нулей и единиц (см. последнюю цифру каждой комбинации): 000111101011001.

Таблица 6.1. **Комбинации чисел в схеме генератора**

| Номер комбинации | Число | Номер комбинации | Число | Номер комбинации | Число |
|---|---|---|---|---|---|
| 1 | 1000 (1) | 6 | 1011 (0) | 11 | 0011 (1) |
| 2 | 1100 (0) | 7 | 0101 (1) | 12 | 1001 (0) |
| 3 | 1110 (1) | 8 | 1010 (1) | 13 | 0100 (0) |
| 4 | 1111 (0) | 9 | 1101 (0) | 14 | 0010 (0) |
| 5 | 0111 (1) | 10 | 0110 (0) | 15 | 0001 (1) |

Из параметров цифрового генератора и характеристик спектральной плотности $M$-последовательности несложно определить, что высота спектральных линий в полосе частот $0{,}05F_c$ практически постоянна: уменьшение мощности на частоте $0{,}05\ F_c$ составляет всего 0,036 дБ. Изменение мощности на – 0,1 дБ имеет место на частоте $0{,}085\ F_c$, а изменение ее на – 3 дБ — на частоте $0{,}45F_c$. Так, если сдвигающий регистр содержит 20 разрядов, т. е. наибольшая длина последовательности $M = 2^m - 1 = 1048575$, то при частоте следования сдвигающих импульсов $F_c$ = 1 МГц полоса «постоянной мощности» составляет 50 кГц, а расстояние между спектральными линиями равно 1 Гц. Полоса, в которой мощность изменяется на 3 дБ, составляет 450 кГц.

Изменение частоты следования сдвигающих импульсов $F_c$ при постоянной длине $M$-последовательности ($M$ = const) пропорционально изменяет расстояние

между спектральными линиями, но не меняет числа линий в лепестке или общей мощности сигнала. Спектральная плотность мощности обратно пропорциональна частоте $F_c$. Изменение длины последовательности $M$ при $F_c$ = const сопровождается обратно пропорциональным изменением расстояния между спектральными линиями. Соответственно изменяется число линий в лепестке, но общая мощность генерируемого сигнала и распределение ее по боковым лепесткам сохраняются. Значит, высоты спектральных линий изменяются так, что спектральная плотность мощности сигнала всегда постоянна.

## 6.6. Генераторы сверхвысоких частот

Генераторы сверхвысоких частот (СВЧ-генераторы) применяют в диапазоне 1...300 ГГц и выполняют на полупроводниковых генераторных диодах, отражательных клистронах, магнетронах, лампах бегущей и обратной волны (ЛОВ). Отдельную группу СВЧ-генераторов составляют оптические квантовые генераторы. Рассмотрим применяемые в передатчиках систем связи полупроводниковые СВЧ-генераторы и исследуем их характеристики и параметры.

### Автогенераторы на элементах с падающим участком ВАХ

Если в дециметровом и длинноволновой части сантиметрового диапазонов волн используют рассмотренные транзисторные автогенераторы, то в короткой части сантиметрового и миллиметровом диапазонах широко применяются диодные генераторы. СВЧ-диоды, используемые в таких автогенераторах, имеют вольт-амперную характеристику с падающим участком, и поэтому на рабочих частотах обладают отрицательным сопротивлением. К генераторным диодам относятся *туннельный диод* (ТД; создан в 1957 г. лауреатом Нобелевской премии японским физиком Лео Эсаки — *Leo Esaki*), *лавинно-пролетный диод* (ЛПД; разработан в 1964 г. лауреатом Ленинской премии А.С. Тагером совместно с А.И. Мельниковым, В.М. Вальд–Перловым и Ю.К. Пожелой) и *диод Ганна* (ДГ; создан в 1963 г. американским физиком Джона Ганном — *John Gunn*,).

В настоящее время разработан новый твердотельный прибор «Quantum Welle-диод» на основе полупроводниковой структуры с «квантовыми ямами», обладающий отрицательным сопротивлением в широком диапазоне частот.



Л. Эсаки А.С. Тагер Д. Ганн

Проанализируем принцип действия СВЧ-генератора на туннельном диоде. *Туннельным диодом* называют полупроводниковый диод, в котором имеется очень узкий потенциальный барьер, препятствующий движению электронов, и поэтому на прямом участке его ВАХ образуется *n*-образная зависимость тока от напряжения.

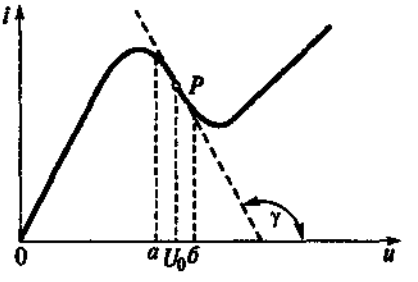



**Рис. 6.14.** ВАХ туннельного диода

На рис. 6.14 показана вольт-амперная характеристика туннельного диода, представляющая зависимость его прямого тока от положительного напряжения смещения. На падающем участке *а — б* дифференциальное сопротивление туннельного диода отрицательно

$$R_{-} = \frac{du}{di} = \operatorname{ctg}\gamma, \tag{6.24}$$

где γ — угол наклона касательной к кривой ВАХ $i = f(u)$ в рабочей точке *P*, когда рабочее напряжение $u = U_0$.

При подключении туннельного диода с подобной ВАХ к колебательной цепи возможна генерация высокочастотных колебаний и полученная схема превращается в *автогенератор с внутренней обратной связью* (рис. 6.15).

Упрощенная электрическая схема генератора на туннельном диоде показана на рис. 6.15, *а*. Колебательная цепь состоит из индуктивности *L* и собственной емкости диода $C_0$; $r_к$ — эквивалентное сопротивление потерь в диоде и элементах контура; $E_0$ — напряжение смещения на диоде. Блокировочные дроссель $L_{бл}$ и конденсатор $C_{бл}$ ($C_{бл} << C_0$) препятствуют протеканию тока высокой частоты через цепи постоянного тока. Схема замещения колебательного контура показана на рис. 6.15, *б*. Контур шунтирован отрицательным дифференциальным сопротивлением туннельного диода, по отношению к которому переменное напряжение $u_к$, действующее на контуре, можно рассматривать как ЭДС. В результате этого действия через диод протекает ток $i_{VD} = -u_к/R_{-}$.

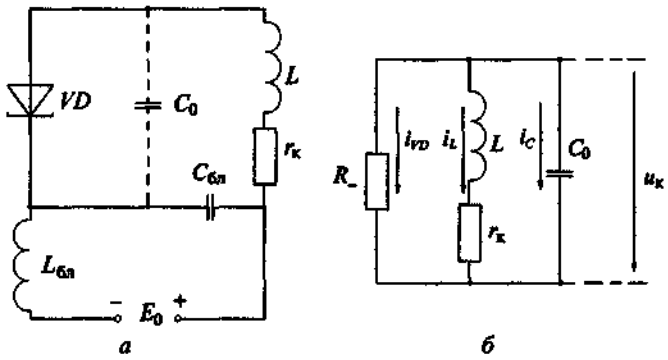



**Рис. 6.15.** Генератор на туннельном диоде:
*а* — упрощенная схема; *б* — схема замещения

Токи в ветвях контура $i_{VD}$, $i_C$, $i_L$ и колебательное напряжение $u_к$ связаны между собой соотношениями (в них аргумент $t$ для упрощения опущен), определяемыми первым и вторым законами Кирхгофа и законом Ома:

$$i_{VD} = i_L + i_C;$$

$$i_C = C_0 \frac{du_к}{dt}; \tag{6.25}$$

$$u_к = r_к i_L + L\frac{di_L}{dt}.$$

В качестве искомой можно выбрать любую из этих функций. Пусть определяется ток $i_L$ в индуктивной ветви колебательного контура. Исключив из первого уравнения (6.25) ток $i_C$, с помощью второго и третьего соотношений находим

$$i_{VD} = i_L + r_к C_0 \frac{di_L}{dt} + LC_0 \frac{d^2 i_L}{dt^2}. \tag{6.26}$$

В то же время ток диода

$$i_{VD} = -\frac{1}{R_-} u_к = -\frac{1}{R_-}\left(r_к i_L + L\frac{di_L}{dt}\right). \tag{6.27}$$

Приравняв токи диода в двух последних выражениях, после группировки слагаемых и несложных выкладок, получим дифференциальное уравнение

$$\frac{d^2 i_L}{dt^2} + \left(\frac{r_к}{L} + \frac{1}{C_0 R_-}\right)\frac{di_L}{dt} + \frac{r_к + R_-}{LC_0 R_-} i_L = 0. \tag{6.28}$$

Определим условия возникновения, нарастания и затухания колебаний в генераторе. Чтобы в колебательной системе возникли высокочастотные колебания и их амплитуда нарастала, коэффициент при первой производной в (6.28) должен быть отрицательным. Отсюда находим условие возникновения колебаний

$$\frac{r_к}{L} - \frac{1}{C_0}\left|\frac{1}{R_-}\right| < 0,$$

или

$$|R_-| < \frac{L}{r_к C_0} = R_0 = \rho Q = \frac{Q}{\omega_р C_0}. \tag{6.29}$$

В двух последних соотношениях $|R_-|$ — абсолютное значение отрицательного сопротивления туннельного диода, зависящая от амплитуды колебаний (рис. 6.16); $R_0$ — резонансное сопротивление, $Q$ — добротность, $\rho = \sqrt{L/C_0}$ — характеристическое сопротивление колебательного контура.

Когда сопротивление туннельного диода $|R_-| = |R_-(U_к)|$ увеличится до

$$|R_-(U_к)| = |R_-(U_{к\,ст})| = R_0, \tag{6.30}$$

в автогенераторе установится стационарная амплитуда колебаний.

Стационарный режим колебаний будет устойчив, если в точке пересечения горизонтальной линии $R_0$ кривая сопротивления $|R_-(U_к)|$ имеет положительный наклон.

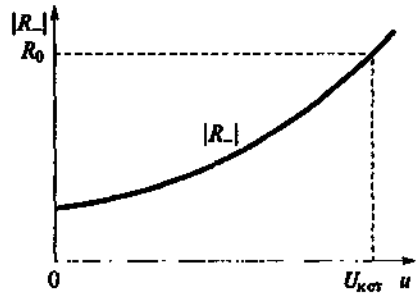



**Рис. 6.16.** К определению стационарной амплитуды в генераторе на туннельном диоде

С физической точки зрения работу такого генератора можно трактовать следующим образом: отрицательное сопротивление $R_-$ туннельного диода вносит в контур генератора больше энергии, чем выделяется энергии в сопротивлении нагрузки, которым в данном случае является сопротивление $r_к$. Туннельный диод — маломощный прибор, через который в рабочем режиме при выборе рабочей точки на падающем участке ВАХ протекает ток порядка $I_0 = 1$ мА. При этом генерируемая мощность, выделяющаяся в нагрузке, составляет 20...30 мкВт. Однако в ряде случаев этой мощности достаточно и приходится создавать генераторы с усилением мощности.

## Контрольные вопросы

1. По каким признакам классифицируют радиотехнические сигналы?
2. Какие сигналы относятся к аналоговым, дискретным и цифровым?
3. В чем заключено основное отличие детерминированных и случайных сигналов?
4. Что собой представляют шумы и помехи?
5. Какие сигналы являются ортогональными?
6. К каким устройствам относят автоколебательные системы?
7. На какие основные типы делят автогенераторы?
8. Приведите структурную схему автогенератора.
9. Чем отличаются условия самовозбуждения автогенератора от условий его стационарного режима?
10. Представьте упрощенную схему *LC*-генератора на операционном усилителе.
11. Каким образом обеспечиваются баланс амплитуд и фаз в *LC*-генераторе?
12. С помощью какого метода анализируются генераторы гармонических колебаний?
13. В чем состоит принципиальное отличие мягкого и жесткого режимов самовозбуждения автогенераторов?
14. Приведите структурные схемы трехточечных автогенераторов.
15. Почему на относительно низких частотах выгоднее применять *RC*-генераторы?
16. Приведите схемы *RC*-генераторов с трехзвенной *RC*-цепью и с мостом Вина; дайте их сравнительный анализ.
17. Какие методы стабилизации частоты применяются в схемах автогенераторов?
18. В чем состоит принцип действия кварцевого резонатора и какие его свойства используются в автогенераторах?
19. Почему кварцевые генераторы имеют высокую стабильность частоты?
20. На каком принципе построено действие цифровых генераторов низких частот?
21. Поясните принцип действия СВЧ-генератора на туннельном диоде.

# Глава 7. ИМПУЛЬСНЫЕ И ЦИФРОВЫЕ УСТРОЙСТВА

В радиотехнических системах, технике связи, телевидении и т. д. широко используют импульсные и цифровые устройства. К импульсным относят устройства, работающие в прерывистом (импульсном) режиме. При этом кратковременное воздействие чередуется с паузой.

Импульсная и цифровая техника базируются на использовании последовательностей импульсов определенной формы. Устройства, с помощью которых формируют и обрабатывают импульсные сигналы, действуют, как правило, в ключевом режиме. Такой режим работы усилительных приборов имеет ряд существенных преимуществ перед активным режимом. Основное преимущество ключевого режима — высокий, близкий к 90...95 % КПД. В ключевом режиме достигается достаточно большая мощность во время действия импульсов при малом значении средней мощности, расходуемой в схеме. Поэтому импульсные и цифровые устройства обладают меньшими массой и габаритными размерами. Транзисторы в импульсных и цифровых схемах либо закрыты, либо полностью открыты (насыщены), поэтому на них рассеивается незначительная мощность и надежность устройств очень высока.

## 7.1. Общие сведения. Логические сигналы и элементы

Будучи тесно и органично связанными, импульсная и цифровая техника отражает различные характеристики функционирования одних и тех же устройств. Прежде всего это связано с тем, что импульсные сигналы являются носителями цифровой информации, а импульсные устройства лежат в основе реализации цифровых схем. Импульсная техника связана с формированием, усилением и преобразованием импульсных сигналов по их физическим параметрам (амплитуде, мощности, длительности и частоте следования). Цифровые устройства выполняют лишь *логические* преобразования информации, которую несут в себе различные последовательности импульсных сигналов, выраженных условными параметрами, например совокупностью разных уровней или чередованием высоких и низких уровней напряжения. Цифровые устройства осуществляют логическое запоминание, суммирование, умножение, шифрацию и дешифрацию кодов, деление частоты следования импульсов и ряд других операций.

Как в цифровой технике осуществляется отражение информации? Люди для обозначения различных предметов, понятий, действий или намерений пользуются словами. Слова строят из букв, которые берут из некоторого набора, назы-

ваемого *алфавитом*. В цифровой технике для этих целей используют *кодовые слова*. Особенность кодовых слов заключается в том, что для конкретной схемы они имеют одинаковую длину (т. е. представляют собой буквенные последовательности одинаковой длины), и для их построения часто применяется простейший алфавит, состоящий из двух букв. Как уже отмечалось, в современной цифровой технике наиболее часто информация (кодовые слова) представляется импульсными сигналами прямоугольной формы, имеющими два фиксированных уровня напряжения. Сигналу 1 (*единице*) соответствует уровень «*высокого*» напряжения — потенциал напряжения питания, а сигналу 0 (*нулю*) — уровень «*нулевого*» напряжения, т. е. потенциал, близкий к потенциалу корпуса (общей шины). Таким образом, кодовое слово в цифровой технике имеет вид последовательности символов 1 и 0 определенной длины, например 10110110. Подобными кодовыми словами может представляться и цифровая, и нецифровая информация. Различие в физической сущности импульсных и цифровых сигналов предопределяет разные методы их описания, теоретического исследования и практического использования.

### Логические сигналы

Цифровые сигналы отражают числовые значения физических величин, например напряжения. Наряду с ними в цифровых устройствах могут действовать сигналы, появление которых связано с наступлением или ненаступлением какого-либо события, т. е. сигналы, отражающие логические задачи. Несмотря на формальную схожесть, между импульсными и цифровыми сигналами существенная разница. Чтобы ее увидеть, рассмотрим цифровое сравнивающее устройство (цифровой компаратор), на входы которого поступают цифровые коды, а результат этого воздействия отражается на двух выходах. Когда коды на входах компаратора окажутся равными, на одном из его выходов появится потенциал определенного уровня (например, импульс, т. е. 1), свидетельствующий о наступлении события, — равенстве кодов. Такой же сигнал появится на этом выходе при равенстве кодов другого значения. Тот же сигнал появится на другом выходе компаратора, если один код меньше другого, причем значение их разности никакого влияния на величину сигнала не оказывает.

Построение компаратора предусмотрено таким образом, что появление описанных сигналов на выходах есть ответ на наступление того или иного события вне зависимости от его содержания; важны лишь истинность этого события или его ложность. Наличие или отсутствие описанных сигналов и порождающие их условия связаны с логическим выражением типа «если ..., то» и другими логическим связями. Поэтому такие сигналы и называют *логическими*. Название объясняется тем, что аналогичные условия между причиной и следствием являются предметом изучения в *формальной логике*. Как известно, формальная логика оперирует с высказываниями вне зависимости от их содержания, учитывая лишь их истинность (*True*) или ложность (*False*). Истинное высказывание: «Постоянный ток не протекает через конденсатор», ложное: «Одиночный видеоимпульс имеет линейчатый характер частотного спектра». По



Д. Буль

структуре высказывания могут быть и простыми, и сложными. Приведенные высказывания являются простыми. Простое высказывание содержат один факт, который не зависит от других фактов, а значит, может быть либо истинным, либо ложным. Сложное высказывание содержит несколько простых. Формальная логика позволяет однозначно представлять сложное высказывание совокупностью простых, а введение символов, обозначающих простые высказывания, — решать логические задачи математическими средствами. Их совокупность составляет содержание алгебры логики, разработанной английским математиком Джорджем Булем (*George Boole*; 815–1864). В соответствии с булевой алгеброй истинному высказыванию (наступлению события) ставится в соответствие символ 1 (логическая единица), а ложному (ненаступлению события) — символ 0 (логический нуль). Значит, булева алгебра оперирует с двумя переменными: 1 и 0, т. е. с двоичными переменными. Функция двоичных переменных принимает те же два значения (1 и 0) и называется *логической функцией*. Может возникнуть вопрос: какая связь между алгеброй логики и двоичным кодированием? Математический аппарат алгебры логики удобен для описания того, как функционируют аппаратные средства цифровых систем, поскольку основной системой счисления в них является двоичная, в которой используются цифры 1 и 0, а значений логических переменных тоже два: 1 и 0. Значит, одни и те же цифровые устройства можно применять для обработки и хранения как числовой информации, представленной в двоичной системе счисления, так и логических переменных. Логические символы 1 и 0 отношения к числовому значению сигнала не имеют. Они лишь описывают качественное состояние события, и к ним не применимы арифметические операции. В электрических цепях эти символы представляют так же, как и аналогичные в цифровой структуре сигнала: логическая 1 — высоким (часто импульсом), а логический 0 — низким (часто нулевым) уровнем потенциала (например, представление двоичных сигналов в компьютерах, рис. 7.1).

В основе алгебры логики Буля лежат три основные элементарные операции: ИЛИ (*логическое сложение*, или *дизъюнкция*; от лат. *disjunctio* — различие), И (*логическое умножение*, или *конъюнкция*), НЕ (*логическое отрицание*, или *инверсия*). С помощью этих логических операций можно описать функционирование электрической цепи, радиотехнического устройства и т. д.

Не вдаваясь в подробное изучение, приведем математические записи основных аксиом и законов алгебры Буля. Применение этих аксиом и законов позволяет существенно упростить представление логических функций, которые имеют различные формы записи: алгебраическое, табличное, графическое и т. д.

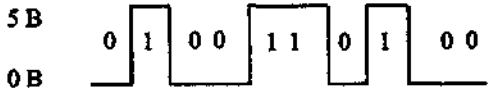



**Рис. 7.1.** Представление двоичных сигналов в компьютерах

В классической математике для задания функции обычно используют два способа — аналитический (формулы) и табличный (таблицы с значениями функций). Подобными же способами можно задавать и логические функции. Наибольшее распространение в теории и практике получило представление логических функций в виде таблиц, называемых *таблицами состояний* (*table state*), *диаграммами состояния* (*state diagram*), *переключательными таблицами* или *таблицами истинности* (*table true*).

*Таблица истинности* — табличное представление логической схемы (операции), в котором перечислены все возможные сочетания значений истинности входных сигналов (операндов) вместе со значением истинности выходного сигнала (результата логической операции) для каждого из этих сочетаний входных сигналов.

**Аксиомы:**

$1 + X = 1$

$0 \cdot X = 0$

$0 + X = X$

$1 \cdot X = X$

$X + X = X$

$X \cdot X = X$

$X + \overline{X} = 1$

$X \cdot \overline{X} = 0$

$\overline{\overline{X}} = X$

**Законы:**

$X_1 + X_2 = X_2 + X_1$

$X_1 \cdot X_2 = X_2 \cdot X_1$

$X_1 + X_2 + X_3 = X_1 + (X_2 + X_3)$

$X_1 \cdot X_2 \cdot X_3 = X_1 \cdot (X_2 \cdot X_3)$

$X_1 \cdot (X_2 + X_3) = (X_1 \cdot X_2) + (X_1 \cdot X_3)$

$X_1 + (X_2 \cdot X_3) = (X_1 + X_2) \cdot (X_1 + X_3)$

$\overline{X_1 + X_2} = \overline{X_1} \cdot \overline{X_2}$

$\overline{X_1 \cdot X_2} = \overline{X_1} + \overline{X_2}$

$X_1 + (X_1 \cdot X_2) = X_1$

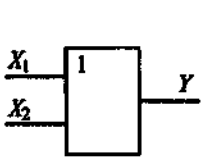
$X_1 \cdot (X_1 + X_2) = X_1$

Используя основные законы булевой алгебры, нетрудно показать, что функционально полным набором, позволяющим осуществить любую логическую операцию, являются функции ИЛИ, И и НЕ.

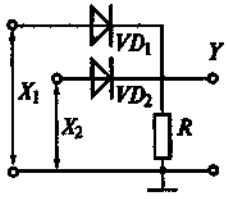
### Логические элементы

Логические элементы выполняют все возможные операции над логическими переменными.

*Логический элемент* — часть электронной логической схемы, которая реализует элементарную логическую функцию. Логическими элементами являются электронные схемы ИЛИ, И, НЕ, ИЛИ-НЕ, И-НЕ и др. (называемые вентилями). К ним в компьютерной технике часто относят триггер. С помощью этих схем можно реализовать любую логическую функцию, описывающую работу устройств компьютера. Обычно у вентилей бывает от двух до восьми входов и один или два выхода. Каждый логический элемент имеет свое условное обозначение, которое выражает его логическую функцию, но не указывает на то, какая именно электронная схема в нем реализована. Это упрощает запись и понимание сложных логических схем.




а




б

| | | | | |
|---|---|---|---|---|
| $X_1$ | 0 | 0 | 1 | 1 |
| $X_2$ | 0 | 1 | 0 | 1 |
| $Y$ | 0 | 1 | 1 | 1 |

в

**Рис. 7.2.** Логический элемент ИЛИ:
*а* — обозначение; *б* — диодная реализация; *в* — таблица истинности

*Логический элемент* ИЛИ имеет два или более входов $X_1$, $X_2$, ... и выход $Y$ (рис. 7.2, *а*); 1 на схеме — от устаревшего обозначения дизъюнкции: >=1 (т. е. значение дизъюнкции равно единице, если сумма значений операндов больше или равна 1). Операция дизъюнкции для двух переменных

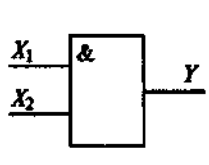
$$Y = X_1 + X_2. \tag{7.1}$$

Рассмотрим диодный двухвходовой элемент ИЛИ (рис. 7.2, *б*). Если единичный уровень напряжения (логическая 1) присутствует или на первом $X_1$, или на втором $X_2$, или на обоих входах, то через открытые диоды $VD_1$ или $VD_2$ это напряжение поступит на выход, создавая единичный уровень выходного напряжения. При нулевых входных напряжениях диоды закрыты, и на выходе элемента будет нулевое напряжение. В соответствии с этими условиями элементу ИЛИ отвечает таблица истинности, показанная на рис. 7.2, *в*.

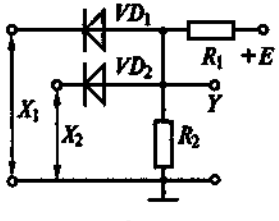
*Логический элемент* И (*логическая схема совпадения*) содержит два или более входов $X_1$, $X_2$, ... и один выход $Y$ (рис. 7.3, *а*). Представленный в обозначении символ & (*and* — И) отражает операцию конъюнкции для двух переменных и определяется выражением

$$Y = X_1 \cdot X_2. \tag{7.2}$$

Обратимся к элементарной, двухвходовой диодной реализации элемента И (рис. 7.3, *б*). Если любое входное напряжение равно нулю, то катод соответствующего диода через источник на входе замкнут на корпус, и выходное напряжение равно нулю. Напряжение на выходе будет, если высокое напряжение, запирающее диоды, будет подано одновременно и на оба входа ($X_1 = 1$, $X_2 = 1$).




а




б

| | | | | |
|---|---|---|---|---|
| $X_1$ | 0 | 0 | 1 | 1 |
| $X_2$ | 0 | 1 | 0 | 1 |
| $Y$ | 0 | 0 | 0 | 1 |

в

**Рис. 7.3.** Логический элемент И:
*а* — обозначение; *б* — диодная реализация; *в* — таблица истинности

В этом случае диоды будут закрыты, и через резисторы $R_1$ и $R_2$ потечет ток, который создаст на резисторе $R_2$ высокое выходное напряжение (сигнал $Y = 1$). Таблица истинности элемента И показана на рис.7.3, *в*.

*Логический элемент* НЕ (*инвертор*) выполняет операцию инверсии, т. е. инвертирует входную логическую величину в соответствии с формулой

$$Y = \overline{X}. \tag{7.3}$$

Словами данная операция трактуется как «*игрек равен не икс*», или «*игрек равен иксу под чертой*».

В обозначении элемента НЕ (рис. 7.4, *а*) операция инверсии отражена кружком на выходе схемы. Работу элемента НЕ иллюстрирует таблица истинности, показанная на рис. 7.4, *б*. Операцию инверсии выполняют *электронные ключи* — транзисторные усилительные каскады (*транзисторные ключи*).

Рассмотрим усилительный ключевой каскад с ОЭ (рис. 7.4, *в*). Транзистор, работающий в ключевом режиме, может работать в двух режимах. Первому из них соответствует точка *1* на выходных характеристиках (рис. 7.4, *г*): при $U_{вх} = 0$ ($X = 0$) транзистор закрыт (устанавливают резистором $R_б$), ток коллектора практически равен нулю, а напряжение на коллекторе (выходе) $U_{вых} \approx E_к$ ($Y = 1$). Второе состояние транзистора (точка *2* на рис. 7.4, *г*) обеспечивают подачей на вход высокого положительного напряжения ($X = 1$). При этом $U_{вых} \approx \Delta U_{нас} \approx 0$, и ток коллектора насыщения возрастает до предельно возможного $I_{кн} \approx E_к/R_к$. Для надежного насыщения (предельного открытия) транзистора его ток базы должен превысить расчетный ток базы насыщения в 1,5…2 раза, т. е. $I_{бн} = (1{,}5\ldots2)I_к/h_{21}$.

Современные ключевые каскады — интегральные схемы, в которых резисторами служат полевые транзисторы — *динамические сопротивления*.

**Пример 7.1.** Определить ток базы насыщения в схеме ключа (рис. 7.4, *в*) при таких параметрах: $E_к = 12$ В, $R_к = 2$ кОм, $h_{21} = 60$, степень насыщения тока 1,5.

Р е ш е н и е. Максимальный ток коллектора насыщения $I_{кн} = E_к/R_к = 6$ мА. Отсюда ток базы насыщения транзистора $I_{бн} = 1{,}5I_{кн}/h_{21} = 0{,}15$ мА.

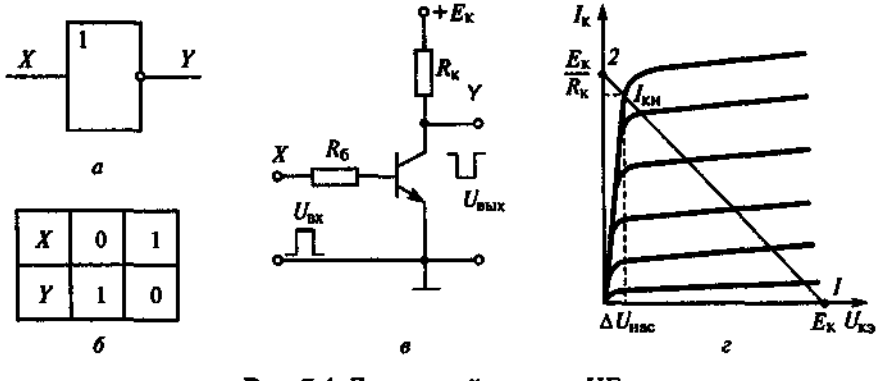
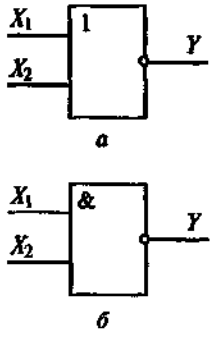



**Рис. 7.4.** Логический элемент НЕ:
*а* — обозначение; *б* — таблица истинности; *в* — транзисторный ключ; *г* — характеристики ключа




| $X_1$ | $X_2$ | $Y$ |
|---|---|---|
| 0 | 0 | 1 |
| 0 | 1 | 0 |
| 1 | 0 | 0 |
| 1 | 1 | 0 |

*в*

| $X_1$ | $X_2$ | $Y$ |
|---|---|---|
| 0 | 0 | 1 |
| 0 | 1 | 1 |
| 1 | 0 | 1 |
| 1 | 1 | 0 |

*г*

**Рис. 7.5.** Базовые элементы:
*а* — ИЛИ-НЕ; *б* — И-НЕ; *в, г* — соответствующие им таблицы истинности

В импульсной и цифровой технике применяют две *базовые схемы* логических элементов и их модификации. В этих микросхемах используют сочетания усложненных элементов ИЛИ и И с усилительным инвертором НЕ (рис. 7.5).

*Элемент* ИЛИ-НЕ (рис. 7.5, *а*) осуществляет логическое сложение двух (и более) цифровых сигналов с инверсией результата:

$$Y = \overline{X_1 + X_2}. \tag{7.4}$$

*Элемент* И-НЕ (рис. 7.5, *б*) производит логическое умножение двух (и более) цифровых сигналов с инверсией результата:

$$Y = \overline{X_1 \cdot X_2}. \tag{7.5}$$

Таблицы истинности элементов ИЛИ-НЕ и И-НЕ показаны на рис. 7.5, *в, г*.

Сейчас имеется ряд семейств цифровых интегральных схем. Среди них наиболее распространены: ДТЛ — *диодно-транзисторная логика*; ТТЛ — *транзисторно-транзисторная логика*; ЭСЛ — *эмиттерно-связанная логика*; $И^2Л$ — *интегральная инжекционная логика*. Для правильного выбора семейств надо знать их возможности, параметры и совместимость с другими элементами.

Заметим, что в булевой алгебре доказана теорема, утверждающая, что функцию любого числа переменных можно получить по принципу наложения из функций двух переменных. Это позволяет строить разнообразные многовходовые комбинации схемы из двухвходовых структур.

**Пример 7.2.** Реализовать логическое умножение десяти переменных на двух-, трех- и четырехвходовых интегральных цифровых элементах.

Р е ш е н и е. Используя два четырехвходовых и один двухвходовый элементы И-НЕ, осуществим следующие логические операции:

$$Y_1 = \overline{X_1 \cdot X_2 \cdot X_3 \cdot X_4}; \quad Y_2 = \overline{X_5 \cdot X_6 \cdot X_7 \cdot X_8}; \quad Y_3 = \overline{X_9 \cdot X_{10}}.$$

На основе трехвходового элемента И-НЕ реализуем требуемую функцию

$$Y = \overline{Y_1 \cdot Y_2 \cdot Y_3} = X_1 \cdot X_2 \cdot \ldots \cdot X_{10}.$$

# 7.2. Триггеры

В 1919 г. известные американские ученые В. Иклс (*W. Eccles*) и Ф. Джордан (*F. Jordan*) и независимо от них русский ученый М.А. Бонч-Бруевич разработали схему триггера; тогда же американские инженеры Я. Абрахам (*J. Abraham*) и Е. Блох (*E. Bloch*) опубликовали описание мультивибратора. Триггер и мультивибратор широко используют в импульсной и цифровой технике.

*Триггер* (от англ. *trigger* — спусковая схема), или *бистабильная* (т. е. *двустабильная*) *полупроводниковая ячейка*, — устройство с двумя устойчивыми состояниями, скачком переходящее из одного состояния в другое под воздействием входных сигналов. Одно из устойчивых состояний цифрового триггера соответствует двоичной единице, а другое — двоичному нулю. Название процесса смены состояний триггера имеет ряд синонимов: *переброс*, *переключение*, *опрокидывание*. В английском языке триггер называют термином «*flip-flop*», т. е. «хлопанье». Это звукоподражательное название схемы указывает на ее способность мгновенно переходить из одного состояния в другое и наоборот.

## Симметричный триггер

Основой большинства цифровых микросхем является *симметричный триггер* (рис. 7.6, *а*), содержащий два транзисторных ключа, охваченных перекрестными положительными обратными связями. Собственно триггер (*триггерная ячейка*) состоит из транзисторов $VT_1$, $VT_2$ и резисторов $R_1$, $R_2$ (в микросхемах резисторы — динамические сопротивления). Транзисторы $VT_3$ и $VT_4$ — вспомогательные ключи, служащие для управления триггерной ячейкой. Они почти все время закрыты и на характеристики триггерной ячейки практически не влияют. Триггер имеет два устойчивых состояния: 1 — транзистор $VT_1$ открыт и насыщен, а транзистор $VT_2$ закрыт; 2 — транзистор $VT_1$ закрыт, а транзистор $VT_2$ открыт и насыщен. Пусть триггер находится в первом устойчивом состоянии, когда транзисторы $VT_3$ и $VT_4$ закрыты.

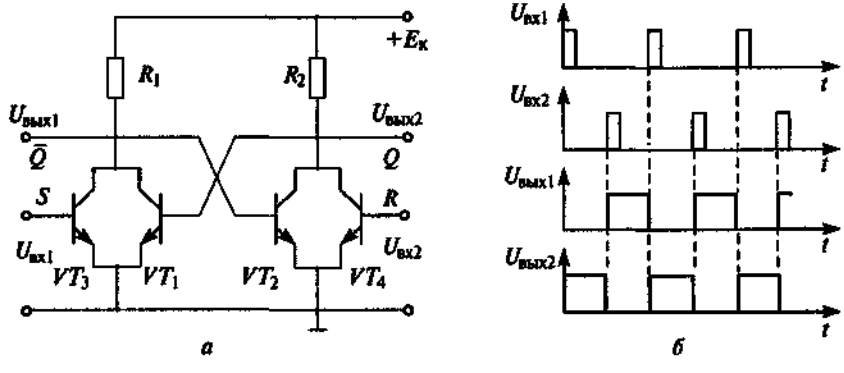



**Рис. 7.6.** Симметричный триггер:
*а* — схема; *б* — временны́е диаграммы

Принцип действия триггера поясняют упрощенные временны́е диаграммы напряжения, показанные на рис. 7.6, *б*. Проанализируем работу триггера по представленным диаграммам. Напряжение $U_{вых2}$ на коллекторе закрытого транзистора $VT_2$ близко к напряжению питания $E_к$. Ток базы транзистора $VT_1$, протекающий через резистор $R_2$ и создающий на нем падение напряжения, удерживает транзистор открытым. В результате напряжение на его коллекторе будет близким к нулю. Поскольку коллектор транзистора $VT_1$ соединен с базой транзистора $VT_2$, то последний будет закрыт. Следовательно, триггер находится в первом устойчивом состоянии. Аналогичная картина наблюдается в симметричном триггере, когда он находится во втором устойчивом состоянии.

Переключение триггера производят входными (отпирающими) импульсами положительной полярности (рис. 7.6, *б*), подаваемыми на базу того вспомогательного транзистора, который включен параллельно соответствующему закрытому основному транзистору. Для переключения триггера, находящегося в первом устойчивом состоянии, входной импульс подают на базу транзистора $VT_4$. При этом транзистор $VT_2$ откроется, и напряжение $U_{вых2}$ на его коллекторе, а значит, и на базе транзистора $VT_1$ упадет до нуля, вызывая запирание последнего. В результате действия положительной ОС транзистор $VT_2$ перейдет из активного режима в режим насыщения. Триггер переключится во второе устойчивое состояние. Для его переключения надо открыть входным импульсом вспомогательный транзистор $VT_3$. Запуск триггера попеременной подачей на входы управляющих импульсов называют *раздельным* (рис. 7.6, *б*). Одновременная подача отпирающих импульсов на оба входа в этой схеме триггера недопустима: если входные импульсы поступают одновременно, то вспомогательные транзисторы откроются и своими выходными сигналами создадут на базах транзисторов $VT_1$ и $VT_2$ нулевые потенциалы, вследствие чего последние закроются. По окончании действия импульсов оба транзистора откроются, и схема временно окажется в неустойчивом положении, из которого под воздействием внутренних флуктуаций триггер может равновероятно перейти в устойчивое состояние.

Основное назначение триггера — *запоминание* (*хранение*) цифровой информации. Приняв одно состояние триггера за 1, а другое — за 0, можно считать, что триггер помнит (хранит) один разряд числа, записанного в двоичном коде. Такое хранение цифровой информации теоретически может продолжаться сколь угодно долго, пока на триггер не поступят сигналы управления или с него не будет снято напряжение питания. Поскольку один триггер может запомнить только один разряд двоичного кода, то для запоминания байта нужно $2^3 = 8$ триггеров, для запоминания килобайта, соответственно, $8 \cdot 2^{10} = 8\,192$ триггеров. Современные микросхемы памяти содержат миллионы триггеров.

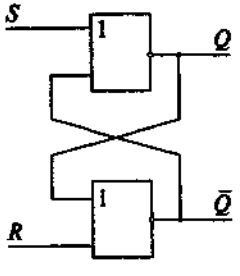
### Интегральные триггеры

Интегральные триггеры состоят из *собственно триггера* и схемы управления его входами (входной логики). В зависимости от метода построения входной логики получают триггеры с разными функциональными возможностями: триггеры типов *RS*, *D*, *T*, *JK* («жи-ка») и др.

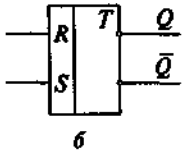
По способу управления триггеры делят на *асинхронные* и *синхронные* (*тактируемые*). Переключение асинхронных триггеров происходит непосредственно с поступлением входного сигнала на *информационный вход*. Синхронные триггеры имеют дополнительный *тактовый вход C* (*C — clock* — фиксация времени) и переключаются лишь при одновременном поступлении входного и *тактового* импульсов. Различают *однотактные* (*T*) и *двухтактные*, или *двухступенчатые* (*TT*), триггеры. Однотактные триггеры переключаются по переднему фронту тактового импульса, а двухтактные — по его заднему фронту. Триггеры имеют два симметричных взаимоинверсных выхода — *прямой* $Q$ и *инверсный* $\overline{Q}$ («не ку», «ку под чертой»). Триггер находится в единичном состоянии, когда $Q = 1$, $\overline{Q} = 0$, и в нулевом, когда $Q = 0$, $\overline{Q} = 1$.

***RS*-триггер.** На рис. 7.7 показана логическая структура (здесь и далее обозначено как схема), условное обозначение и таблица истинности асинхронного *RS*-триггера на цифровых логических элементах ИЛИ-НЕ. Этот триггер имеет в своей структуре два симметричных информационных входа $R$ и $S$ и два симметричных выхода $Q$ и *инверсный* $\overline{Q}$. На каждый из входов $R$ и $S$ могут подаваться входные сигналы в виде кратковременных импульсов (⎍). Наличие импульса на входе считают единицей, а его отсутствие — нулем. Вход, по которому триггер устанавливается в единичное состояние, называют входом $S$ (*set* — установка), а в нулевое — входом $R$ (*reset* — сброс).

На двух входах *RS*-триггера возможны четыре комбинации входных сигналов. Если $S = 1$ и $R = 0$, то триггер переходит в *единичное состояние*; при $S = 0$ и $R = 1$ — в *нулевое состояние*. При установке на входах логических нулей ($S = 0$, $R = 0$) триггер сохраняет предыдущее состояние. Такую комбинацию входных сигналов называют *нейтральной*, или *памятью*. Если на входы одновременно поступают единичные сигналы ($S = 1$, $R = 1$), то на обоих выходах появляются логические нули $Q = 0, \overline{Q} = 0$, и триггер утрачивает свои свойства, поскольку под воздействием внутренних факторов он может равновероятно перейти как в единичное, так и нулевое состояние. Такая комбинация входных сигналов является *запрещенной* для триггера.




*а*

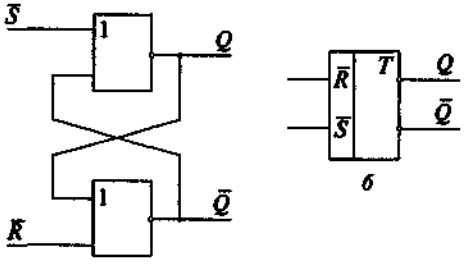



*б*

| $S$ | $R$ | $Q$ | $\overline{Q}$ |
|---|---|---|---|
| 1 | 0 | 1 | 0 |
| 0 | 1 | 0 | 1 |
| 0 | 0 | Память | |
| 1 | 1 | Запрет | |

*в*

**Рис. 7.7.** *RS*-триггер с прямыми входами:
*а* — схема; *б* — обозначение; *в* — таблица истинности




*а* *б*

| $\bar{S}$ | $\bar{R}$ | $Q$ | $\bar{Q}$ |
|---|---|---|---|
| 0 | 1 | 1 | 0 |
| 1 | 0 | 0 | 1 |
| 0 | 0 | Запрет | |
| 1 | 1 | Память | |

*в*

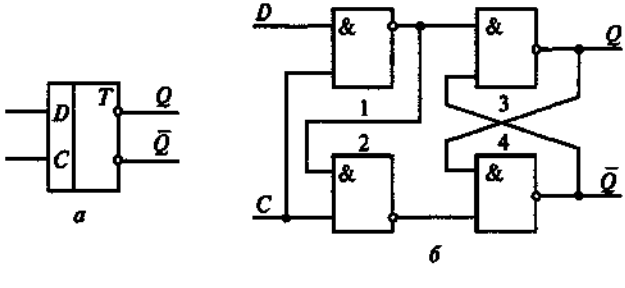
**Рис. 7.8.** *RS*-триггер с инверсными входами:
*а* — схема; *б* — обозначение; *в* — таблица истинности

*RS*-триггер, построенный на логических элементах ИЛИ-НЕ и переключающийся единичными входными сигналами, называют *триггером с прямыми входами*. *RS*-триггер, выполненный на базовых элементах И-НЕ (рис. 7.8), — *триггером с инверсными входами*. Аналогом интегрального *RS*-триггера с прямыми входами может служить симметричный триггер на биполярных транзисторах (см. рис. 7.6, *а*), в котором $U_{вх1}$ служит входом $S$, а $U_{вх2}$ — входом $R$. При этом $U_{вых1}$ соответствует инверсному выходу $\bar{Q}$, а $U_{вых2}$ — прямому выходу $Q$.

Чтобы *RS*-триггер на элементах И-НЕ (рис. 7.8, *а*) вырабатывал те же выходные сигналы, что и триггер с прямыми входами, надо подать на его входы инверсные сигналы. Поэтому в условном обозначении *RS*-триггера с инверсными входами (рис. 7.8, *б*) на них показаны знаки инверсии. Таблица истинности *RS*-триггера с инверсными входами представлена на рис. 7.8, *в*.

***D*-триггер** имеет два входа — информационный $D$ (от англ. *delay* — задержка) и тактовый $C$ (рис. 7.9, *а*). Запись информации производится только в моменты поступления тактовых импульсов на вход $C$ ($C = 1$). В этом случае выходной сигнал $Q$ повторяет входной $D$: $Q = 1$, если $D = 1$, и $Q = 0$, если $D = 0$.

В схеме D-триггера на элементах И-НЕ (рис. 7.9, б) элементы 1 и 2 образуют схему входной логики, а 3 и 4 — триггерную ячейку. В паузах между тактовыми импульсами элементы 1 и 2 закрыты, и на их выходах существуют логические единицы, что служит нейтральной комбинацией для триггерной ячейки.




*а* *б*

**Рис. 7.9.** *D*-триггер:
*а* — обозначение; *б* — схема

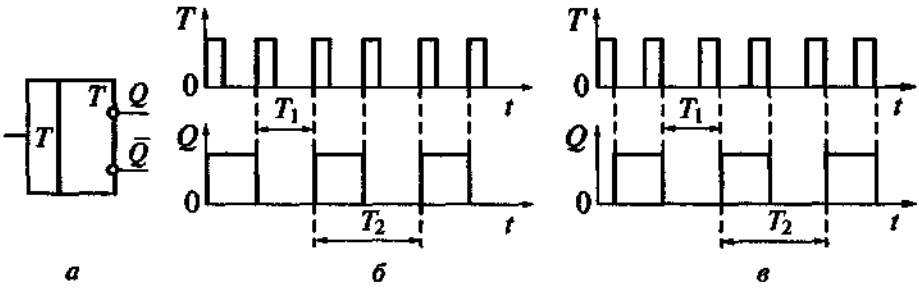


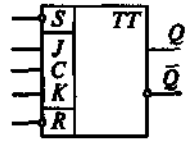

**Рис. 7.10.** *T*-триггер:
*а* — обозначение; *б*, *в* — временные диаграммы однотактного и двухтактного триггеров

***T*-триггер** называют *счетным*, он имеет один информационный вход *T* (рис. 7.10, *а*), а переключение происходит в момент поступления очередного входного импульса. Период повторения выходных импульсов *T*-триггера в два раза больше периода повторения входных импульсов (рис. 7.10, *б*). Поскольку частота следования выходных импульсов уменьшается вдвое, то *T*-триггеры применяют в *делителях частоты*. *T*-триггер создают и по двухтактной схеме, например на базе двух *RS*-триггеров. Тогда в триггере происходит задержка выходного импульса на длительность входного и он срабатывает по его заднему фронту (рис. 7.10, *в*).

***JK*-триггер** имеет информационные входы *J* (от англ. *jump* — переброс, прыжок), *K* (от англ. *keep* — держать, сохранять), тактовый вход *C* (рис. 7.11, *а*) и является универсальным. Если тактовый сигнал *C* = 1, то *JK*-триггер функционирует как *RS*-триггер с прямыми входами, при этом вход *J* может устанавливать триггер в состояние «1», а вход *K* — в состояние «0».

В отличие от *RS*-триггера, комбинация входных сигналов *J* = *K* = 1 не запрещена для *JK*-триггера — в этом случае происходит его переключение. Таблица истинности показана на рис. 7.11, *б*. В *JK*-триггере имеются два дополнительных инверсных входа $\overline{S}$ и $\overline{R}$, используемых для его принудительной установки в состояние «1» или «0». При этом состояние входов *J*, *K* и *C* могут быть произвольны. Если необходимо исключить влияние входов принудительной установки $\overline{S}$ и $\overline{R}$, на них подают логическую 1, т. е. подсоединяют через ограничивающие токи резисторы к источнику питания.

**Пример 7.3.** На *JK*-триггерах реализовать делитель частоты на 8. С помощью временных диаграмм напряжений пояснить работу и составить таблицу истинности.

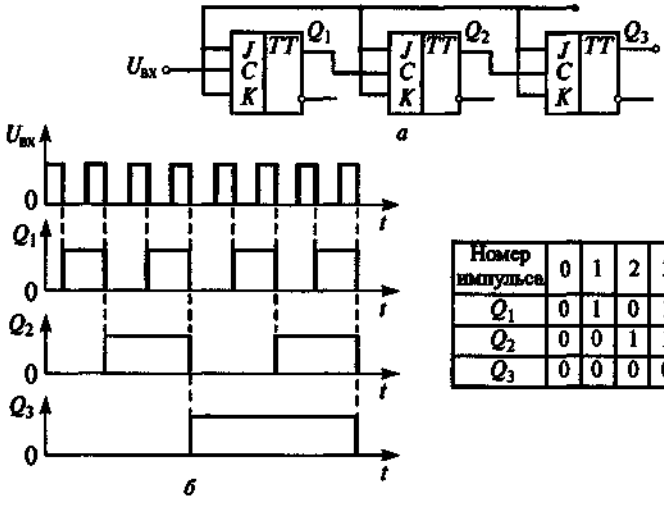



а

| *J* | *K* | *Q* | $\overline{Q}$ |
|---|---|---|---|
| 0 | 1 | 1 | 0 |
| 1 | 0 | 0 | 1 |
| 0 | 0 | Память | |
| 1 | 1 | Переброс | |

б

**Рис. 7.11.** *JK*-триггер:
*а* — обозначение; *б* — таблица истинности




| Номер импульса | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|---|
| $Q_1$ | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 |
| $Q_2$ | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 |
| $Q_3$ | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 |

в

**Рис. 7.12.** Делитель частоты на 8:
*а* — схема; *б* — временные диаграммы; *в* — таблица истинности

Р е ш е н и е. В качестве делителей частоты выберем *JK*-триггеры. Поскольку триггер делит частоту следования импульсов на 2, то для деления частоты на 8 требуется три последовательно соединенных триггера (рис. 7.12, *а*). Входные импульсы подают на тактовый вход *C* первого триггера, а входы *C* второго и третьего триггеров соединены соответственно с выходами *Q* первого и второго триггеров. На остальные входы поданы 1 для установки триггеров в исходное состояние. На рис. 7.12, *б* показаны временные диаграммы работы делителя частоты. Из таблицы истинности (рис. 7.12, *в*) следует, что каждому из восьми входных импульсов однозначно соответствуют определенные значения сигналов на выходах $Q_1$, $Q_2$ и $Q_3$.

## 7.3. Счетчики, шифраторы и дешифраторы

**Счетчик** — цифровое устройство (цифровая микросхема или ее часть), выполненное на интегральных триггерах и осуществляющее счет поступающих на его вход цифровых кодов (импульсов). Счет кодового числа в счетчике представляется определенными комбинациями состояний триггеров. При поступлении на вход схемы очередной логической единицы в счетчике устанавливается новая комбинация состояний триггеров, соответствующая числу, превышающему предыдущее на единицу. Такие счетчики называют *суммирующими*. В цифровой технике широкое применение находят *вычитающие счетчики*, в которых в процессе счета кодовые числа (входные импульсы) последовательно убывают на единицу, а также *реверсивные счетчики*, способные переключаться из режима суммирования в режим вычитания и наоборот.

С и с т е м ы  с ч и с л е н и й. Для представления чисел в интегральных счетчиках обычно используются *двоичная, десятичная* и *двоично-десятичная системы счисления*.

Совокупность приемов и правил, по которым числа записывают и читают, называют системой счисления. В кодировании сообщений используют различные системы счисления. Из всех систем счисления практическое значение при кодировании сообщений имеют двоичная ($R = 2$), троичная ($R = 3$), четверичная ($R = 4$), восьмеричная ($R = 8$) и десятичная ($R = 10$). Применяется также двенадцатиричная система ($R = 12$) для отсчитывания времени и шестидесятиричная ($R = 60$) для отсчета углов. Во всех системах используется различное число цифр. Напомним, что в двоичной системе используют цифры 0 и 1 (например, 1011001, причем первый разряд — старший по весу). Представление компьютерных кодов производят в шестнадцатеричной системе счисления, где для кодирования используют сочетание цифр (0, 1, 2, …, 9) и букв (A, B, C, D, E, F).

Основанием счисления в десятичной системе является число 10. Форма записи числа в десятичной системе счисления имеет следующий вид:

$$N_{10} = \sum_{i=0}^{n-1} a_i 10^i = a_{n-1}10^{n-1} + a_{n-2}10^{n-2} + \ldots + a_2 10^2 + \ldots + a_1 10^1 + a_0 10^0, \qquad (7.6)$$

где $10^i$ — десятичный разряд; $a_i$ — значение символа в соответствующем разряде, которое может быть любым от 0 до 9.

Каждое из слагаемых ряда (7.6) и называется *разрядом*. Поэтому десятичное число $N_{10}$ может быть одно-, двух-, трех- и $n$-разрядным. Например, число 582 с помощью трех десятичных разрядов запишется как

$$N = 582 = 5 \cdot 10^2 + 8 \cdot 10^1 + 2 \cdot 10^0.$$

Аналогично записывается целое число и в двоичной системе счисления

$$N_2 = \sum_{i=0}^{n-1} b_i 2^i = b_{n-1}2^{n-1} + b_{n-2}2^{n-2} + \ldots + b_2 2^2 + \ldots + b_1 2^1 + b_0 2^0, \qquad (7.7)$$

где коэффициенты $b_i$ принимают лишь два значения: 0 и 1. Например, число 582 в двоичной системе запишется в виде

$$N_2 = 1 \cdot 2^9 + 0 \cdot 2^8 + 0 \cdot 2^7 + 1 \cdot 2^6 + 0 \cdot 2^5 + 0 \cdot 2^4 + 0 \cdot 2^3 + 1 \cdot 2^2 + 1 \cdot 2^1 + 0 \cdot 2^0.$$

Итак, числу 582 в десятичной системе счисления соответствует число в двоичной системе 1001000110, которое называют *кодом* двоичного числа.

Теоретически в качестве основания счисления можно принять любое целое число $m$ и тогда представить число $N$ как:

$$N_m = \sum_{i=0}^{n-1} d_i m^i = d_{n-1}m^{n-1} + d_{n-2}m^{n-2} + \ldots + d_2 m^2 + \ldots + d_1 m^1 + d_0 m^0, \qquad (7.8)$$

где $d_i$ – символ в соответствующем разряде, принимающий значения от 1 до $n$.

При использовании десятичной системы счисления для образования кода требуется десять различных импульсов, например отличающихся амплитудой, длительностью и т. д. Такое представление кодов применяют редко, поскольку для образования и его распознавания требуется сложная аппаратура.

Наиболее просты арифметические действия в двоичной системе. Так сложение двоичных чисел осуществляют по правилам:

$$0 + 0 = 0; \quad 0 + 1 = 1; \quad 1 + 0 = 1; \quad 1 + 1 = 10.$$

Применяют еще поразрядное сложение без переноса в старший разряд, т. е. сложение по модулю 2 (обозначается символом $\oplus$). Правила сложения таковы:

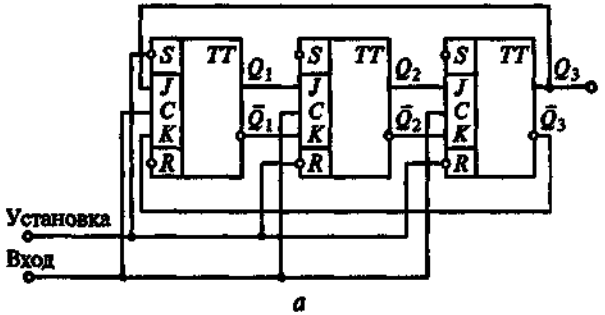
$$0 \oplus 0 = 0; \quad 1 \oplus 1 = 0; \quad 0 \oplus 1 = 1; \quad 1 \oplus 0 = 1.$$

Напомним, что при преобразовании последовательности элементов сообщения в двоичный код для передачи его по каналу связи достаточно передавать всего два различных сигнала. В частности, символы 0 и 1 можно передавать колебаниями с различными частотами либо импульсами тока (напряжения) разной полярности или амплитуды. При этом для образования и обработки двоичного кода можно использовать двоичные цифровые элементы, имеющие два состояния: *единица* и *нуль*. При использовании двоичной системы логические уровни на прямых выходах триггеров определяют цифры двоичных разрядов числа. В этом случае каждый разряд числа в счетчике обеспечивается состоянием одного триггера. Максимальное число импульсов, которое можно записать, $N = 2^n - 1$, где $n$ — число разрядов (число триггеров) в счетчике.

При реализации кодов каждому символу используемой системы счисления (комбинации чисел) должен соответствовать свой элемент. В зависимости от очередности передачи элементов кода все коды можно разделить на последовательные и параллельные. У *последовательного кода* элементы передаются последовательно во времени (и могут передаваться по одному каналу). У *параллельного кода* элементы передаются одновременно по различным каналам.

Рассмотрим *двоичный кольцевой счетчик*, содержащий несколько замкнутых в кольцо триггеров, по которым под воздействием входных импульсов циркулирует одна или несколько кодовых единиц.

В кольцевом счетчике на трех *JK*-триггерах (рис. 7.13, *а*) прямой выход каждого предыдущего триггера соединен с входом *J* последующего триггера.




*а*

| Номер импульса | Выходы | | |
|---|---|---|---|
| | $Q_1$ | $Q_2$ | $Q_3$ |
| 0 | 1 | 0 | 0 |
| 1 | 0 | 1 | 0 |
| 2 | 0 | 0 | 1 |
| 3 | 1 | 0 | 0 |
| 4 | 0 | 1 | 0 |
| 5 | 0 | 0 | 1 |
| 6 | 1 | 0 | 0 |
| 7 | 0 | 1 | 0 |
| 8 | 0 | 0 | 1 |

*б*

**Рис. 7.13.** Кольцевой счетчик:
*а* — схема; *б* — таблица истинности

Тактовые входы $C$ всех триггеров объединены (узел «Вход») и на них поступают счетные импульсы. Перед началом счета первый триггер импульсным сигналом «Установка» устанавливается в состояние «1», остальные триггеры — в состояние «0». Этому состоянию счетчика соответствует 0 на выходе $Q_3$ последнего триггера. После отключения сигнала «Установка» начинается счет, и триггеры функционируют как обычные $JK$-триггеры. Поскольку на информационных входах первого триггера было $J = 0$ и $K = 1$, то в момент окончания первого входного импульса он перейдет в состояние логического 0. Второй триггер примет состояние «1», так как на его входах было $J = K = 1$. Третий триггер не изменит своего предыдущего состояния. Таким образом, кодовая 1 перешла с первого триггера на второй. По окончании каждого следующего входного импульса 1 будет перемещаться от предыдущего триггера к последующему, т. е. переходить по схеме вправо (см. рис. 7.13, *б* таблица истинности).

В широко применяемых в цифровой технике *десятичных* (*декадных*) счетчиках, цифры разрядов десятичного числа представляют в четырехразрядной двоичной форме, т. е. используют двоично-кодированную десятичную систему исчисления. При этом для каждого разряда десятичного числа включают четыре триггера, и если число десятичных разрядов $k$, то для регистрации чисел в счетчике необходимо задействовать $4k$ триггеров. Максимальное число записанных импульсов в счетчике составляет $N = 10^k - 1$. Десятичные счетчики находят широкое применение в случаях, когда число поступающих импульсов необходимо представить в привычной для человека десятичной системе.

### Шифраторы и дешифраторы

**Шифратор** (*иначе кодер*) — устройство, преобразующее десятичные числа в двоичную систему счисления. Пусть в шифраторе имеется $m$ входов, последовательно пронумерованных десятичными числами (0, 1, 2, 3, ..., $m - 1$), и $n$ выходов. Поступление сигнала на один из входов вызывает появление на выходах двоичного $n$-разрядного числа, соответствующего номеру возбужденного входа. Известно, что построить шифраторы с большим числом входов технически трудно, поэтому их используют для преобразования в двоичную систему счисления небольших десятичных чисел. Для преобразования больших десятичных чисел используются специальные методы.

Шифраторы широко применяют в устройствах ввода информации в цифровые системы. Такие устройства часто снабжают клавиатурой, каждая клавиша которой связана с определенным входом шифратора, и на его выходе воспроизводится двоичное число, соответствующее написанному на клавише символу. На рис. 7.14, *а* показано обозначение шифратора, преобразующего первые десять цифр десятичного счисления в двоичное представление. Символ $CD$ в обозначении образован из букв, входящих в английское слово «*CODER*». Слева на схеме показано 10 входов, представленных соответствующими десятичными числами 0, 1, 2, ..., 9. Справа показаны выходы шифратора: цифрами 1, 2, 4, 8 обозначены весовые коэффициенты двоичных разрядов, соответствующих отдельным его выходам.

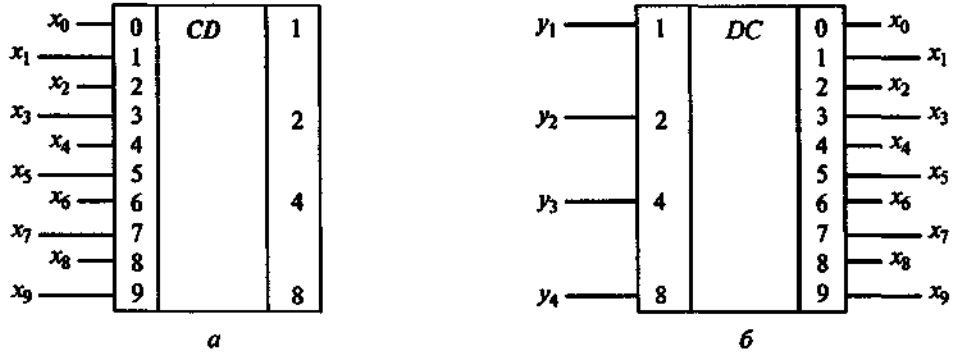



**Рис. 7.14.** Обозначение цифровых схем:
*а* — шифратора; *б* — дешифратора

**Дешифратор** (иначе *декодер*) предназначен для обратного преобразования двоичных чисел в небольшие по значению десятичные числа. Входы дешифратора служат для подачи двоичных чисел, а выходы последовательно нумеруются десятичными числами. Дешифраторы широко используют в цифровой и вычислительной технике. В частности, их применяют в системах, печатающих на бумаге выводимые из цифровой схемы текст или числа. В таких устройствах двоичное число, поступая на соответствующий вход дешифратора, вызывает появление сигнала на его определенном выходе.

На рис. 7.14, *б* приведено обозначение структурной схемы дешифратора. Символ *DC* образован из букв английского слова «*DECODER*». Слева представлены входы дешифратора: цифрами 1, 2, 4, 8 обозначены весовые коэффициенты двоичных разрядов. Справа на схеме показаны 10 выходов, пронумерованных десятичными числами 0, 1, 2, ..., 9, соответствующими отдельным комбинациям входного двоичного числа.

## 7.4. Мультивибраторы

Если в автогенераторе гармонических колебаний глубину положительной обратной связи теоретически приблизить к 100 % ($\beta = 1$), то синусоидальное выходное напряжение должно превратиться в сигнал прямоугольной формы. Такое физическое явление объясняется тем, что в колебательную систему за счет глубокой положительной ОС вводится энергии больше, чем в ней расходуется. Избытки этой энергии накапливаются в емкости и индуктивности и при достижении определенного значения, колебательная система автоматически резко переключается. При этом накопленная в реактивных элементах электрическая энергия начинает расходоваться в активной нагрузке, и автогенератор гармонических колебаний превращается в *релаксационный автогенератор* (от англ. *relaxation* — релаксация, высвобождение, в данном случае энергии).

### Автогенераторы релаксационных (импульсных) колебаний

Применение в релаксационных автогенераторах колебательных систем (контуров) возможно лишь при малом характеристическом сопротивлении реактивных элементов, т. е. при низкой добротности контуров (поскольку необходима широкая полоса частот для спектров импульсных сигналов). Поэтому в релаксационных автогенераторах вместо колебательных контуров надо использовать другую систему, способную работать с импульсными сигналами. Такой системой может быть, например, апериодическое электрическое звено с одним энергоемким (накопительным) элементом (чаще всего конденсатором или варикапом), например *RC*-цепь.

В отличие от автогенераторов гармонических колебаний, в которых процесс самовозбуждения и установления стационарного режима происходит в течение нескольких периодов колебаний, в релаксационных автогенераторах сразу устанавливается стационарное значение амплитуды. При этом ни на какой фиксированной частоте колебаний нельзя говорить об условии баланса фаз — оно выполняется в достаточно широкой полосе частот.

Упрощенная структурная схема генератора релаксационных колебаний состоит из источника питания, электронного ключа, накопителя энергии с цепью положительной обратной связи и поглощающей (активной) нагрузки (рис. 7.15). В этой схеме энергия источника питания через электронный (транзисторный) ключ поступает в накопитель энергии. Когда накапливаемая энергия достигнет определенного (достаточно высокого) уровня, ключ, благодаря наличию положительной ОС, автоматически переключит накопительный элемент на нагрузку. При уменьшении накопленной энергии до некоторого малого значения ключ вновь подключит источник питания к накопителю. Как правило, накопитель энергии является элементом времязадающей цепи и цепи ПОС автогенератора, определяющими длительность и период следования импульсов.

Следует отметить, что быстропротекающие (лавинообразные, регенеративные) процессы формирования фронтов и срезов выходных импульсов определяются параметрами транзисторов, а сравнительно медленные процессы, связанные с формированием длительности импульсов и паузы между ними, — в основном параметрами времязадающих цепей.

Рассмотрим некоторые виды релаксационных автогенераторов.




**Рис. 7.15.** Структурная схема автогенератора релаксационных колебаний

**Мультивибраторы.** Автогенератор прямоугольных импульсов, длительность и частота следования которых определяется параметрами времязадающих *RC*-цепей, называют *мультивибратором* (от уже упоминаемого *мульти* и лат. *vibro* — колеблю). Мультивибраторы имеют два *временно устойчивых* (*квазиустойчивых*) состояния, и их переключение обусловлено процессами, протекающими во времязадающих *RC*-цепях.

Симметричный мультивибратор — это классическая схема на дискретных усилительных элементах, содержит два транзисторных ключа, взаимно охваченных с входов на выходы петлями положительных ОС: одна из них образована резистивно-емкостной цепью $R_1C_1$, а другая — идентичной цепью $R_2C_2$ (рис. 7.16, *а*). Рассмотрим работу мультивибратора, полагая, что транзисторы неинерционные (т. е. достаточно высокочастотные) и переключаются мгновенно. Временны́е диаграммы напряжений на базах и коллекторах транзисторов показаны на рис. 7.16, *б*. Пусть в момент времени $t = 0$ анализируемый мультивибратор находится в квазиустойчивом состоянии, при котором транзистор $VT_1$ открыт и насыщен, а транзистор $VT_2$ закрыт. В схеме, в этом случае, имеет место следующее состояние: напряжение на коллекторе транзистора $VT_1$ $U_{к1} = 0$; напряжение на коллекторе транзистора $VT_2$ составляет $U_{к2} \approx E_к$; конденсатор $C_2$ разряжен, и напряжение на нем и на базе транзистора $VT_1$ будет $U_{б1} \approx 0$; конденсатор $C_1$ заряжен, и напряжение на нем и на базе транзистора $VT_2$ равно $U_{б2} \approx E_к$. Такое состояние мультивибратора не является устойчивым, и в схеме одновременно имеют место два процесса. Первый из них связан с зарядом конденсатора $C_2$ током, протекающим по цепи: источник $+E_к$, резистор $R_{к2}$, конденсатор $C_2$, цепь база–эмиттер транзистора $VT_1$. Второй процесс обусловлен перезарядом конденсатора $C_1$ током, протекающим по цепи: источник $+E_к$, резистор $R_1$, конденсатор $C_1$, цепь коллектор–эмиттер открытого транзистора $VT_1$. Поскольку в мультивибраторе сопротивления $R_1 > R_{к2}$ (соответственно $R_2 > R_{к1}$), то постоянная времени перезаряда конденсатора $C_1$: $\tau_п = R_1C_1$ больше постоянной времени заряда конденсатора $C_2$, т. е. $\tau_з = R_{к2}C_2$.

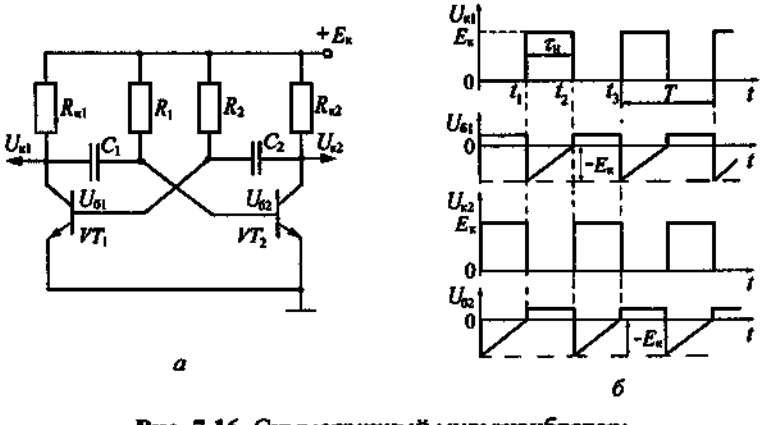



**Рис. 7.16.** Симметричный мультивибратор:
*а* — схема; *б* — временны́е диаграммы

В момент $t_1$ напряжение на конденсаторе $C_1$, а значит и на базе транзистора $VT_2$ станет равным нулю. Транзистор откроется, через него потечет коллекторный ток, вызывая уменьшение напряжения $U_{к2}$. Отрицательный скачок напряжения через конденсатор $C_2$ передастся на базу транзистора $VT_1$, приводя последний к призакрыванию и увеличению его коллекторного напряжения $U_{к1}$. Положительный скачок напряжения $U_{к1}$ через конденсатор $C_1$ поступит на базу транзистора $VT_2$ и еще больше его откроет. В схеме начнется лавинообразный процесс, заканчивающийся переключением мультивибратора в другое квазиустойчивое состояние, когда транзистор $VT_2$ открыт и насыщен, а транзистор $VT_1$ закрыт. При этом конденсатор $C_1$ будет разряжен, а конденсатор $C_2$ заряжен до напряжения $U_{б1} \approx -E_к$. На интервале $t_1 - t_2$ происходит перезаряд конденсатора $C_2$ и заряд конденсатора $C_1$, и в момент времени $t_2$ произойдет новое переключение мультивибратора. Далее процессы начнут периодически повторяться, а на коллекторах транзисторов будут формироваться выходные импульсы.

Длительность импульсов напряжения на коллекторах транзисторов соответствует *времени перезаряда* соответствующего конденсатора. Можно показать, что в симметричном мультивибраторе, у которого $C_1 = C_2 = C$, $R_1 = R_2 = R$ и $R_{к1} = R_{к2} = R_к$, длительность импульса определяется формулой

$$\tau_и = RC \ln 2 \approx 0{,}7RC. \qquad (7.9)$$

Очевидно, что период повторения импульсов

$$T = 2\tau_и \approx 1{,}4RC. \qquad (7.10)$$

**Пример 7.4.** Длительность выходного импульса мультивибратора (см. рис. 7.16, *а*) должна составлять $\tau_и = 14$ мкс. Определить параметры времязадающей цепочки.

Р е ш е н и е. Зададим приемлемое сопротивление резистора $RC$-цепи $R = 1$ кОм. Тогда, согласно формуле (7.9), находим, что емкость

$$C = \tau_и / (0{,}7R) = 0{,}02 \text{ мкФ}.$$

М у л ь т и в и б р а т о р ы  н а  л о г и ч е с к и х  э л е м е н т а х. Такие схемы мультивибраторов широко используют в качестве задающих генераторов в цифровых устройствах. Простейшая схема симметричного мультивибратора на логических элементах И-НЕ ($DD1$ и $DD2$), взаимно охваченных положительными ОС с помощью времязадающих цепочек $R_1C_1$ и $R_2C_2$, показана на рис. 7.17, *а*.

Положим, что на интервале времени $0 - t_1$ (рис. 7.17, *б*) мультивибратор находится в состоянии, когда элемент $DD1$ закрыт, и на его выходе логическая 1, а элемент $DD2$ открыт, и на его выходе логический 0. Конденсатор $C_2$ будет заряжаться выходным током элемента $DD1$, протекающим через резистор $R_2$ (ведь диод $VD2$ закрыт). Напряжение $U_{вх2}$, выделяемое на резисторе $R_2$, уменьшается по экспоненте (рис. 7.17, *в*) и определяется известной формулой

$$U_{вх2} = U_1 e^{-t/\tau_2}, \qquad (7.11)$$

где $\tau_2 = R_2C_2$ — постоянная времени цепи $R_2C_2$.

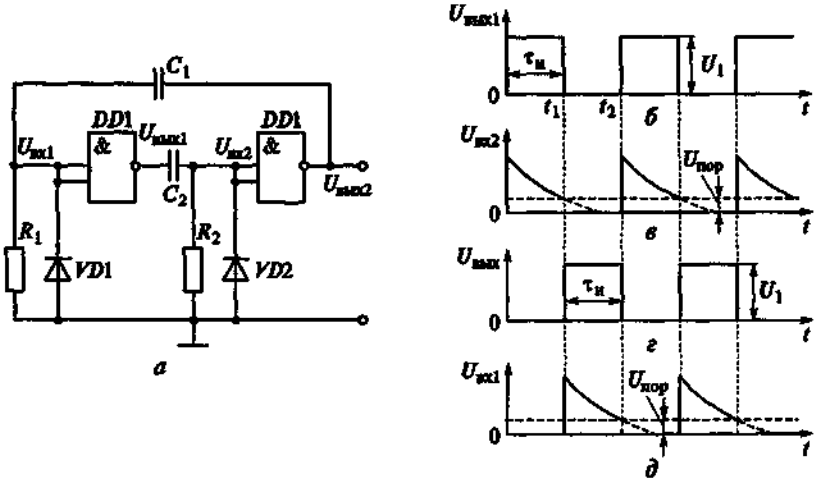



**Рис. 7.17.** Мультивибратор на логических элементах:
*а* — схема; *б – д* — временны́е диаграммы

В момент времени $t = t_1$ напряжение $U_{вх2}$ достигает *порогового* значения $U_{пор}$, при котором элемент *DD*2 переходит из состояния «0» в состояние «1», изменяя свое выходное напряжение. Скачок этого напряжения $U_{вых2} = U_1$ через конденсатор $C_1$ подается на вход элемента *DD*1, переводя его в состояние «0». Так как напряжение на выходе элемента *DD*1 при этом уменьшилось до нуля, то конденсатор $C_2$ начнет разряжаться. В результате данного процесса на резисторе $R_2$ возникнет отрицательное напряжение, открывающее диод *VD*2. Вследствие этого конденсатор $C_2$ быстро разрядится через открытый диод до нулевого напряжения. После окончания процесса мультивибратор в момент времени $t = t_1$ скачкообразно изменит свое состояние, при котором $U_{вых2} = U_1$ (рис. 7.17, *г*).

Одновременно, начиная с момента времени $t = t_1$, происходит заряд конденсатора $C_1$, и затем напряжение на входе логического элемента *DD*1 уменьшается. Когда в момент времени $t = t_2$ напряжение $U_{вх1} = U_{пор}$ (рис. 7.17, *д*), мультивибратор опять скачкообразно изменит свое состояние. Далее процессы в схеме мультивибратора начнут периодически повторяться.

Длительность импульса при $R_1 = R_2 = R$, $C_1 = C_2 = C$ определяют по формуле

$$\tau_и = RC \ln (U_1/U_{пор}). \qquad (7.12)$$

Схемотехнически современные мультивибраторы выполняются в виде отдельных интегральных микросхем.

Ждущий мультивибратор. В системах передачи информации для ослабления влияния случайных флуктуаций часто требуется из коротких импульсов получить более широкие, определенной, заданной длительности. Эта задача легко реализуется с помощью ждущего мультивибратора (иначе, одновибратора). Одновибратор является триггерной схемой и генерирует одиночный импульс под действием внешнего управляющего сигнала. При этом длитель-

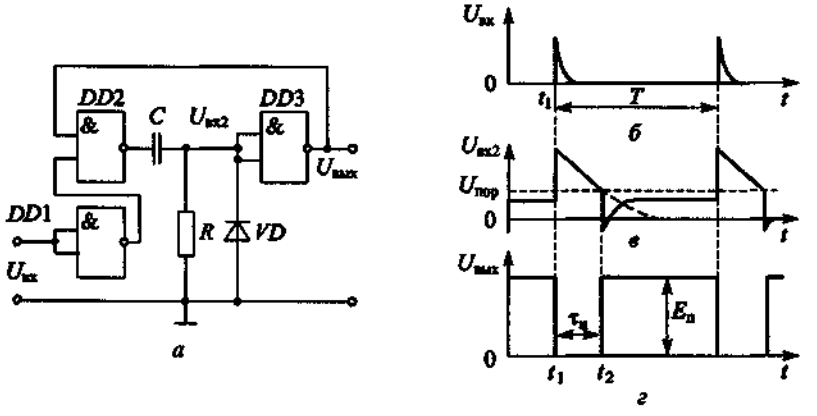



**Рис. 7.18.** Одновибратор на логических элементах:
*а* — схема; *б – г* — временны́е диаграммы

ность запускающего импульса особой роли не играет, лишь бы она была не больше длительности вырабатываемого одновибратором импульса. В одновибраторе одно из состояний является *локально устойчивым* (проще, устойчивым), а другое *квазиустойчивым*. В устойчивом состоянии одновибратор может находиться неограниченное время («ждать» запускающего импульса). Квазиустойчивое состояние наступает с приходом импульса запуска, и его длительность определяется постоянной времени формирующей *RC*-цепочки.

Наиболее просто одновибратор реализуют на базовых логических элементах (рис. 7.18, *а*). Для этого в рассмотренную схему мультивибратора вводят цепь запуска, выполненную на логическом элементе *DD*1. В исходном состоянии логический элемент *DD*3 закрыт, и напряжение на выходе одновибратора $U_{вых} = E_п$, т. е. равно уровню логической 1 (рис. 7.18, *б – г*). Логический элемент *DD*1 цепи запуска одновибратора в исходном состоянии закрыт, и на его выходе присутствует логическая 1.

Уровни логических 1 с выходов закрытых элементов *DD*1 и *DD*3 поступают на входы элемента *DD*2, поддерживая его в открытом состоянии. На выходе открытого элемента *DD*2 имеет место логический 0, и поэтому конденсатор *C* разряжен через этот элемент и открытый диод *VD* до нулевого потенциала. При поступлении в момент времени $t = t_1$ на вход одновибратора положительного импульса запуска (рис. 7.18, *б*) элемент *DD*1 открывается, а элемент *DD*2 переходит в закрытое состояние. На выходе закрытого элемента *DD*2 возникает положительный скачок напряжения, который через конденсатор *C* передается на объединенный вход логического элемента *DD*3. Этот элемент открывается, и на его выходе устанавливается логический 0.

После переключения одновибратора конденсатор *C* начинает разряжаться через сопротивление *R*, и напряжение $U_{вх2}$ на входе элемента *DD*3 снижается. В момент $t = t_2$, когда напряжение на конденсаторе $U_{вх} = U_{пор}$ (см. рис. 7.18, *в*), одновибратор переключится и вновь перейдет в устойчивое состояние. Длительность выходного импульса $\tau_и = t_2 - t_1$ зависит от времени разряда конденсатора *C*

на резистор $R$ и определяется выражением (7.9). Отметим, что в данной схеме выходной импульс имеет низкий потенциал (т. е. уровень логического 0) и для получения высокого потенциала (логической 1) необходимо на выходе одновибратора включить инвертор.

### Мультивибраторы на операционных усилителях

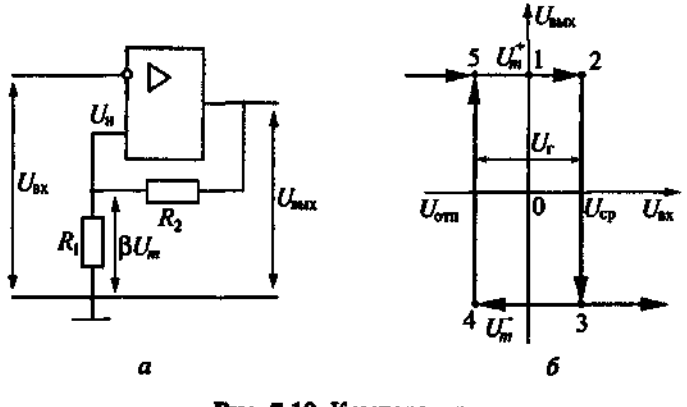
**Компаратор.** Основой мультивибратора на ОУ служит *компаратор* — устройство, осуществляющее сравнение двух напряжений. В простейшем компараторе входное напряжение сравнивают с некоторым *опорным*, в качестве которого используют часть выходного напряжения (рис. 7.19, *а*). В результате на инвертирующий вход ОУ поступает входное напряжение, а на неинвертирующий вход — опорное напряжение, снимаемое с делителя $R_1$, $R_2$. ОУ охвачен положительной обратной связью по неинвертирующему входу, и выходное напряжение скачком изменяет свою полярность при сравнении входного и опорного напряжений.

Принцип действия компаратора рассмотрим с помощью передаточной характеристики — зависимости выходного напряжения от входного (рис. 7.19, *б*). Положим, что входное напряжение $U_{вх} = 0$, а выходное $U_{вых} = U_m^+$ (точка 1 на рис. 7.19, *б*). Напряжение на неинвертирующем входе при этом будет

$$U_н = \beta U_m^+, \tag{7.13}$$

где $\beta = R_1/(R_1 + R_2)$ — коэффициент передачи резистивной цепи $R_1R_2$ положительной ОС в компараторе.

Если входное напряжение компаратора положительно (больше нуля) и увеличивается, то при его сравнении с опорным, равным *напряжению срабатывания* $U_{ср} = \beta U_m^+$, устройство переключится и произойдет скачкообразное изменение выходного напряжения со значения $U_m^+$ на $U_m^-$ (переход от точки 2 к точке 3 на рис. 7.19, *б*). Дальнейшее увеличение входного напряжения не изменит состояния компаратора, и напряжение на неинвертирующем входе ОУ будет постоянным: $U_н = \beta U_m^-$.




*а* *б*

**Рис. 7.19.** Компаратор:
*а* — схема; *б* — передаточная характеристика

При уменьшении входного напряжения до уровня опорного (*напряжения отпускания*) $U_{вх} = U_{отп} = \beta U_m^-$, произойдет скачкообразный возврат компаратора в исходное состояние. Выходное напряжение при этом изменится с $U_m^-$ на $U_m^+$ (переход от точки 4 к точке 5 на рис. 7.19, *б*). Таким образом, передаточная характеристика компаратора имеет вид *петли гистерезиса*. Компаратор обладает *триггерным* (переключающим) *эффектом* и его называют *триггером Шмитта* (*Schmit trigger*; часто *триггер Шмидта*; назван в честь О.Ю. Шмидта).

Сумма напряжений срабатывания и отпускания

$$U_г = U_{ср} + U_{отп} = 2\beta U_m \qquad (7.14)$$

является *напряжением гистерезиса*.

Напряжение гистерезиса вводится для повышения помехоустойчивости, что позволяет устранить «*дребезг*» триггера, т. е. случайное его переключение напряжением помех при отсутствии входного сигнала.

В компараторе на ОУ амплитуда выходного напряжения практически равна напряжению питания: $U_{вых} = U_m^{\pm} = \pm U_п$. Компараторы применяют для формирования сигналов прямоугольной формы из различных видов непрерывных сигналов. В частности, при подаче на вход схемы компаратора синусоидального напряжения (рис. 7.20), на его выходе формируется симметричное прямоугольное колебание — *меандр*. Пусть в момент $t = 0$ напряжение на выходе компаратора $U_{вых} = U_m^+$. В таком состоянии компаратор будет находиться, пока амплитуда входного напряжения $U_{вх} < U_{ср}$. В момент $t = t_1$ входное напряжение станет $U_{вх} = U_{ср}$, и компаратор переключится. При этом выходное напряжение $U_{вых}$ скачком изменится со значения $U_m^+$ на значение $U_m^-$. В момент $t = t_2$ входное напряжение станет равным $U_{отп}$, и произойдет новое переключение компаратора.

**Пример 7.5.** Рассчитать напряжения срабатывания и гистерезиса компаратора на ОУ (см. рис.7.19, *а*), если сопротивления $R_1 = 1$ кОм, $R_2 = 9$ кОм, а напряжение $U_m = U_п = 10$ В.

Р е ш е н и е. Коэффициент передачи по напряжению петли ПОС этого компаратора $\beta = R_1/(R_1+R_2) = 0{,}1$. Тогда, согласно соотношениям (7.13) и (7.14), имеем $U_{ср} = \beta U_m = 1$ В; $U_г = 2\beta U_m = 2$ В.

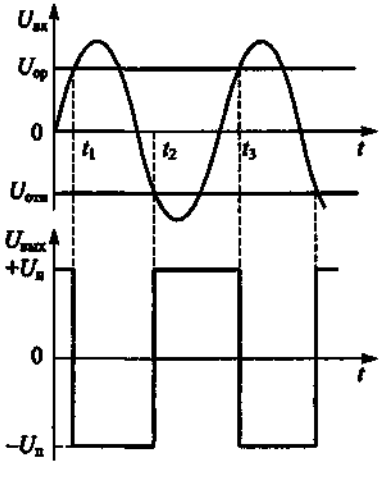



**Рис. 7.20.** Формирование меандра из синусоиды компаратором

**Мультивибратор на ОУ**. Для превращения компаратора в мультивибратор вводят времязадающую *RC*-цепь отрицательной ОС (рис. 7.21, *а*). После включения питания схема, благодаря положительной ОС, остается в одном из квазиустойчивых состояний.

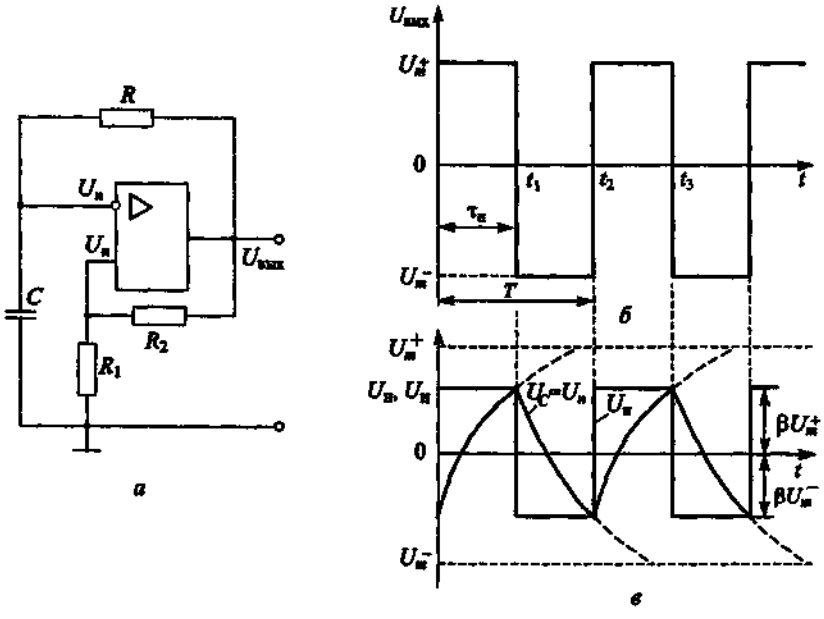



**Рис. 7.21.** Мультивибратор на ОУ:
*а* — схема; *б, в* — временны́е диаграммы напряжений на выходе и входах

Положим, что в момент времени $t = 0$ напряжение на выходе стало равным $U_m^+$, напряжение на неинвертирующем входе ОУ $\beta U_m^+$, а конденсатор $C$ заряжен до некоторого отрицательного напряжения $U_C = \beta U_m^-$ (рис. 7.21, *в*). С этого момента под действием выходного напряжения конденсатор $C$ перезаряжается через резистор $R$, и напряжение на нем будет стремиться к значению $U_m^+$. В момент $t = t_1$ напряжение на конденсаторе, а значит, и на инвертирующем входе $U_и$ станет равным напряжению на неинвертирующем входе $U_н$, что, как в компараторе, вызовет скачкообразное изменение полярности выходного напряжения с $U_m^+$ на $U_m^-$. С момента $t = t_1$ конденсатор вновь начинает перезаряжаться, и напряжение на нем будет повышаться до $U_m^-$. Однако в момент $t = t_2$ произойдет сравнение амплитуд напряжений на инвертирующем и неинвертирующем входах ОУ (рис. 7.21, *в*), и мультивибратор вновь переключится. Далее процессы в схеме будут периодически повторяться.

Длительность импульса на выходе мультивибратора определяется временем перезаряда конденсатора от уровня $\beta U_m^+$ до $\beta U_m^-$. Можно показать, что

$$\tau_и \approx 2\beta RC. \tag{7.15}$$

Тогда период повторения

$$T = 2\tau_н \approx 4\beta RC. \tag{7.16}$$

Заменив резистор $R$ цепью из двух параллельно включенных диодов (диод соединен последовательно с резистором), получим *несимметричный мультивибратор* с разной длительностью положительных и отрицательных импульсов.

## 7.5. Схемотехника импульсных и цифровых детекторов

В современных радиоэлектронных системах приемные устройства представляют собой весьма сложную структуру, реализующую обработку сигналов, как правило, цифровыми методами. Одними из основных их элементов являются импульсные и цифровые детекторы, которые выполняют на логических и цифровых интегральных элементах.

### Импульсный фазовый детектор на логических элементах

В импульсном фазовом детекторе на логических элементах ФМ-колебание преобразуется в импульсное напряжение, скважность импульсов которого зависит от фазы входного сигнала. На рис. 7.22 показаны схема фазового детектора и диаграммы, поясняющие его работу.

Детектор содержит два входа: на один вход поступает ФМ-сигнал $u_{ФМ}(t)$ (рис. 7.22, *б*), а на другой — импульсное опорное напряжение $u_{оп}(t)$ (рис. 7.25, *г*). ФМ-сигнал и опорное напряжение подаются на *формирующие устройства* $УФ_1$ и $УФ_2$ соответственно, в качестве которых и используются компараторы. На выходах УФ возникают последовательности прямоугольных импульсов $u_1$ и $u_2$ (рис. 7.22, *в*, *д*), длительности которых равны соответственно полупериодам входных колебаний — ФМ-сигнала и опорного напряжения.

Сформированные импульсные напряжения $u_1$ и $u_2$ поступают на логический элемент И-НЕ. Импульсное колебание $u_и$ амплитудой $U_0$ на выходе этого элемента формируется только при одновременном действии напряжений $u_1$ и $u_2$ (рис. 7.22, *е*). ФНЧ выделяет из этого напряжения постоянную составляющую, амплитуда которого $U_c$ определяют формулой (можно вывести самостоятельно)

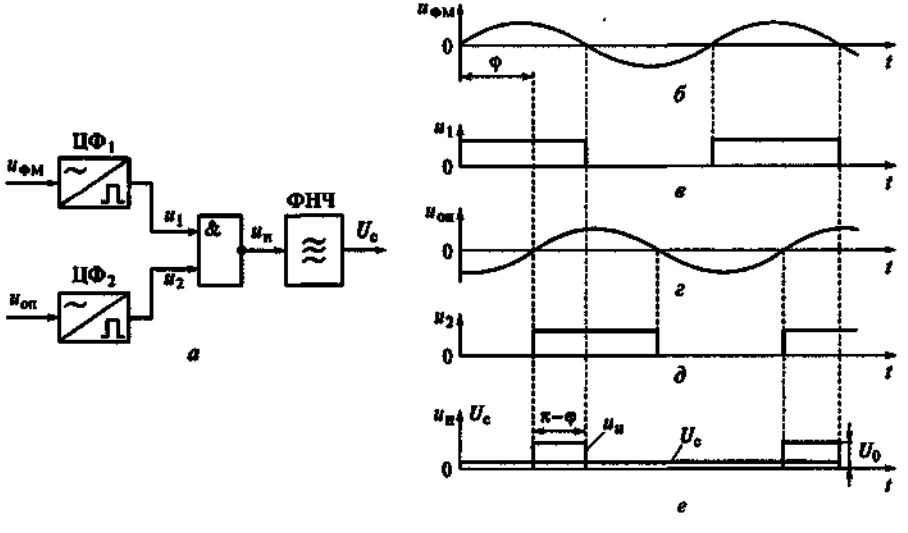



**Рис. 7.22.** Импульсный фазовый детектор на логических элементах:
*а* — схема; *б – е* — временны́е диаграммы

$$U_c = \frac{U_0|\pi - \varphi|}{2\pi} = \frac{U_0}{2}\left|1 - \frac{\varphi}{\pi}\right|. \qquad (7.17)$$

Согласно (7.17), выходное напряжение $U_c$ фазового детектора линейно зависит от сдвига фазы ФМ-сигнала относительно фазы опорного напряжения.

### Цифровой фазовый детектор на *JK*-триггере

Проанализируем процессы детектирования так называемого *знакового сигнала*, представляющего собой последовательность потенциальных импульсов (единиц) и пауз (нулей). Аналогами таких колебаний являются сигналы с ШИМ, или ФИМ.

Рассмотрим фазовое детектирование периодической последовательности прямоугольных импульсов. Заметим, что задержка на некоторое время $\tau$ периодического сигнала с периодом следования $T$ эквивалентна повороту его фазы на определенный угол $\varphi = 2\pi\tau/T$. Простейшая схема *цифрового фазового детектора* (ЦФД) приведена на рис. 7.23, *а*. ЦФД выполнен на *JK*-триггере, к выходу которого подключен фильтр нижних частот в виде интегрирующей *RC*-цепи. На рис. 7.23, *б* показаны временны́е диаграммы знакового сигнала $u_{ФМ}$ (отражающего ФМ-колебание), тактовой последовательности импульсов $u_{оп}$ (т. е. опорного напряжения, с фазой которого сравнивается фаза знакового сигнала) и сигнал $U(t)$ на выходе ЦФД. Импульсный сигнал $Q$ на выходе *JK*-триггера соответствует его таблице истинности (см. рис. 7.11, *б*). Как следует из временны́х диаграмм напряжений, длительность выходных импульсов триггера пропорциональна временно́му (или фазовому) сдвигу между сигналами $u_{ФМ}$ и $u_{оп}$. Напряжение на выходе ЦФД $U(t)$ образуется сглаживанием импульсов $Q$ в ФНЧ.

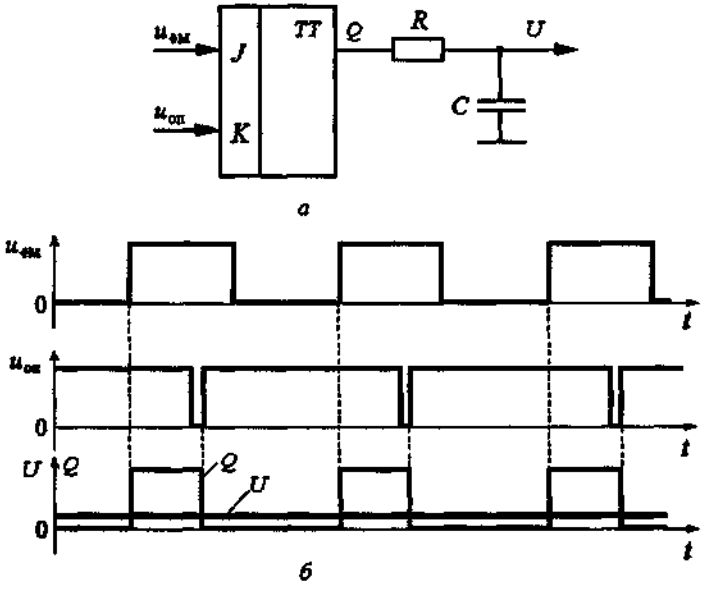



**Рис. 7.23.** Цифровой фазовый детектор на *JK*-триггере:
*а* — схема; *б* — временны́е диаграммы

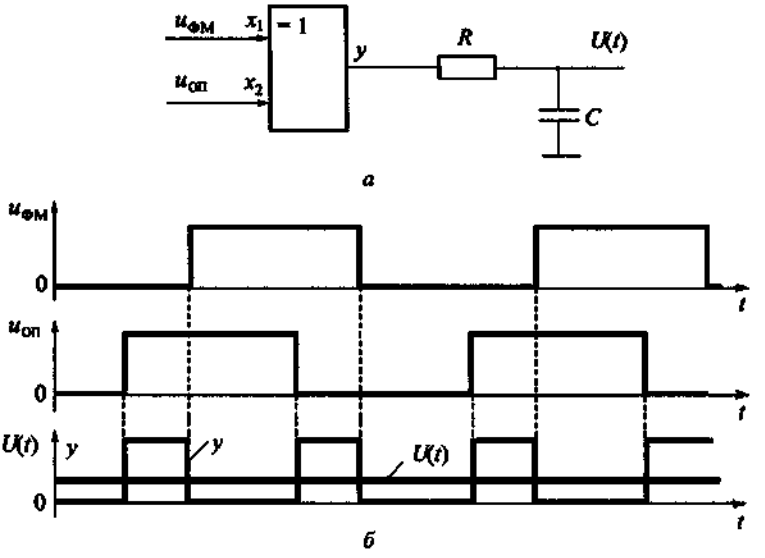



**Рис. 7.24.** Цифровой фазовый детектор на элементе «Исключающее ИЛИ»:
*а* — схема; *б* — временны́е диаграммы

Цифровые фазовые детекторы можно построить не только на интегральном *JK*-триггере, но и на других логических схемах: элементе «Исключающее ИЛИ», *RS*-триггере и пр. С помощью данных схем достаточно просто удается получить длительность выходных импульсов, прямо пропорциональную временно́й задержке между модулированным $u_{ФМ}$ и опорным $u_{оп}$ сигналами, после чего сгладить эти импульсы в ФНЧ.

На рис. 7.24, *а* в качестве примера приведена схема ЦФД на элементе «Исключающее ИЛИ» (сумматор по модулю 2). Временны́е диаграммы работы ЦФД показаны на рис. 7.24, *б*. В схеме импульсное напряжение *y*, сформированное в схеме «Исключающее ИЛИ», подается на ФНЧ. Напряжение $U(t)$ на выходе ФНЧ пропорционально сдвигу ФМ-сигнала относительно опорного $u_{оп}$.

Этот детектор более помехоустойчив, чем ЦФД на триггере. Дело в том, что триггеры срабатывают по фронтам импульсов, поэтому в случае «дребезга» этих фронтов выходной сигнал ЦФД может оказаться существенно искаженным. Напротив, схема «Исключающее ИЛИ» работает по уровням входных сигналов, поэтому короткие шумовые или помеховые импульсы, приводящие к «дребезгу» фронтов этих сигналов, не могут заметно исказить выходное напряжение.

### Цифровой частотный детектор

Процедура цифрового частотного детектирования знакового сигнала оказывается очень простой. Схема цифрового частотного детектора (ЦЧД) приведена на рис. 7.25, *а*. Она включает в себя формирователь импульсов нормированной длительности на логических элементах И-НЕ, работающий по одному или обоим фронтам импульсного знакового сигнала, и фильтр нижних частот. Временны́е диаграммы напряжений в схеме показаны на рис. 7.25, *б*.

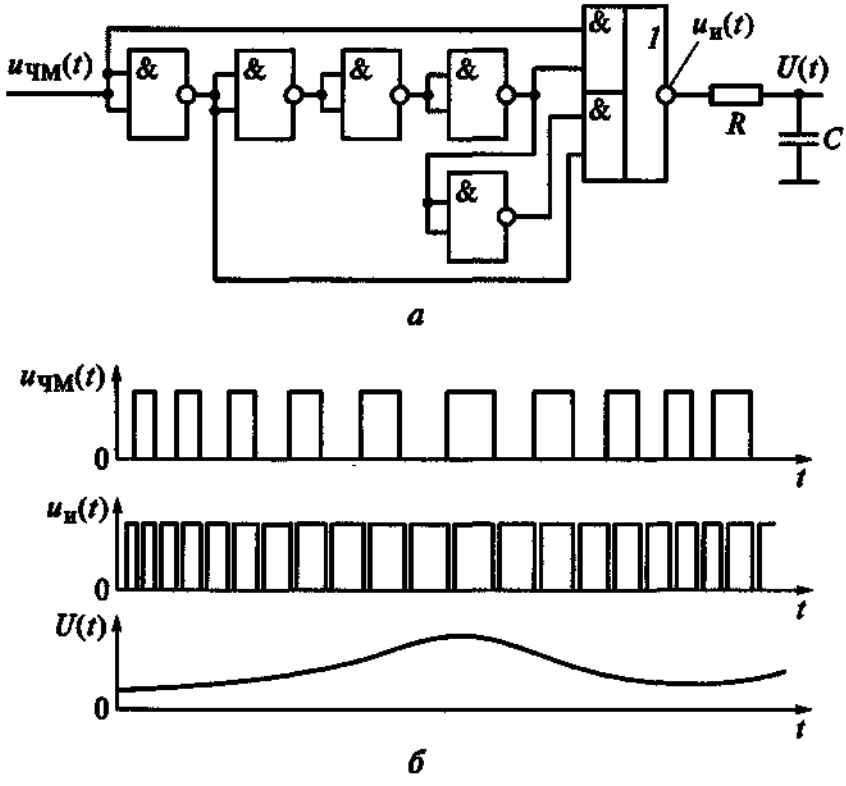



**Рис. 7.25.** Цифровой частотный детектор на элементах «И-НЕ»:
*а* — схема; *б* — временны́е диаграммы

Формирование цифровым частотным детектором импульсов нормированной длительности может выполняться несколькими способами. В данной схеме детектора это формирование основано на задержке распространения импульсно-модулированного сигнала $u_{ЧМ}(t)$ по цепочке последовательно включенных логических элементов И-НЕ.

Верхняя ветвь формирователя обеспечивает генерирование импульса по переднему фронту, а нижняя — по заднему. Сформированные по переднему и заднему фронтам последовательности импульсов подаются на схему И-ИЛИ-НЕ. Выходной сигнал $u_{и}(t)$ этого логического элемента поступает на ФНЧ. Нетрудно заметить, что низкочастотная составляющая выходного напряжения оказывается пропорциональной частоте входного сигнала.

Отметим, что рассмотренный ЦЧД не обязательно требует перевода непрерывного (модулированного) сигнала в знаковый, а последнего в импульсный. Часто непрерывный сигнал преобразуется непосредственно в импульсный. Такой детектор принято называть *импульсно-счетным*.

## Контрольные вопросы

1. В чем сходство и различие импульсных, цифровых и логических сигналов?
2. Запишите основные законы и аксиомы алгебры Дж. Буля.
3. Изобразите простейшие реализации логических схем ИЛИ, И, НЕ, ИЛИ-НЕ, И-НЕ и поясните принцип их действия. Представьте их таблицы истинности.
4. Изобразите схему симметричного триггера на дискретных элементах и поясните принцип его действия.
5. Сравните схемы *RS*-триггеров на элементах ИЛИ-НЕ и И-НЕ и укажите, какими сигналами осуществляется их переключение.
6. Представьте схемы *D*- и *T*-триггеров на базовых логических элементах.
7. Объясните назначение и принцип действия *JK*-триггера.
8. На каком принципе строятся делители частоты и счетчики?
9. Изобразите схему кольцевого счетчика и поясните его работу.
10. Объясните принцип действия классической схемы симметричного мультивибратора на транзисторах.
11. Изобразите компаратор на ОУ и поясните принцип его работы.
12. Изобразите мультивибратор на ОУ и поясните принцип его работы.
13. Поясните принцип действия одновибратора на ОУ.
14. Изобразите мультивибратор и одновибратор на логических элементах.
15. Параметрами каких элементов и процессов определяются длительности импульсов мультивибратора?
16. Параметрами каких элементов и процессов определяются длительности импульсов одновибратора?
17. Поясните принцип действия импульсного детектора.
18. Изобразите схемы и поясните принцип работы цифровых фазовых детекторов с аналоговым выходом.
19. Изобразите схему и поясните принцип работы цифрового частотного детектора.

# ЧАСТЬ III

# ЦИФРОВАЯ ОБРАБОТКА, ПЕРЕДАЧА И ПОМЕХОУСТОЙЧИВЫЙ ПРИЕМ СИГНАЛОВ

## Глава 8. ЦИФРОВАЯ ОБРАБОТКА СИГНАЛОВ

Бурное развитие цифровых технологий изменило смысл понятия «радиотехника», поскольку особое значение приобрели дискретные и цифровые сигналы. Это вызвано широким внедрением связи и усовершенствованием способов цифровой обработки сигналов высокоскоростными сигнальными микропроцессорами. *Цифровая обработка сигналов* (ЦОС; *digital signal processing — DSP*) — область науки и техники, изучающая общие для разных дисциплин алгоритмы и средства обработки сигналов на основе численных методов с использованием цифровой техники. Сущность цифровой обработки состоит в том, что сигнал преобразуется в последовательность *чисел*, которая затем подвергается *математическим преобразованиям* в вычислительном устройстве. Фундаментальной базой этого служит теорема Котельникова (*теорема отсчетов*). Большую роль в радиотехнике сыграли алгоритмы дискретного и быстрого преобразований Фурье, позволяющие с высокой скоростью определять спектральные характеристики сигналов. Заметные успехи достигнуты в создании и применении для обработки цифровых сигналов цифровых фильтров.

### 8.1. Общие сведения

Большинство реальных сигналов являются непрерывными функциями (если пренебречь квантовыми эффектами). Поэтому в течение длительного времени в радиотехнике использовали преимущественно аналоговые сигналы. С их помощью удавалось решать достаточно сложные проблемы, имеющие место в связи, радиолокации, телевидении и т. д. Аналоговые сигналы сравнительно просто генерировать, усиливать, преобразовывать и обрабатывать с помощью электронных устройств непрерывного действия — ламповых и транзисторных приборов.

В начале 60-х гг. XX в. стали разрабатываться радиосистемы, базирующиеся на обработке дискретных сигналов. Эти системы позволяют располагать в интервалах времени между отсчетами одних дискретных сигналов отсчеты других дискретных сигналов. В результате появилась возможность по одной радиолинии передавать одновременно несколько сообщений, осуществляя многоканальную связь с разделением каналов по времени. Дальнейшим развитием техники дискретных сигналов являются цифровые радиотехнические системы. Существенное влияние на разработку цифровых систем оказало расширение функциональных возможностей микросхемотехники. Широкое внедрение цифровой обработки сигналов в радиотехнике было обусловлено увеличением дальности связи и требованием высокой ее помехоустойчивости. Цифровые методы приобрели также первостепенное значение в обработке, преобразовании и хранении двумерных сигналов и изображений (телевизионных и пр.). Следует также отметить высокую экономическую эффективность цифровых устройств и решение ряда задач, практически не выполнимых на базе аналоговой техники. Последнее касается, в частности, простоты перепрограммирования цифровых устройств на реализацию других функций. Еще одно важнейшее преимущество цифровой техники связано с принципиальной возможностью устранения накопления шумов путем восстановления формы импульсов в каскадных системах, например связных, ретрансляционных, спутниковых и т. д. Однако за все это приходится платить более широкой полосой частот, занимаемых цифровыми сигналами.

## 8.2. Цифровое представление сигналов. Теорема Котельникова

Как уже отмечалось, при переходе от аналогового сигнала к цифровому осуществляют три преобразования: дискретизацию по времени, квантование по уровню и кодирование (оцифровку). Это представление сигналов называют *аналого-цифровым преобразованием.*

*Дискретизация* — процесс превращения во времени непрерывного сигнала в дискретную последовательность *отсчетов* (*выборок*), следующих с некоторым временны́м интервалом $\Delta t = T$, и по которым с заданной точностью можно вновь восстановить исходный сигнал. Точки, в которых определены дискретные значения сигнала, называют *узловыми.*

Возникающую при этом задачу вычисления значений функции в промежутках между узловыми точками называют восстановлением или *интерполяцией* (*interpolation* — восстановление значения функции в промежуточной точке по известным ее значениям в соседних точках), т. е. для восстановления непрерывного сигнала нужно проинтерполировать цифровой сигнал «между отсчетами» (узловыми точками) специальными функциями. Эти функции могут и не иметь аналитических выражений и представляться табличными числовыми значениями $u_i$ для дискретных значений независимых переменных $t_i$. Аналогичными таблицами $\{u_i, t_i\}$ можно представлять и экспериментальные данные.

Задачу вычисления за пределами семейства узловых точек вперед или назад по переменным называют *задачей экстраполяции* (от *экстра* и лат. *polio* — приглаживаю, выправляю, изменяю — особый тип аппроксимации, при котором функция аппроксимируют не *между* заданными значениями, а *вне* заданного интервала) или *прогнозирования*. Решение этих задач обычно выполняют также с использованием аппроксимирующих функций. При кусочно-линейной интерполяции определение дополнительных точек выполняют по линейной зависимости. Графически это означает соединение узловых точек отрезками прямых.

Процесс дискретизации можно реализовывать по-разному, а наиболее удобной является операция *выборки-хранения* (*sample-and-hold*). В этом случае при дискретизации непрерывного сигнала формируется множество его *отсчетных значений* соответствующей амплитуды (в виде теоретически бесконечно коротких импульсов), взятых через определенный, заданный интервал времени $\Delta t$ (рис. 8.1). При этом амплитуда $k$-го отсчета $u_{Tk}(t)$ равна значению непрерывного сигнала $u(t)$ в дискретный момент времени $t = k\Delta t$ (рис. 8.1, *а*, *б*). Фактически результатом процесса дискретизации является сигнал в форме сигнала с *амплитудно-импульсной модуляцией*.

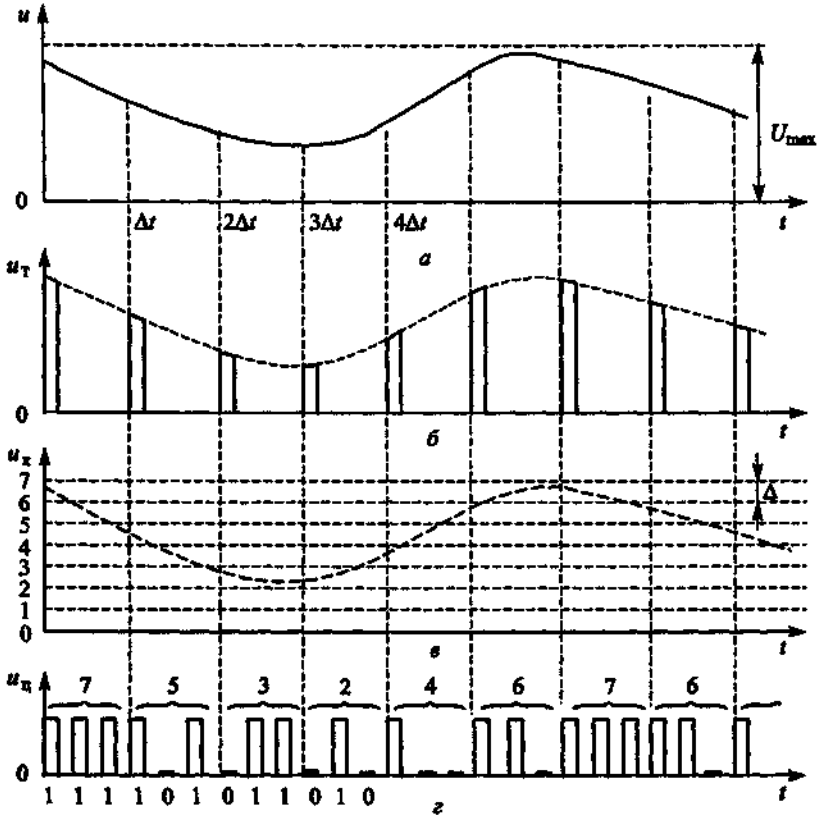



**Рис. 8.1.** Аналого-цифровое преобразование сигнала:
*а* — аналоговый; *б* — дискретизированный; *в* — квантованный; *г* — цифровой

Такое название возникло потому, что выходной дискретизированный сигнал можно описать как последовательность импульсов с амплитудами, определяемыми выборками входного сигнала.

Для представления дискретных отсчетов цифровыми сигналами (*кодирования*) их предварительно квантуют по уровню напряжения. В процессе квантования весь диапазон возможных изменений амплитуд аналогового сигнала от 0 до $U_{max}$ (или от $U_{min}$ до $U_{max}$ в случае разнополярного сигнала) разбивается на определенное число одинаковых или различных фиксированных уровней напряжения $\Delta$, называемых *шагом квантования* (рис. 8.1, *в*). Различают *равномерное* (уровень $\Delta$ — постоянная величина) и *неравномерное* ($\Delta$ — переменная величина) квантование. Неравномерное квантование применяется достаточно редко, в специфических случаях, например при большом динамическом диапазоне квантуемой величины.

Процесс квантования дискретного сигнала описывается уравнением

$$u_{T_k}(t) = u_k(t) = u_k = N(t_k)\Delta\,\sigma(t - t_k), \tag{8.1}$$

где $u_{T_k}(t)$ — квантованный сигнал; $N = N(t_k)$ — число квантов; $\sigma(t - t_k)$ — единичная функция.

При равномерном квантовании каждому фиксированному уровню сигнала $u_k(t)$ присваивают определенное значение в форме числа цифрового кода. С точки зрения удобства технической реализации и обработки используют (это уже отмечалось) двоичные цифровые коды, составленные из $n$ ($n$ — целое натуральное число) разрядов, каждый из которых представлен 1 — импульсом или 0 — паузой. Общее число уровней квантования составляет $2^n$. Значение одного шага квантования (рис. 8.1, *в*) связано с числом разрядов двоичного кода формулой

$$\Delta = U_{max}/2^n. \tag{8.2}$$

На рис. 8.1 в качестве примера показано квантование простейшего однополярного аналогового сигнала на $2^n = 2^3 = 8$ уровней, что соответствует трехразрядному коду. На временно́й оси трехразрядный код $u_ц$ представляется различными комбинациями из трех импульсов и пауз. Каждый из этих импульсов на одном интервале дискретизации $\Delta t$ в соответствии с занимаемой позицией, отвечающей разряду $2^2$, $2^1$, $2^0$, имеет множитель 1 или 0. Наличие на данном интервале дискретизации импульсов с тем или иным множителем определяет уровень квантования. В частности, при кодировании значения напряжения $u(0) = 7$ каждый разряд имеет множитель 1, чему соответствует присутствие всех трех импульсов на интервале дискретизации — 111. Аналогично значение $u(2\Delta t) = 3$ представлено двоичным кодом 011, т. е. паузой и двумя импульсами.

Отметим, что преобразуемый в цифровую форму непрерывный сигнал может иметь и положительное и отрицательное значение. В этом случае максимальному значению отрицательного потенциала сигнала $u_ц$ соответствует нулевой двоичный код, т. е. 000. Проведя «нумерацию» уровней квантования, можно передавать не сами уровни, а их значения по шкале уровней в двоичном виде.

В цифровой технике часто используют один из методов кодирования отсчетов непрерывного сигнала, при котором разрядные импульсы и паузы следуют в одном интервале дискретизации $\Delta t$ без временны́х промежутков. При этом несколько очередных соседних импульсов могут сливаться в один более широкий, суммарный импульс по оси времени.

Заметим, что в процессе квантования каждый отсчет не всегда точно совпадает с одним из фиксированных значений непрерывного сигнала, поэтому его увеличивают или уменьшают до ближнего заданного уровня (см. рис. 8.1, *в*). Этот процесс аналогичен процедуре округления чисел. Если амплитуда отсчета сигнала в пределах двух соседних разрешенных значений превышает половину шага квантования $\Delta/2$, ее значение увеличивается в большую сторону, если меньше половины шага квантования — в меньшую сторону. В результате этого при восстановлении аналогового сигнала из дискретного возникают систематические погрешности, характеризующие степень отклонения полученного сигнала от исходного. Эти погрешности квантования играют заметную роль в теории дискретных и цифровых сигналов и будут специально рассмотрены далее.

### Теорема Котельникова (теорема отсчетов)

Проблема дискретизации широко освещена в научной и учебной литературе. Считается, что первыми фундаментальными трудами в этой области были работа В.А. Котельникова «О пропускной способности «эфира» и проволоки в электросвязи», вышедшая в 1933 г., и статья К. Шеннона «Связь при наличии шума», опубликованная в 1949 г. Статья К. Шеннона была написана на основе работы математика Е.Т. Уттакера «Функции, представляемые распространением теории интерполяции», вышедшей в 1915 г. В математическом плане проблема представления функции ее отдельными значениями и восстановление ее при помощи интерполяции начала решаться еще в XVIII в. в работах О. Коши, Лапласа и т. д., а позднее была описана повторно Д. Карсоном и Р. Хартли.

Для того чтобы восстановить исходный непрерывный сигнал из дискретизированного с малыми искажениями (погрешностями), необходимо рационально выбрать шаг дискретизации. Поэтому при преобразовании аналогового сигнала в дискретный обязательно возникает вопрос о величине шага дискретизации $\Delta t$. Интуитивно нетрудно понять следующую разумную идею. Если аналоговый сигнал обладает низкочастотным спектром, ограниченным некоторой верхней частотой $F_в$ (т. е. функция $u(t)$ имеет вид плавно изменяющейся кривой, без резких изменений амплитуды), то вряд ли на некотором небольшом временно́м интервале дискретизации $\Delta t$ эта функция может существенно изменяться по амплитуде.

Совершенно очевидно, что точность восстановления аналогового сигнала по последовательности отсчетов зависит от величины интервала дискретизации $\Delta t$. Чем он короче, тем меньше будет отличаться функция $u(t)$ от плавной кривой, проходящей через точки отсчетов. Однако с уменьшением интервала дискретизации $\Delta t$ существенно возрастает сложность и объем обрабатывающей аппаратуры. При достаточно большом интервале дискретизации $\Delta t$ возрастает вероятность искажения или потери информации при восстановлении аналогового сигнала.

Оптимальный интервал дискретизации устанавливается теоремой Котельникова (другие названия — *теорема отсчетов, теорема Шеннона, теорема Найквиста — Nyquist theorem, теорема дискретизации — sampling theorem, теорема о выборках — sampling theorem*; в зависимости от пристрастий автора встречаются все мыслимые комбинации этих имен).

Согласно одной из наиболее известных и простых интерпретаций теоремы Котельникова, *произвольный сигнал $u(t)$, спектр которого ограничен некоторой частотой $F_в$, может быть полностью восстановлен по последовательности своих отсчетных значений, следующих с интервалом времени*

$$\Delta t = \frac{1}{2F_в}. \qquad (8.3)$$

В формулировке В.А. Котельникова (1933 г.) данной теоремы имеются такие слова: «... можно непрерывные сигналы передавать с любой точностью при помощи *чисел*, следующих друг за другом через $0{,}5/F_в$ секунды», т. е. по существу впервые им был поставлен вопрос о цифровом способе передачи сигналов. Нередко специалисты приводят частную интерпретацию теоремы отсчетов, верную для непрерывных сигналов, полоса частот которых начинается с нулевой частоты, «... чтобы не присутствовали частоты большие, нежели половина частоты дискретизации».

Интервал дискретизации $\Delta t$ и верхнюю частоту $F_в = F_д$ (8.3) в радиотехнике часто называют соответственно *интервалом* (*барьером*) и *частотой Найквиста* (*Nyquist frequency*; в отечественной литературе интервал $\Delta t$ — *интервалом Котельникова*).

Аналитически теорема Котельникова представляется рядом:

$$u(t) = \sum_{k=-\infty}^{\infty} u(k\Delta t)\frac{\sin\omega_в(t - t\Delta t)}{\omega_в(t - k\Delta t)}, \qquad (8.4)$$

где $k$ — номер отсчета; $u(k\Delta t)$ — значения непрерывного сигнала $u(t)$ в точках отсчета; $\omega_в = 2\pi F_в = \pi/\Delta t$ — верхняя частота спектра сигнала.

Для доказательства теоремы Котельникова (в современных учебных изданиях приводится, как правило, вывод теоремы отсчетов по Шеннону, который несколько отличен от вывода по Котельникову) рассмотрим произвольный непрерывный сигнал $u(t)$, спектральная плотность $S(\omega)$ которого сосредоточена в полосе частот $-\omega_в \le \omega \le \omega_в$ (сплошная линия на рис. 8.2).

Мысленно дополним диаграмму спектральной плотности исходного сигнала $S(\omega)$ симметрично формами графиков, повторяющимися с периодом $2\omega_в$ (штриховые линии на рис. 8.2). Полученную таким образом периодическую функцию разложим в ряд Фурье, заменив в формуле (2.39) аргумент $t$ на $\omega$, частоту $\omega_1 = \omega_в$ на $\Delta t$ и (формально) $n$ на $k$.

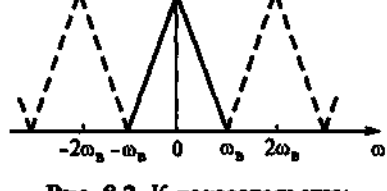



**Рис. 8.2.** К доказательству теоремы Котельникова

Тогда

$$S(\omega) = \sum_{k=-\infty}^{\infty} C_k e^{j\omega k\Delta t}. \qquad (8.5)$$

Полагая, что в соотношении (2.40) аналог периода $T$ — $2\omega_в$, а интервал дискретизации $\Delta t = \pi/\omega_в$, запишем

$$C_k = \frac{1}{2\omega_в} \int_{-\omega_в}^{\omega_в} S(\omega) e^{-j\omega k\pi/\omega_в} d\omega. \qquad (8.6)$$

Воспользуемся формулой обратного преобразования Фурье (2.55) и представим исходный непрерывный сигнал в следующем виде:

$$u(t) = \frac{1}{2\pi} \int_{-\omega_в}^{\omega_в} S(\omega) e^{j\omega t} d\omega. \qquad (8.7)$$

Таким же образом запишем значение дискретизированного сигнала для некоторого $k$-го отсчета времени. Поскольку время $t = k\Delta t = k\pi/\omega_в$, то

$$u(k\Delta t) = \frac{1}{2\pi} \int_{-\omega_в}^{\omega_в} S(\omega) e^{j\omega k\pi/\omega_в} d\omega.$$

Сравнив эту формулу с (8.6), замечаем, что $C_k = \Delta t\, u(-k\Delta t)$. С учетом этого соотношения функция (8.5), после несложных преобразований, примет вид:

$$S(\omega) = \sum_{k=-\infty}^{\infty} \Delta t\, u(-k\Delta t) e^{j\omega k\Delta t} = \sum_{k=-\infty}^{\infty} \Delta t\, u(k\Delta t) e^{-j\omega k\Delta t}. \qquad (8.8)$$

Затем проделаем следующее: подставим выражение (8.8) в соотношение (8.7), изменим порядок интегрирования и суммирования, представим отношение $\pi/\Delta t = \omega_в$ и вычислим интеграл. В результате получим такую формулу:

$$u(t) = \sum_{k=-\infty}^{\infty} \frac{\Delta t}{2\pi} u(k\Delta t) \int_{-\omega_в}^{\omega_в} e^{j\omega(t-k\Delta t)} d\omega = \sum_{k=-\infty}^{\infty} u(k\Delta t) \frac{\sin\omega_в(t-k\Delta t)}{\omega_в(t-k\Delta t)}. \qquad (8.9)$$

Из (8.9) следует, что непрерывная функция $u(t)$ определяется совокупностью ее значений амплитуд в отсчетные моменты времени $t = k\Delta t$, что и доказывает теорему Котельникова. Формулу (8.9) называют *интерполяционным рядом Котельникова* (точнее, Котельникова–Шеннона) и, по существу, она является разложением сигнала по системе ортогональных функций. С другой стороны, (8.9) представляет свертку дискретной функции отсчетов $u(k\Delta t)$ с непрерывной функцией интегрального синуса. Для больших массивов данных точность восстановления исходного сигнала ограничивают интервалом задания функции интегрального синуса, по которому устанавливается интервал суммирования.

Простейшие сигналы вида

$$s_k(t) = \frac{\sin \omega_в (t - k\Delta t)}{\omega_в (t - k\Delta t)}, \qquad (8.10)$$

ортогональные на временно́м интервале $[-\infty, \infty]$, называют *функциями отсчетов, базисными функциями, или функциями Котельникова.*

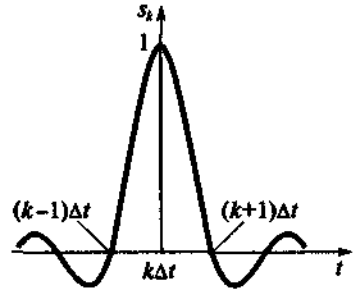



**Рис. 8.3.** График функции отсчетов

График $k$-й функции Котельникова представлен на рис. 8.3. Каждая базисная функция Котельникова $s_k(t)$ сдвинута относительно соседней функции $s_{k-1}(t)$ или $s_{k+1}(t)$ на интервал дискретизации $\Delta t$. Элементарный анализ соотношения (8.10) и ее графика показывает, что сигнал $s_k(t)$ отражается известной функцией $\sin x/x$, характеризующей огибающую спектральной плотности прямоугольного видеоимпульса.

Представление заданного аналогового сигнала $u(t)$ рядом Котельникова (8.3) поясняется диаграммами на рис. 8.4. На графике (здесь для упрощения базисные функции $s_k$ показаны без аргумента $t$) построены четыре первых члена ряда Котельникова, соответствующие отсчетам непрерывного сигнала в моменты времени 0, $\Delta t$, $2\Delta t$ и $3\Delta t$, взятым в соответствии с теоремой Котельникова. При суммировании этих членов ряда в любые отсчетные моменты времени $k\Delta t$ непрерывный сигнал абсолютно точно аппроксимируется (восстанавливается) независимо от числа выбранных отсчетов. В интервале же между любыми отсчетами сигнал $u(t)$ аппроксимируется тем точнее, чем больше суммируется членов ряда Котельникова (8.3). Заметим, что соединить дискретные отсчеты сигнала на графике прямыми линиями было бы не совсем верно, так как при восстановлении непрерывного сигнала по дискретному используют более сложные интерполирующие функции.

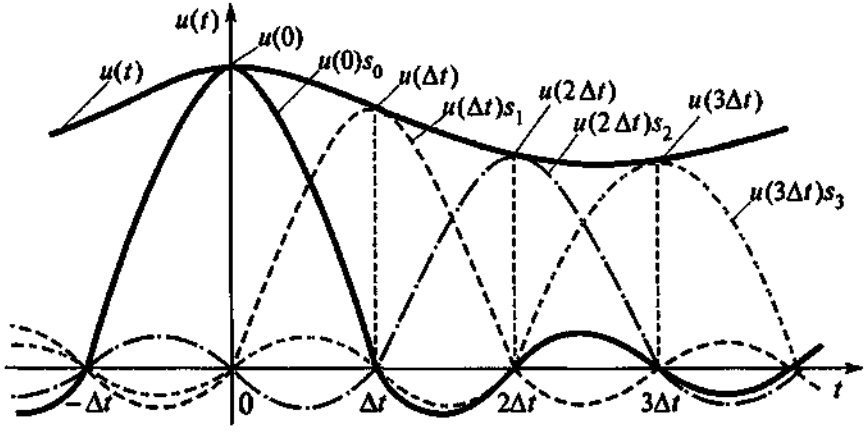



**Рис. 8.4.** Аппроксимация непрерывного сигнала рядом Котельникова

На практике эта теорема имеет огромное значение. Например, как уже отмечалось, большинство звуковых сигналов можно с некоторой степенью точности считать сигналами с ограниченным спектром. Их спектр в основном лежит ниже 20 кГц. Это значит, что при дискретизации с частотой не ниже 40 кГц можно более или менее точно восстановить исходный аналоговый звуковой сигнал по его цифровым отсчетам. Абсолютной точности достичь не удастся, поскольку в природе не бывает сигналов с идеально ограниченным спектром.

**Пример 8.1.** Сигнал звукового сопровождения в телевизионном канале ограничен верхней частотой $f_в$ = 12 кГц. Определить интервал $\Delta t$ между отсчетами этого сигнала, необходимый для неискаженного воспроизведения сигнала при передаче его дискретным способом.

Р е ш е н и е. Определяем интервал дискретизации

$$\Delta t = 1/(2f_в) = 1/(2\cdot 12\cdot 10^3) = 41{,}67\cdot 10^{-6}\ \text{с}.$$

Теореме отсчетов посвящено множество работ, в которых были сделаны ее обобщения для следующих функций:

- многомерных (случай, характерный для телевизионных сигналов);
- отсчеты которых берутся в произвольные моменты времени;
- у которых берутся отсчеты как самой функции, так и ее производной;
- для стохастических и т. п.

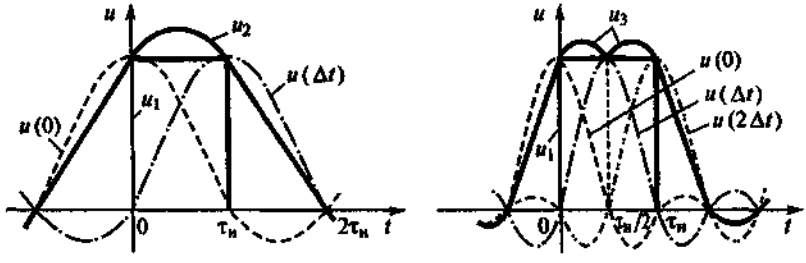
Теорема Котельникова сыграла величайшую роль в развитии теории связи и информации. Она наглядна при спектральном представлении сигнала. Ее преимуществом является *методологический характер*. Из нее вытекает не только частота дискретизации, но и метод восстановления рядом Котельникова, и способ восстановления при помощи идеального фильтра.

Оценим возможность применения теоремы Котельникова к импульсному сигналу $u(t)$ конечной длительности $T_и$. Как известно, такие сигналы теоретически обладают бесконечно широким спектром. Однако на практике всегда можно ограничиться некоторой верхней частотой $F_в$, за пределами которой в спектре содержится пренебрежительно малая доля энергии по сравнению с энергией всего исходного сигнала. В радиотехнике таким критерием является содержание 90 %-ной средней мощности сигнала в границах спектра. При этом сигнал $u(t)$ длительностью $T_и$ с верхней граничной частотой спектра $F_в$ можно представить рядом Котельникова с определенным, ограниченным числом отсчетов,

$$u(t) = \sum_{k=0}^{N} u(k\Delta t)\frac{\sin\omega_в(t-k\Delta t)}{\omega_в(t-k\Delta t)}, \qquad (8.11)$$

где $N = T_и/\Delta t$ — число отсчетов.

**Пример 8.2.** Представить рядом Котельникова прямоугольный импульс напряжения единичной амплитуды и длительности $\tau_и$ для двух случаев: спектр аппроксимирующей функции ограничен значениями верхней частоты $F_{в1} = 1/(2\tau_и)$ и $F_{в2} = 1/\tau_и$.




*а* *б*

**Рис. 8.5.** Представление прямоугольного импульса отсчетами:
*а* — двумя; *б* — тремя

Р е ш е н и е. Для первого случая интервал дискретизации $\Delta t = 1/(2F_в) = \tau_и$, а значит, импульс будет представлен всего двумя отсчетными значениями — в начале и конце импульса. Подставив в (8.11) значения амплитуды и длительности импульса, запишем математическую модель аппроксимирующей функции

$$u_2(t) = \frac{\sin(\pi t/\tau_и)}{\pi t/\tau_и} + \frac{\sin[\pi(t-\tau_и)/\tau_и]}{\pi(t-\tau_и)/\tau_и}.$$

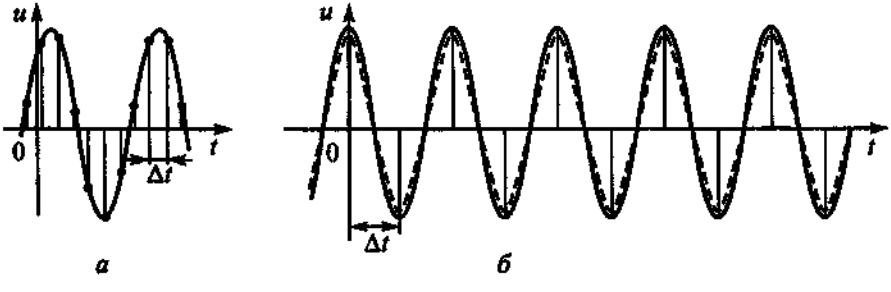
Во втором случае импульс дискретизируется тремя равными отсчетами, производимыми в моменты времени $t = 0$, $\tau_и/2$ и $\tau_и$, т. е. в начале, середине и конце импульса. Тогда аппроксимирующая функция

$$u_3(t) = \frac{\sin(2\pi t/\tau_и)}{2\pi t/\tau_и} + \frac{\sin[2\pi(t-\tau_и/2)/\tau_и]}{2\pi(t-\tau_и/2)/\tau_и} + \frac{\sin[2\pi(t-\tau_и)/\tau_и]}{2\pi(t-\tau_и)/\tau_и}.$$

Временны́е диаграммы аппроксимирующих функций $u_2(t)$ и $u_3(t)$ и образующие их члены ряда Котельникова представлены на рис. 8.5.

**Пример 8.3.** Определить минимальную частоту дискретизации по Котельникову, при которой гармонический сигнал $u(t) = \cos(2\pi F_0 t + \varphi_0)$ может быть полностью восстановлен по своим отсчетным значениям (выборкам).

Р е ш е н и е. При выборе интервала дискретизации $\Delta t = 1/(2F_в)$, где $F_в$ — верхняя граничная частота спектра, исходный сигнал $u(t)$ можно однозначно восстановить по отсчетам (рис. 8.6, *а*). Если соотношение частот $F_0 < F_в$, то к гармоническому сигналу тоже применима теорема отсчетов.




*а* *б*

**Рис. 8.6.** Восстановление синусоиды:
*а* — при $F_0 < F_в$; *б* — по двум отсчетам

При этом отсчетные значения сигнала

$$u_k = \cos(k\pi F_0 / F_в + \varphi_0).$$

В предельном случае, когда частота гармонического сигнала $F_0$ стремится к частоте дискретизации $F_в$ слева, т. е.

$$F_0 = \lim_{\mu \to 0} (F_в - \mu),$$

на каждом периоде исходного гармонического сигнала должно производиться две выборки (два отсчета).

Восстановление функции зависит от фазы отсчетов сигнала по отношению к временны́м выборкам. Если максимум синусоидальной функции приходится на середину интервала между отсчетами, то погрешность наибольшая, если же на отсчет, то наименьшая. Очевидно, что выборки могут попадать на нулевые значения синусоиды, экстремумы или промежуточные значения. Поскольку априорно фаза выборок относительно дискретизируемой синусоиды не известна, то после восстановления сигнала фильтром синусоиду вообще можно не увидеть. В данном примере самая высокая точность восстановления синусоиды будет тогда, когда обе выборки взяты в ее максимальных значениях. При этом колебание на входе ФНЧ имеет пилообразную форму той же частоты, что и частота исходного сигнала (штриховые линии на рис. 8.6, *б*).

Если же отсчеты во времени производят недостаточно часто и условия теоремы Котельникова нарушаются, то однозначное восстановление исходного гармонического сигнала принципиально невозможно. В этих случаях через отсчетные моменты времени можно провести бесчисленное множество кривых, спектральные плотности которых отличны от нуля вне полосы частот $-F_в \le F \le F_в$.

Поэтому можно утверждать, что погрешность восстановления исходной синусоиды при частоте выборок равной удвоенной частоте этой синусоиды может составлять 100 %. Уже этого достаточно для подтверждения правильности изложенных выводов по теореме Котельникова.

**Пример 8.4.** Дискретизированный в соответствии с теоремой Котельникова исследуемый непрерывный сигнал $u(t)$ определен двумя отсчета на временно́й оси: на $t = 0$ мкс и $t = 2$ мкс (рис. 8.7). Вычислить мгновенное значение исходного сигнала в момент времени $t = 1$ мкс.

Р е ш е н и е. По рис. 8.7 определяем, что интервал дискретизации $\Delta t = 2 \cdot 10^{-6}$ с. Отсюда верхняя круговая частота спектра исходного сигнала $\omega_в = \pi/\Delta t = 1{,}57 \cdot 10^6$ с$^{-1}$.

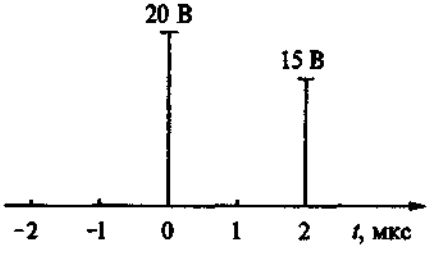



**Рис. 8.7.** К примеру 8.4

Согласно (8.11) ряд Котельникова для данного случая примет вид

$$u(t) = 20\frac{\sin\omega_в t}{\omega_в t} + 15\frac{\sin(\omega_в t - \pi)}{\omega_в t - \pi}.$$

Из этого соотношения определяем мгновенное значение непрерывного сигнала в момент времени $t = 1$ мкс

$$u(t = 1 \text{ мкс}) = 22{,}28 \text{ В}.$$

**Восстановление непрерывного сигнала по дискретным значениям**

Задачу восстановления (интерполяцию) непрерывного сигнала по его дискретным значениям решают прежде всего при графическом представлении результатов расчета на компьютере — от дискретных значений переходят к непрерывному графику. В математическом плане восстановление исходного сигнала сводится к отысканию функции, наилучшим образом описывающей сигнал. Эта функция должна в отсчетные моменты времени принимать заданные значения (равные дискретным значениям), а в остальные моменты описывает сигнал с требуемой степенью точности. Указанная операция без конкретизации физического содержания рассматриваемой функции составляет содержание общей задачи интерполяции. Интерполировать временну́ю функцию сигнала приходится и тогда, когда известно аналитическое описание сигнала, но функция является сложной, и имеются определенные трудности ее расчета. В этом случае вычисляют значения сигнала для нескольких моментов времени и по ним строят более простую (интерполирующую) функцию, с помощью которой определяют значения сигнала в остальных точках.

В целом известные методы интерполяции можно разделить на три основных группы, в которых применяют:

- степенны́е полиномы (ряды);
- ортогональные функции;
- стохастический анализ.

**Интерполяция степенны́ми полиномами.** Сущность методов степенны́х рядов заключается в предположении, что исходная функция может быть представлена степенны́м рядом вида

$$f(x) = a_0 + a_1 x + a_2 x + a_n x^n$$

при $n \to \infty$.

Для определения ряда $f(x)$ необходимо определить $\{a_i\}$, для чего нужно решить систему уравнений при известных значениях $\{x_i\}$. Способы определения $\{a_i\}$ по известным $\{x_i\}$ бывают различными, что и определяет различие методов интерполяции полиномами.

Одним из наиболее известных интерполяционных алгебраических полиномов является полином Лагранжа. Известные разновидности полинома Лагранжа: интерполяционные формулы Ньютона, Гаусса, Бесселя, Стирлинга, Эферетта, сплайн-функции (сплайны) и др. Заметим, что полином Лагранжа обладает хорошими интерполяционными свойствами при сравнительно небольшом числе узлов интерполяции.

В радиотехнике до сих пор применяют интерполяционный ряд Котельникова (8.11). Одно из основных его достоинств связано с тем, что в качестве коэффициентов ряда используют дискретные значения, которыми задается сигнал.

Основным достоинством методов степенны́х полиномов является их простота при реализации на компьютерах, поэтому сейчас они являются основными при цифровой обработке информации. Недостатками методов степенны́х поли-

номов являются появление эффекта увеличения погрешности при увеличении степени полинома и недостаточная обоснованность применения их к случайным функциям. Это связано с тем, что методы интерполяции первоначально разрабатывались для потребностей астрономии.

Что же касается случайных функций, то интерполяция конечными полиномами связана с погрешностями, которые существенно влияют не только на качество восстановления сигнала, но и на саму возможность применять интерполяцию для восстановления исходного сигнала.

Одним из видов интерполирующей функции является сплайновая функция, или сплайн (сплайновый метод разработан в 1948 г.). *Сплайн* (от англ. *spline* — гибкая линейка) — линия, форма которой определяется типом вершин узловых точек, через которые проходит эта линия (наиболее общие типы сплайна — кривые Безье; *Bezier curves*). В более широком смысле под сплайнами понимают плоские фигуры, служащие заготовками для создания трехмерных объектов. Сплайн описывает интерполируемую функцию на каждом частном интервале дискретизации. Интерполяция сплайнами — метод, дающий в результате гладкую функцию. Достигается это следующим образом: между каждыми двумя точками функция интерполируется кубической параболой (т. е. полиномом третьей степени вида $y = ax^3 + bx^2 + cx + d$), причем функцию подбирают так, чтобы на концах отрезка (как раз в точках задания значений) совпадали с какими-то заранее заданными числами и значения функции, и значения производной. Совпадение самих значений функции требуется для непрерывности получаемой приближающей функции, а совпадение значений производной — для непрерывности производной, т. е. для «гладкости» приближающей функции. Таким образом, результат интерполяции получается составленным не из прямых, как при кусочно-линейной интерполяции, а из «склеенных» кусков разных кубических парабол. Причем эта кривая получается еще и гладкой, так как «склеиваются» не только сами значения функции, но и значения ее производной.

При сплайновой интерполяции чаще всего используют локальные полиномы не выше третьей степени. Так, например, *кубические* сплайны проходят через три смежные узловые точки (текущие опорные точки вычислений), при этом в граничных точках совпадают как значения полинома и функции, так и значения их первых и вторых производных. Коэффициенты полиномов, проходящих через три смежные узловые точки, рассчитывают так, чтобы непрерывными были первая и вторая его производные. Линия, которую в этом случае описывает сплайн-функция, напоминает по форме гибкую линейку, закрепленную в узловых точках (линия *1* на рис. 8.8). Это создает высокую плавность сплайнового интерполирующего полинома по сравнению с другими методами аппроксимации, например полиномом Лагранжа (линия *2* на рис. 8.8).

**Интерполяция ортогональными функциями.** Методы применения ортогональных разложений начали развиваться со времени разработки Ж. Фурье разложения функций по гармоническим составляющим.

**Интерполяция на основе стохастического анализа.** Методами стохастического анализа определяют математическое ожидание погрешности дискрети-

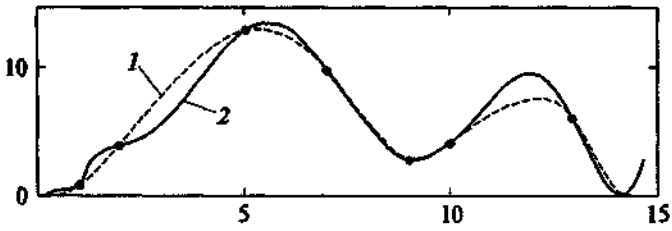



**Рис. 8.8.** Интерполяция сигнала по дискретным отсчетам:
*1* — сплайновая; *2* — полиномом Лагранжа

зации. В настоящее время при метрологическом анализе процесса дискретизации непрерывных сигналов чаще всего идут этим путем. Такие методы пока не нашли широкого применения вследствие трудностей определения стохастических характеристик исходной информации. Кроме того, эти методы применимы для стационарных случайных процессов, в то время как реально измеряемые параметры таковыми, как правило, не являются.

**Алиасинг**. Что же произойдет при восстановлении исходного сигнала, если попытаться оцифровать сигнал с недостаточной для него частотой дискретизации (или если спектр сигнала не ограничен)? В этом случае по полученной цифровой выборке не возможно верно восстановить исходный сигнал. Восстановленный сигнал будет выглядеть так, как если бы частоты, лежащие выше половины частоты дискретизации, отразившись от половины частоты дискретизации, перешли в нижнюю часть спектра и наложились на частоты, уже присутствующие там. Эти частоты суть «призраки», а эффект называют *наложением спектров*, эффектом *маскировки* (подмены) *частот*, *элайзингом*, или *алиасингом* (от англ. — *aliasing*). Употребляется также термин «алиасы».

Положим, что с частотой 45 кГц производится оцифровка музыкального сигнала, спектр которого ограничен частотой 20 кГц. При записи какой-то электронный прибор (например, источник питания) сгенерировал сильную помеху с ультразвуковой частотой 40 кГц (*ультразвуком* называют механические колебания упругой среды с частотой, превышающей верхний предел слышимости человеком частоты, т. е. выше 20 кГц), которая наложилась на спектр аналогового звукового сигнала. Пусть спектр звука, лежащий ниже частоты 45/2 = 22,5 кГц, будет записан по теореме отсчетов. Но поскольку ультразвуковая помеха лежит выше частоты 22,5 кГц, то возникнет алиасинг, и помеха «отразится» в нижнюю часть спектра сигнала на частоту 45 – 40 = 5 кГц. Если теперь восстановить с помощью ЦАП аналоговый сигнал и прослушать его, то услышим на фоне музыки помеху на частоте 5 кГц. Таким образом, помеха переместилась из неслышимой человеком ультразвуковой области в слышимую область.

Эффект маскировки частот называют стробоскопическим. *Стробоскопический эффект* — зрительная иллюзия замедленного движения (или неподвижности), возникающая из-за инерции зрения человека, когда какой-либо предмет наблюдают не непрерывно, а в течение отдельных периодически следующих один за другим интервалов времени. При этом движущийся предмет периодиче-

ски (с частотой $F_1$ — частота смены кадров — аналог частоты дискретизации) занимает прежнее положение (проявляется в кинематографе и телевидении, когда быстро вращающееся колеса движущегося автомобиля выглядят неподвижными либо медленно поворачивающимися, причем в любую сторону). Для иллюзии полной неподвижности необходимо, чтобы частота наблюдения $F_0$ была равна частоте смены кадров $F_1$. Если $F_0$ и $F_1$ не равны, но близки, то кажущееся движение предмета характеризуется *частотой биений* $F_б = F_0 - F_1$.

Как избежать алиасинга? Первый способ — использовать более высокую частоту дискретизации, чтобы весь спектр записываемого сигнала уместился ниже половины частоты дискретизации. Второй способ — искусственно ограничить спектр сигнала перед оцифровкой. Для этого применяют фильтры низких частот. В этом случае ФНЧ называют *антиалиасинговыми*. Частоту среза антиалиасинговых ФНЧ выбирают равной половине частоты дискретизации.

**Пример 8.5.** Известно, что для получения разборчиво звучащей человеческой речи достаточно оцифровывать ее с частотой 8 кГц. Какой диапазон частот может быть правильно передан такой цифровой записью? Что необходимо предпринять при оцифровке для правильной передачи этого диапазона?

Р е ш е н и е. В соответствии с теоремой Котельникова, при частоте дискретизации 8 кГц можно правильно передать диапазон частот 0...4 кГц. Однако так как человеческая речь содержит частоты и выше 4 кГц, то для предотвращения алиасинга необходимо перед оцифровкой пропустить сигнал через антиалиасинговый фильтр с частотой среза 4 кГц.

## 8.3. ДИСКРЕТНЫЕ СИГНАЛЫ. АЛГОРИТМЫ ДИСКРЕТНОГО И БЫСТРОГО ПРЕОБРАЗОВАНИЙ ФУРЬЕ

**Дискретизация непрерывного сигнала.** Процедуру получения *дискретизированного* (термин «дискретизированный» в данном контексте подчеркивает, что последовательность отсчетов получена именно в результате дискретизации непрерывного сигнала; поэтому далее проще, *дискретного*) сигнала $u_Д(t)$ удобно рассматривать как непосредственное умножение непрерывного сигнала $u(t)$ на вспомогательную последовательность $y(t)$ коротких дискретизирующих прямоугольных импульсов единичной амплитуды

$$u_Д(t) = u(t)y(t). \qquad (8.12)$$

На практике эту операцию осуществляют с помощью электронного ключа $К$ и генератора прямоугольных импульсов $Г$ (рис. 8.9, *а*). Длительность импульсов дискретизации $\tau_и$ должна быть небольшой (теоретически дельта-функция), причем много меньше интервала $\Delta t$. Принцип формирования дискретного сигнала показан на рис. 8.9, *б – г*, где приведены графики функций $u(t)$, $y(t)$ и $u_Д(t)$. При этом реальный дискретный сигнал $u_Д(t)$ имеет вид импульсно-модулированного колебания, т. е. АИМ-сигнала (рис. 8.9, *г*).

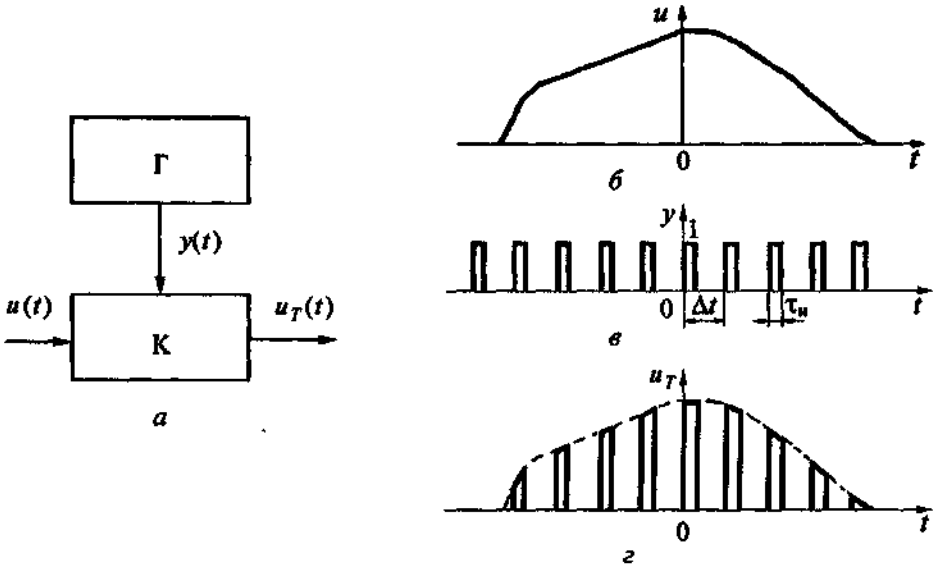



**Рис. 8.9.** Дискретизация сигнала:
*а* — дискретизатор; *б* — непрерывный сигнал;
*в* — последовательность дискретизирующих импульсов; *г* — дискретный сигнал

### Спектр дискретного сигнала

Чтобы дать оценку требованиям к длительности дискретизирующих импульсов, определим спектральный состав дискретного сигнала $u_T(t)$. Пусть непрерывный сигнал $u(t)$ имеет спектральную плотность S(ω) (рис. 8.10, *а*). Представим последовательность дискретизирующих прямоугольных импульсов $y(t)$ рядом Фурье (2.27), в котором частота $\omega_1 = 2\pi/\Delta t$

$$y(t) = \frac{\tau_и}{\Delta t}\left(1 + 2\sum_{n=1}^{\infty} I_n \cos n\omega_1 t\right). \tag{8.13}$$

Здесь коэффициенты

$$I_n = \frac{\sin(n\omega_1 \tau_и / 2)}{n\omega_1 \tau_и / 2}. \tag{8.14}$$

Подставив формулу (8.13) в (8.12), получим

$$u_T(t) = \frac{\tau_и}{\Delta t} u(t) + 2\frac{\tau_и}{\Delta t}\sum_{n=1}^{\infty} I_n u(t)\cos n\omega_1 t. \tag{8.15}$$

Проанализируем первое и второе слагаемые этого выражения отдельно. Первому слагаемому соответствует спектральная плотность S(ω) исходного сигнала $u(t)$. К произведению $u(t)\cos n\omega_1 t$ второго слагаемого применим прямое преобразование Фурье (2.54).

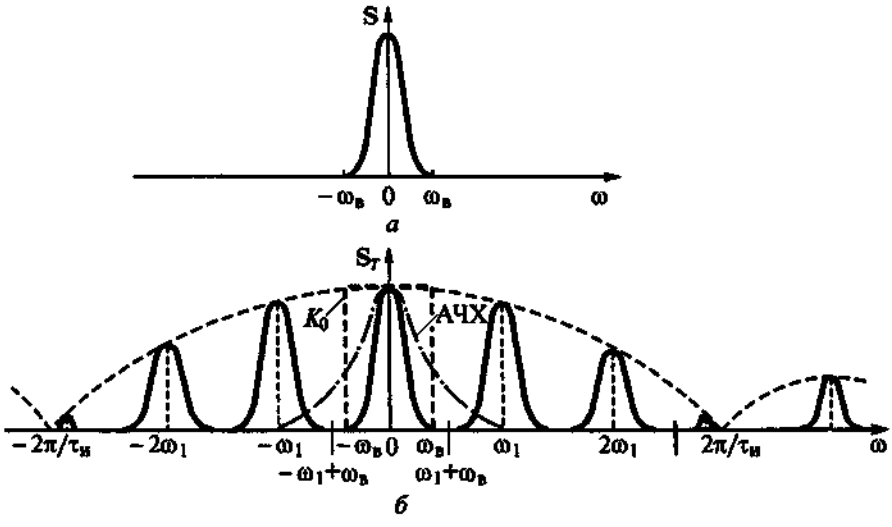



**Рис. 8.10.** Спектры сигналов:
*а* — непрерывного; *б* — дискретного

Используя формулу Эйлера и проведя несложные математические выкладки, запишем

$$\int_{-\infty}^{\infty} u(t)\cos n\omega_1 t\, e^{-j\omega t}dt = \frac{1}{2}\int_{-\infty}^{\infty} u(t)e^{-j(\omega-n\omega_1)t}dt + \frac{1}{2}\int_{-\infty}^{\infty} u(t)e^{-j(\omega+n\omega_1)t}dt.$$

В этом выражении первый интеграл представляет собой спектральную плотность сигнала $u(t)$ на частотах $\omega - n\omega_1$, а второй — ту же спектральную плотность, но на частотах $\omega + n\omega_1$, поэтому

$$\int_{-\infty}^{\infty} u(t)\cos n\omega_1 t\, e^{-j\omega t}dt = \frac{1}{2}\left[S(\omega - n\omega_1) + S(\omega + n\omega_1)\right].$$

Следовательно, дискретному сигналу вида (8.15) соответствует спектральная плотность

$$S_T(\omega) = \frac{\tau_и}{\Delta t}\left[S(\omega) + \sum_{n=1}^{\infty} I_n S(\omega - n\omega_1) + \sum_{n=1}^{\infty} I_n S(\omega + n\omega_1)\right]. \qquad (8.16)$$

Учитывая, что при $n = 0$ коэффициент $I_n = 1$, запишем:

$$S_T(\omega) = \frac{\tau_и}{\Delta t}\sum_{n=-\infty}^{\infty} I_n S(\omega - n\omega_1) = \frac{\tau_и}{\Delta t}\sum_{n=-\infty}^{\infty} \frac{\sin(n\omega_1\tau_и/2)}{n\omega_1\tau_и/2} S(\omega - n\omega_1). \qquad (8.17)$$

График спектра дискретного сигнала $S_T(\omega)$, который сформирован на основе теоремы Котельникова из непрерывного, показан на рис. 8.10, *б*.

Полученные результаты позволяют сделать фундаментальные выводы для теории дискретных сигналов:

• *спектральная плотность* $S_T(\omega)$ *дискретного сигнала* $u_T(t)$ *представляет собой бесконечную последовательность спектральных плотностей* $S(\omega)$ *исходного непрерывного сигнала* $u(t)$, *сдвинутых друг относительно друга на частоту дискретизации* $\omega_1$;

• *огибающая спектральной плотности* $S_T(\omega)$ *дискретного сигнала* $u_T(t)$ *с точностью до коэффициента* $1/\Delta t$ *повторяет огибающую спектральной плотности дискретизирующего прямоугольного импульса.*

Чтобы восстановить непрерывный сигнал $u(t)$ из дискретного $u_T(t)$, достаточно выделить небольшую центральную часть спектра $S_T(\omega)$. На практике это осуществляют идеальным ФНЧ, имеющим коэффициент передачи вида (штриховая линия прямоугольной формы на рис. 8.10, *б*):

$$K(\omega) = K_0, \qquad -\omega_в \le \omega \le \omega_в. \tag{8.18}$$

Вместе с тем из курса основ теории цепей известно, что идеальный фильтр нижних частот физически нереализуем и может служить лишь теоретической моделью для пояснения принципа восстановления аналогового сигнала на основе теоремы отсчетов. Реальный ФНЧ имеет АЧХ, которая либо охватывает несколько лепестков спектра (штрихпунктирная линия АЧХ на рис. 8.9, *б*), либо имеет конечную длительность крутизны ската характеристики и не полностью охватывает центральный лепесток. Этот недостаток накладывает определенные ограничения на применение теоремы отсчетов, требуя уменьшения интервала дискретизации. На практике интервал дискретизации сигналов, определяемый формулой (8.3), уменьшают в 2...5 раз. Тогда отдельные составляющие спектра дискретизированного сигнала $S_T(\omega)$ не перекрываются, как это и показано на рис. 8.10, *б*, и могут быть разделены фильтрами.

С уменьшением длительности дискретизирующего импульса $\tau_и$ амплитуды спектральных составляющих с ростом частоты убывают медленнее. В предельном случае, при $\tau_и \to 0$, спектр дискретного сигнала будет представлять собой бесконечную последовательность «копий» спектров исходного сигнала, имеющих равную амплитуду. Если одновременно с уменьшением длительности импульса увеличивать его амплитуду так, чтобы его площадь оставалась неизменной и равной единице (фактически этот импульс — дельта-функция), то теоретически дискретизирующим сигналом может быть бесконечная последовательность дельта-функций

$$y(t) = \sum_{k=-\infty}^{\infty} \delta(t - k\Delta t).$$

В этом случае формула (8.12) запишется следующим образом:

$$u_T(t) = u(t) \sum_{k=-\infty}^{\infty} \delta(t - k\Delta t) = \sum_{k=-\infty}^{\infty} u(k\Delta t)\delta(t - k\Delta t). \tag{8.19}$$

Итак, согласно выражению (8.19), дискретный сигнал $u_T(t)$ при подобной дискретизации представляет собой последовательность дельта-функций, следующих с интервалом времени $\Delta t$. Эти функции имеют амплитудные коэффициенты, равные выборкам (амплитудам, уровням) непрерывного сигнала $u(t)$ в точках дискретизации $t = k\Delta t$.

Аналитическое выражение (8.17) для спектральной плотности дискретного сигнала в этом случае примет вид:

$$S_T(\omega) = \frac{1}{\Delta t}\sum_{n=-\infty}^{\infty} S(\omega - n\omega_1). \qquad (8.20)$$

**Пример 8.6.** Непрерывный сигнал, имеющий форму прямоугольного импульса напряжения с единичной амплитудой и длительностью $\tau_и$, дискретизирован десятью отсчетами. Определить спектр дискретного сигнала.

**Решение.** Для нахождения спектра воспользуемся формулой (8.20). В ней частота $\omega_1 = 2\pi/\Delta t = 20\pi/\tau_и$, интервал дискретизации $\Delta t = \tau_и/10$, а спектральная плотность дискретизируемого прямоугольного импульса определяется выражением (2.61).

Тогда из (8.21) получим

$$S_T(\omega) = \frac{10}{\tau_и}\sum_{n=-\infty}^{\infty}\tau_и \frac{\sin\left(\frac{\omega\tau_и}{2} - \frac{20n\pi}{\tau_и}\cdot\frac{\tau_и}{2}\right)}{\frac{\omega\tau_и}{2} - \frac{20n\pi}{\tau_и}\cdot\frac{\tau_и}{2}} = 10\sum_{n=-\infty}^{\infty}\frac{\sin\left(\frac{\omega\tau_и}{2} - 10n\pi\right)}{\frac{\omega\tau_и}{2} - 10n\pi}.$$

Возможность представления дискретных сигналов $u_T(t)$ в форме (8.20) существенно упрощает их анализ. В частности, спектральную плотность $S_T(\omega)$ можно вычислить непосредственно по совокупности временны́х отсчетов $\{u(k\Delta t)\} = u_k$. Действительно, применив прямое преобразование Фурье (2.54) к соотношению (8.20) для отсчетов только с положительными номерами $k = 0, 1, ..., \infty$, получим с учетом фильтрующего свойства дельта-функции:

$$S_T(\omega) = \int_0^{\infty}\sum_{k=0}^{\infty} u(k\Delta t)e^{-j\omega t}\delta(t - k\Delta t)dt = \sum_{k=0}^{\infty} u(k\Delta t)e^{-j\omega k\Delta t}. \qquad (8.21)$$

При этом существенно сокращается время обработки реальных сигналов.

### Дискретное преобразование Фурье

Как и при анализе аналоговых сигналов, дискретные сигналы можно представить во временно́й и частотной областях. В настоящее время обработку дискретных сигналов чаще всего проводят в частотной области, что диктуется значительными сокращениями объема цифровой аппаратуры и времени обработки.

Положим, что сигналы, которые нужно преобразовать, являются дискретными и, возможно, непериодическими. В таком случае преобразование Фурье

использовать нельзя, поскольку оно предназначено для непрерывных функций. Однако существует преобразование, которое можно применять к дискретным сигналам, — дискретное преобразование Фурье. Пусть дискретной обработке подвергается импульсный сигнал $u(t)$ длительностью $T_и$, имеющий спектральную плотность $S(\omega)$ (рис. 8.11, *а*, *б*). Теоретически можно предположить, что дискретизация сигнала производится последовательностью дельта-функций

$$y(t) = \sum_{k=0}^{N-1} \delta(t - k\Delta t), \tag{8.22}$$

где $N = T_и/\Delta t$ — требуемое число отсчетов сигнала, отвечающее теореме отсчетов.

Подставив в (8.19) пределы суммирования от 0 до $N-1$ и заменив здесь и далее для упрощения и уменьшения объема формул $u(k\Delta t) = u_k$, запишем выражение для дискретного сигнала (рис. 8.11, *в*):

$$u_T(t) = u(t)\sum_{k=0}^{N-1} \delta(t - k\Delta t) = \sum_{k=0}^{N-1} u_k \delta(t - k\Delta t). \tag{8.23}$$

На основании формулы (8.23) можно сделать вывод, что спектр данного дискретного сигнала имеет периодическую структуру с периодом по оси частот $\omega_1 = 2\pi/\Delta t$ (рис. 8.11, *г*). Мысленно продолжим дискретный сигнал периодически с интервалом $T_и$ (рис. 8.10, *д*)

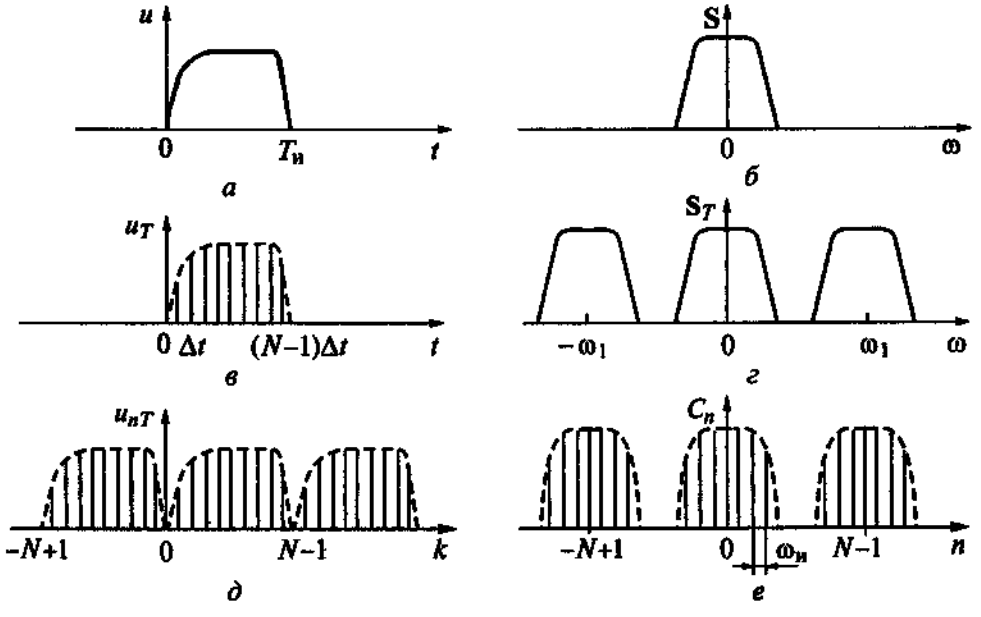



**Рис. 8.11.** К выводу алгоритма дискретного преобразования Фурье:
*а*, *б* — аналоговый сигнал и его спектр; *в*, *г* — дискретный сигнал и его спектр; *д* — периодическая последовательность дискретного сигнала; *е* — ДПФ сигнала

$$u_{nT}(t + nT_н) = u_T(t), \quad n = 0, \pm 1, \pm 2, ...$$

По аналогии с представлением периодических непрерывных сигналов дискретную функцию $u_{nT}(t)$ можно разложить в комплексный ряд Фурье (2.39)

$$u_{nT}(t) = \sum_{n=-\infty}^{\infty} \mathbf{C}_n e^{jn\omega_н t}, \tag{8.24}$$

где $\omega_н = 2\pi/T_н = 2\pi/(N\Delta t)$ — частота дискретизации спектра дискретного сигнала.

Согласно (2.40), комплексные коэффициенты этого ряда:

$$\mathbf{C}_n = \frac{1}{T_н}\int_0^{T_н} u_T(t) e^{-jn\omega_н t} dt = \frac{1}{T_н}\int_0^{T_н} u_T(t) e^{-j2\pi nt/T_н} dt. \tag{8.25}$$

Для определения этих коэффициентов проделаем следующее. Подставим формулу (8.23) в (8.25) и, заменив параметр $T_н = N\Delta t$, а также введя безразмерную переменную $y = t/\Delta t$, запишем

$$\mathbf{C}_n = \frac{1}{T_н}\int_0^{T_н}\sum_{k=0}^{N-1} u_k \delta(t - k\Delta t) e^{-j2\pi nt/T_н} dt = \frac{1}{N}\sum_{k=0}^{N-1} u_k \int_0^{N} \delta(y - k) e^{-j2\pi ny/N} dy.$$

Используя фильтрующее свойство дельта-функции, находим

$$\int_0^{N} \delta(y - k) e^{-j2\pi ny/N} dy = e^{-j2\pi nk/N}.$$

Тогда окончательная формула для вычисления коэффициентов ДПФ

$$\mathbf{C}_n = \frac{1}{N}\sum_{k=0}^{N-1} u_k e^{-j2\pi nk/N}. \tag{8.26}$$

Это фундаментальное для дискретных сигналов соотношение, показанное на рис. 8.11, *е*, называют *дискретным преобразованием Фурье* (ДПФ; англ. — *Discrete Fourier Transform, DFT*). Дискретное преобразование Фурье по существу представляет собой алгоритм вычисления гармонических составляющих спектра $\mathbf{C}_n$ по заданным дискретным отсчетам $u_k$ аналогового сигнала $u(t)$, что значительно сокращает время обработки.

Следует отметить ряд специфических свойств дискретного преобразования Фурье, которые нетрудно сформулировать из свойств преобразований Фурье для непрерывных сигналов.

1. Дискретное преобразование Фурье обладает свойством линейности — сумме (разности) дискретных сигналов отвечает сумма (разность) их ДПФ.

2. Коэффициент $\mathbf{C}_0$ представляет собой среднее значение (постоянную составляющую) всех дискретных отсчетов сигнала

$$\mathbf{C}_0 = \frac{1}{N}\sum_{k=0}^{N-1} u_k.$$

3. Число определяемых коэффициентов $\mathbf{C}_n$ равно числу отсчетов $N$ за длительность сигнала $T_н$; при $n = N$ коэффициент $\mathbf{C}_N = \mathbf{C}_0$.

**Пример 8.7.** Найти коэффициенты ДПФ дискретизированного прямоугольного импульса единичной амплитуды, заданного четырьмя отсчетами $\{u_k\} = (1, 1, 1, 1)$.

Решение. Используя основную формулу (8.26), вычислим коэффициенты ДПФ в пределах периода $N = 4$:

$$\mathbf{C}_0 = 4/4 = 1; \qquad \mathbf{C}_1 = \frac{1}{4}\left(1 + e^{-j\pi/2} + e^{-j\pi} + e^{-j3\pi/2}\right) = 0;$$

$$\mathbf{C}_2 = \frac{1}{4}\left(1 + e^{-j\pi} + e^{-j2\pi} + e^{-j3\pi}\right) = 0; \qquad \mathbf{C}_3 = \frac{1}{4}\left(1 + e^{-j3\pi/2} + e^{-j3\pi} + e^{-j9\pi/2}\right) = 0.$$

Итак, в пределах периода лишь одна спектральная составляющая с амплитудой 1 отлична от нуля. Интересно подтверждение 3-го свойства ДПФ, согласно которому коэффициент $\mathbf{C}_0 = \mathbf{C}_4$, поскольку $\mathbf{C}_4 = \frac{1}{4}\left(1 + e^{-j2\pi} + e^{-j4\pi} + e^{-j6\pi}\right) = 1$.

**Обратное дискретное преобразование Фурье.** При изучении теории ДПФ возникает естественный вопрос: можно ли по известным коэффициентам ДПФ вычислить отсчетные значения $u_k$ непрерывного сигнала? По аналогии с периодическими сигналами представим заданную периодическую последовательность отсчетов комплексным рядом Фурье.

Заменив в (8.24) $t = k\Delta t$, $\omega_н = 2\pi/(N\Delta t)$ и учитывая, что суммируется конечное число членов ряда, запишем:

$$u_k = \sum_{n=0}^{N-1} \mathbf{C}_n e^{j2\pi nk/N}. \qquad (8.27)$$

Данное соотношение отражает алгоритм *обратного дискретного преобразования Фурье* (ОДПФ, англ. — *Return Discrete Fourier Transform, RDFT*). Формулы (8.26) и (8.27) являются аналогами прямого (2.54) и обратного (2.55) преобразований Фурье для непрерывных сигналов.

**Пример 8.8.** Вычислить обратное дискретное преобразование Фурье последовательности четырех отсчетов сигнала $\mathbf{C}_0 = 1$; $\mathbf{C}_1 = 0$; $\mathbf{C}_2 = 0$; $\mathbf{C}_3 = 0$.

Решение. Используя формулу (8.27) для ОДПФ, вычислим отсчеты:

$$u_k = \sum_{n=0}^{N-1} \mathbf{C}_n e^{j2\pi nk/N} = \sum_{n=0}^{3} \mathbf{C}_n e^{j\pi nk/2} = \sum_{n=0}^{3} \mathbf{C}_0 e^{j\pi \cdot 0 \cdot k/2}.$$

Отсюда имеем: $u_0 = 1$; $u_1 = 1$; $u_2 = 1$; $u_3 = 1$.

Отметим, что известное применение находят и варианты дискретного преобразования Фурье. Ограничиваясь только косинусоидами, приходим к *дискретному косинусному преобразованию* (ДКП), позволяющему представлять только четные сигналы, т. е. $f(x) = f(-x)$, но избавляясь при этом от разрывов. Есть своя ниша и у *дискретного синусного преобразования* (ДСП), применимого только для нечетных функций.

Две другие задачи, где с успехом употребляется ДПФ, используют одно из его важнейших свойств, а именно, что ДПФ от свертки двух сигналов равно произведению их ДПФ. Поскольку выходной сигнал какой-либо линейной системы есть свертка входного сигнала и отклика этой системы, ДПФ от выхода системы получается перемножением ДПФ входного сигнала и ДПФ отклика. При этом, если доступны вход и выход системы и можно вычислить Фурье от них, передаточная функция получается простым делением. Точно так же, если нужно вычислить свертку, причем длина свертываемых отрезков столь велика, можно, посчитав ДПФ от свертываемых векторов, перемножив и найдя обратное ДПФ, получить искомое.

### Быстрое преобразование Фурье

Формула (8.26) показывает, что для определения одного коэффициента ДПФ последовательности из $N$ отсчетов, необходимо выполнить около $N$ операций умножения на комплексное число и столько же сложений, а для нахождения всех коэффициентов — $N^2$ вычислений. В частности, при $N = 2^{10} = 1\,024$ надо осуществить чуть более миллиона ($1024^2$) умножений и сложений. Если длины обрабатываемых массивов превышают тысячу, то дискретная спектральная обработка в реальном масштабе времени требует увеличения мощности и усложнения вычислительных комплексов.

Однако разработан способ выполнить отмеченные преобразования значительно быстрее. Достигается это искусным применением принципа, позволяющего не повторять уже проделанной обработки, и прилагательное «быстрое» отнюдь не следует трактовать, как «упрощенное» или «неточное»: результаты расчетов согласно новому алгоритму и по алгоритму ДПФ совпадают. Этот способ называется *быстрым преобразованием Фурье* (БПФ, англ. *fast Fourier transform — FFT*,), и его в 1965 г. разработали американские специалисты Д. Кули (*J. Gooley*) и Д. Тьюки.

Нетрудно заметить, что среди множителей (синусов) ДПФ есть много повторяющихся значений (в силу, например, периодичности функции синуса). Алгоритм БПФ группирует слагаемые с одинаковыми множителями, значительно сокращая число операций умножения. В результате быстродействие БПФ может в зависимости от $N$ в сотни раз превосходить быстродействие стандартного алгоритма ДПФ (кстати отметим, что число $N$ называется *размером*, или *длиной БПФ — FFT size*).

В основу алгоритма БПФ положен принцип разбиения (прореживания во времени; англ. — *decimation in time, DIT*), или *децимации* (*decimation* — от греч.

*децu* — доля; операция, заключающаяся в выбрасывании отсчетов, чей порядковый номер кратен определенному числу), заданной последовательности отсчетов исходного дискретного сигнала на ряд промежуточных последовательностей (подпоследовательностей). Это значит, что число отсчетов дискретного сигнала $N$ разделяют на множители (например, $N = 8 = 2 \cdot 2 \cdot 2$, $N = 60 = 3 \cdot 4 \cdot 5$). Затем определяют спектры этих промежуточных последовательностей отсчетов, с помощью которых и находится спектр всего сигнала.

В зависимости от состава, числа и порядка следования указанных множеств данных создают различные алгоритмы БПФ. В цифровой технике удобнее обрабатывать сигнальные последовательности со значениями числа отсчетов $N$, являющимися степенью с основанием два (4, 8, 16 и т. д.). Это позволяет многократно делить входную последовательность отсчетов на более мелкие подпоследовательности.

Пусть требуется вычислить ДПФ входной последовательности (массива отчетов) дискретного сигнала $\{u(k\Delta t)\} = \{u_k\}$, имеющей четное число отсчетов (рис. 8.12, *а*), причем $N = 2^r$, где $r$ — целое число (если это условие не выполняется, то последовательность искусственно дополняют нулями до требуемого значения $N$).

Представим входную последовательность в виде двух подпоследовательностей с четными и нечетными номерами и половинным числом членов в каждой (рис. 8.12, *б*, *в*):

$$u_{чт} = u_{2k}; \quad u_{нч} = u_{2k+1}; \quad k = 0, 1, 2, ..., N/2 - 1.$$

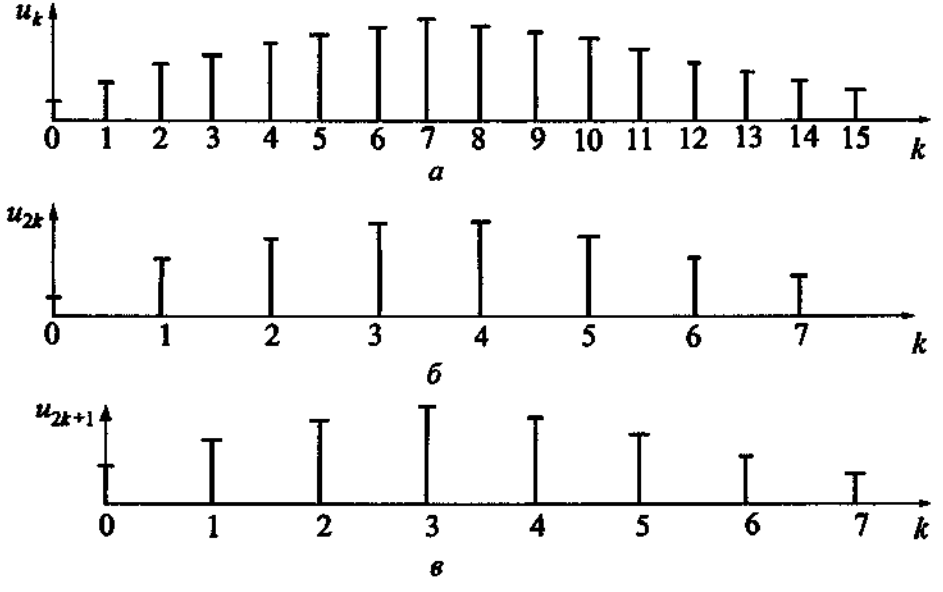



**Рис. 8.12.** Последовательности и подпоследовательности дискретного сигнала:
*а* — входная; *б* — с четными номерами; *в* — с нечетными номерами

Коэффициенты ДПФ для подпоследовательностей с четными и нечетными номерами запишем отдельно:

$$\frac{1}{N}\sum_{k=0}^{N/2-1} u_{2k} e^{-j2\pi n 2k/N} = \frac{1}{N}\sum_{k=0}^{N/2-1} u_{2k} e^{-\frac{j2\pi nk}{N/2}} = \mathbf{C}_{n\text{чт}};$$

$$\frac{1}{N}\sum_{k=0}^{N/2-1} u_{2k+1} e^{-j2\pi n(2k+1)/N} = \frac{1}{N} e^{-j2\pi n/N}\sum_{k=0}^{N/2-1} u_{2k+1} e^{-\frac{j2\pi nk}{N/2}} = e^{-j2\pi n/N}\mathbf{C}_{n\text{нч}}. \qquad (8.28)$$

Коэффициенты $\mathbf{C}_n$ результирующего ДПФ входной последовательности можно выразить через параметры $\mathbf{C}_{n\text{чт}}$ и $\mathbf{C}_{n\text{нч}}$ двух вновь введенных подпоследовательностей. Из последней формулы нетрудно заметить, что в диапазоне номеров отсчетов от 0 до $N/2-1$ ДПФ входной последовательности определяется соотношением

$$\mathbf{C}_n = \mathbf{C}_{n\text{чт}} + e^{-j2\pi n/N}\mathbf{C}_{n\text{нч}}, \qquad n = 0, 1, 2, ..., N/2 - 1. \qquad (8.29)$$

Так как ДПФ четной и нечетной подпоследовательностей являются периодическими, имеющими период следования $N/2$, то

$$\mathbf{C}_{n\text{чт}} = \mathbf{C}_{n\ -\ N/2\ \text{чт}}; \qquad \mathbf{C}_{n\text{нч}} = \mathbf{C}_{n\ +\ N/2\ \text{нч}}.$$

Входящий в формулу (8.29) экспоненциальный множитель при $n \geq N/2$ можно записать в виде:

$$e^{-j\frac{2\pi(N/2+n)}{N}} = e^{-j\pi} e^{-j\frac{2\pi n}{N}} = -e^{-j\frac{2\pi n}{N}}.$$

С учетом двух последних выражений находим ДПФ входной последовательности для отсчетов с номерами от $N/2$ до $N-1$

$$\mathbf{C}_{N/2+n} = \mathbf{C}_{n\text{чт}} - e^{-j2\pi n/N}\mathbf{C}_{n\text{нч}}, \qquad n = 0, 1, 2, ..., N/2 - 1. \qquad (8.30)$$

Соотношения (8.29) и (8.30) представляют собой *алгоритмы БПФ*. Отметим, что экспоненциальные фазовые множители $e^{-j2\pi n/N}$ в этих алгоритмах учитывают влияние сдвига нечетной подпоследовательности относительно четной.

Чтобы еще уменьшить число вычислений, четную и нечетную подпоследовательности ДПФ также разбивают каждую на две промежуточные части. Разбиение продолжают вплоть до получения простейших двухэлементных последовательностей. Определив ДПФ данных простейших пар отсчетов, можно вычислить ДПФ четырехэлементных, восьмиэлементных и так далее подпоследовательностей. При объединении в ДПФ четной и нечетной подпоследовательностей используют алгоритмы (8.29) и (8.30), подставляя в них соответствующие значения номеров $N$ и $n$.

Нетрудно заметить, что вычисление по формулам (8.28) не требует операций умножения, а только сложение и вычитание комплексных чисел. Учитываться же

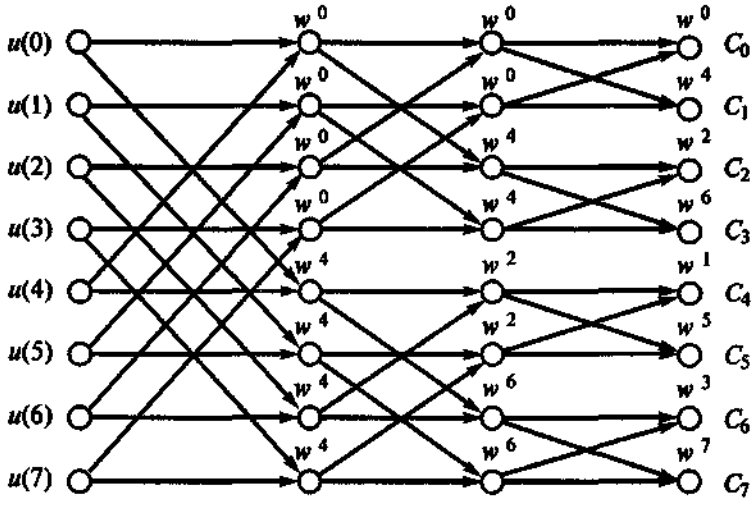



**Рис. 8.13.** Сигнальный граф БПФ для $N = 8$ отсчетов входного сигнала

должны лишь операции умножения в алгоритмах (8.29) и (8.30) для различных $n$ при разбиениях массива отчетов на мелкие подпоследовательности. Число этих операций при первом разбиении составляло $N/2$. Такое же число $N/2$ операций требуется выполнить при каждом следующем разбиении. Таким образом, вдвое увеличивается число подпоследовательностей и вдвое сокращается наибольшее число $n$ в формулах (8.29), (8.30).

Таким образом, в основу алгоритмов БПФ заложены операции сложения/вычитания с умножением одного из компонентов на экспоненциальный множитель $e^{-j2\pi n/N}$. Это *базовая операция* алгоритма БПФ с прореживанием по времени, получившая в отечественной цифровой технике жаргонное название «*бабочка*» (*butterfly*), поскольку диаграмма похожа на стилизованное изображение этого насекомого, а также по ассоциации с изображением ее *направленного графа*. Кружок на линиях графа обозначает арифметическую операцию сложения/вычитания, верхний выход соответствует сумме, нижний — разности, стрелка обозначает операцию умножения на поворачивающий множитель, стоящий над стрелкой (рис. 8.13).

Вычисление коэффициентов ДПФ последовательности из $N$ отсчетов сигнала по алгоритмам БПФ требует примерно $N\log_2 N$ операций умножения. Значит, алгоритмы БПФ сокращают число операций по сравнению с алгоритмами ДПФ в $N^2/(N\log_2 N) = N/\log_2 N$ раз. В частности, при числе отсчетов обрабатываемого сигнала $N = 2^{10}$, имеем $\log_2 N = 10$ и сокращение числа вычислительных операций составляет $N/\log_2 N \approx 100$.

При очень больших массивах отсчетов сигнала выигрыш в скорости обработки алгоритмом БПФ достигает нескольких тысяч. При этом следует подчеркнуть, что алгоритм БПФ является очень точным. Он даже точнее стандартного ДПФ, так как сокращая число операций, приводит к меньшим ошибкам округления.

Особенность построения сигнального графа по алгоритмам БПФ рассмотрим на примере вычисления коэффициентов ДПФ для восьмиэлементной ($N = 8$) последовательности входного сигнала (см. рис. 8.13). Узлы графа соответствуют исходным отсчетам сигнала $u(0)$, ..., $u(7)$, а также промежуточным и конечным коэффициентам ДПФ. Входящие в любой узел стрелки (дуги графа) обозначают переменные, участвующие в его формировании. Такое формирование основано на операции суммирования указанных переменных с определенными весовыми экспоненциальными множителями, соответствующими определенному узлу. Например, сокращенная запись в виде $w^0$ означает, что суммирование в узле должно осуществляться с множителем $w^0 = e^0 = 1$, при $w^4$ — с множителем $-1$: тогда $w^4 = e^{-j2\pi 4/8} = -1$, и т. д.

Как видно из структуры сигнального графа на рис. 8.13, все коэффициенты ДПФ $C_0$, $C_1$, ..., $C_7$ восьмиэлементной последовательности вычисляются через две четырехэлементные, которые, в свою очередь, определяются через четыре двухэлементные ДПФ.

## 8.4. Дискретная свертка сигналов. Теория $z$-преобразования

Свертка — основной процесс в цифровой обработке сигналов и ее часто и по существу называют *цифровой фильтрацией*. Поэтому важно уметь эффективно ее вычислять. Прямое вычисление свертки требует $NM$ умножений, где $N$ — количество отсчетов исходного сигнала; $M$ — число отсчетов импульсной характеристики используемого дискретного фильтра — длина *ядра свертки* (*kernel convolution* — дискретная импульсная характеристика фильтра). Часто длина ядра свертки достигает нескольких тысяч точек, и число умножений становится огромным. В теории цифровой обработки сигналов особенно важное значение имеет дискретная свертка.

### Дискретная свертка

При больших длинах ядра свертки сигналов существует специальный алгоритм, позволяющий вычислить ее значительно быстрее. Этот алгоритм основан на следующей важной теореме, которую приведем в достаточно полной, но не очень нестрогой формулировке.

**Теорема свертки**: *свертка во временно́й области эквивалентна умножению в частотной области; умножение во временно́й области эквивалентно свертке в частотной области*. Это значит, что для выполнения свертки двух сигналов можно перевести их в частотную область, перемножить их спектры и перевести результат обратно во временну́ю область. Такая операция выглядит громоздко. Однако с появлением алгоритмов БПФ, позволяющих быстро вычислять преобразования Фурье, вычисление свертки через частотную область резко сокращает число операций и поэтому стало широко использоваться в радиотехнике. При значительных длинах ядра свертки такой подход позволяет в сотни раз сократить время вычисления свертки.

По аналогии со сверткой двух непрерывных функций (2.67) имеем для сигналов $u(t)$ и $h(t)$

$$y(t) = \int_{-\infty}^{\infty} u(\tau)h(t-\tau)d\tau = \int_{-\infty}^{\infty} h(\tau)u(t-\tau)d\tau = u(t) * h(t)$$

в системах цифровой обработки вводят *линейную дискретную свертку*, представляющую собой вещественный дискретный сигнал, отсчеты которого связаны с отсчетами двух вещественных дискретных сигналов $\{u_k\}$ и $\{h_m\}$ формулой

$$y_m = \sum_{k=0}^{\infty} u_k h_{m-k} = \sum_{k=0}^{\infty} h_k u_{m-k}, \quad m = 0, 1, 2, \ldots \tag{8.31}$$

В (8.31) суммирование по номерам ведется от $k = 0$, поскольку исследуются вещественные сигналы. Если первый сигнал $\{u_k\}$ является обрабатываемым дискретным сигналом с числом отсчетов $k$, а второй сигнал $\{h_m\}$ — импульсной характеристикой обрабатывающей цифровой системы с числом отсчетов (ядром свертки; *kernel convolution*) $m$, то число выходных отсчетов в дискретной свертке сигнала будет $N = k + m - 1$, т. е. операция свертки *расширяет выходной сигнал* на $m - 1$ точку. Это фундаментальное свойство линейных дискретных систем. Знание дискретной импульсной характеристики позволяет оценить прохождение через дискретную систему любого сигнала. Нетрудно убедиться, что операция свертки коммутативна, т. е. допускает изменение порядка следования функций:

$$u[n] * h[n] = h[n]\ y(t) = \int_{-\infty}^{\infty} u(\tau)h(t-\tau)d\tau = \int_{-\infty}^{\infty} h(\tau)u(t-\tau)d\tau = u(t) * h(t)\ u[n],$$

т. е. можно переставлять местами исходный сигнал и ядро свертки.

Заметим, что линейную свертку (8.31), являющейся основой алгоритма дискретной фильтрации, не следует путать с *круговой сверткой* периодических последовательностей при исследовании свойств дискретного преобразования Фурье.

**Пример 8.9.** Посчитать, сколько умножений нужно произвести для вычисления свертки сигнала длины $n$ с ядром импульсной характеристики длины $m$.

Р е ш е н и е. Для вычисления свертки первым рассмотренным способом нужно каждую из $n$ точек исходного сигнала перевести в $m$ точек импульсной характеристики и сложить соответствующие точки импульсной характеристики. Таким образом, для каждой точки исходного сигнала нужно умножить на ее значение вектор из $m$ точек импульсной характеристики ($m$ умножений). Итак, всего для вычисления свертки требуется $nm$ умножений.

**Пример 8.10.** Для двух дискретных сигналов (в реальности один из них $h_k$ — отсчеты импульсной характеристики фильтра), заданных соответственно отсчетами $\{u_k\} = \{1, 2, 3\}$ и $\{h_k\} = \{4, 3, 2, 1\}$, вычислить дискретную свертку.

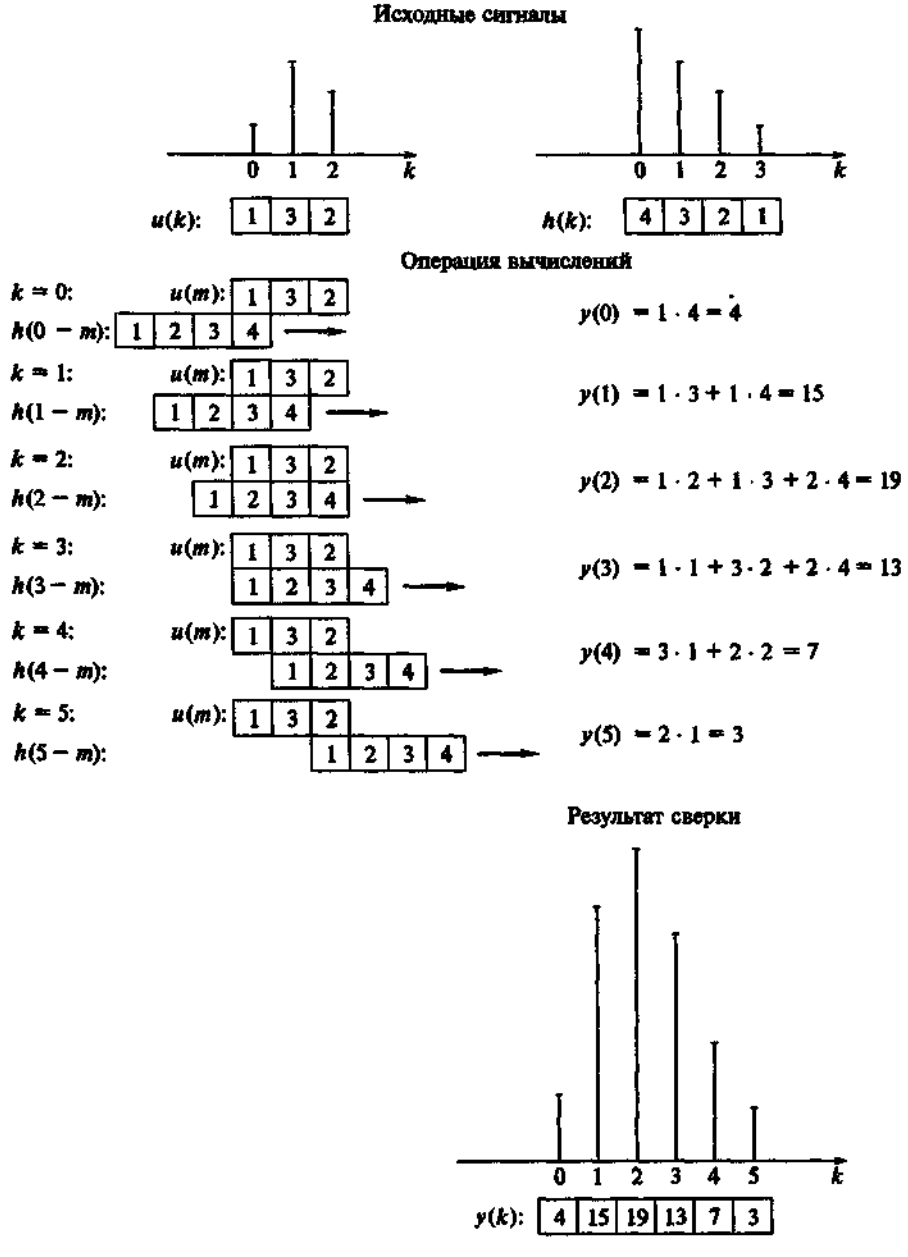



**Рис. 8.14.** К вычислению дискретной свертки

Р е ш е н и е. Воспользовавшись алгоритмом дискретной свертки (8.31), осуществим непосредственное вычисление ее отсчетов.

Для этого на одной полоске тетрадной клетчатой бумаги запишем отсчеты первого сигнала $\{u_k\}$, а на другой — отсчеты второго сигнала $\{h_m\}$, причем в последней позиции элементов второго сигнала отсчеты должны быть расположены «зеркально», т. е. справа налево (рис. 8.14).

Чтобы определить нулевой отсчет свертки, совместим первые позиции сигналов и перемножим отсчеты, находящиеся друг под другом. Суммируя результаты, имеем $y(0) = y_0 = 1 \cdot 4 = 4$. Для вычисления следующего отсчета $y_1$ сдвинем любую полоску на одну позицию. В данном случае, после перемножения отсчетов и сложения результатов, получим $y_1 = 1 \cdot 3 + 3 \cdot 4 = 15$. Проделав аналогичные операции до момента, когда отсчеты перестанут накладываться, находим значения свертки:

$$\{y_k\} = \{4, 15, 19, 13, 7, 3\}.$$

### Представление дискретных сигналов с помощью z-преобразования

При описании дискретных цепей широко применяют *z-преобразование*, играющее по отношению к дискретным сигналам такую же роль, как интегральные преобразования Фурье и Лапласа для аналоговых сигналов. Применение основных свойств z-преобразования резко упрощает анализ дискретных сигналов. Z-преобразование — одна из форм преобразования Лапласа и поэтому обладает подобными свойствами. Польза z-преобразования в том, что математика непрерывных величин разработана лучше дискретной, что позволяет использовать мощь дифференциального и интегрального исчисления, алгебры и т. п.

По аналогии с преобразованием Лапласа (4.10) для аналогового сигнала $u(t)$ для дискретного сигнала $\{u_k\}$ изображение по Лапласу определяется формулой

$$\mathrm{U}_T(p) = \sum_{k=0}^{\infty} u_k \mathrm{e}^{-pk\Delta t}. \tag{8.32}$$

Обратное дискретное преобразование Лапласа заданной последовательности, аналогично (4.13), имеет вид:

$$u_k = \frac{1}{2\pi j} \int_{\alpha_1 - j\infty}^{\alpha_1 + j\infty} U_T(p) \mathrm{e}^{pk\Delta t} dp, \tag{8.33}$$

где $\alpha_1$ — вещественная переменная на комплексной плоскости.

Изображения по Лапласу дискретных сигналов, в которые сомножителем входит экспоненциальный член $e^{p\Delta t}$, являются трансцендентными функциями аргумента $p$, что существенно усложняет анализ. Его можно упростить, переходя к z-преобразованию. Для этого в дискретном преобразовании Лапласа вводят новую переменную $z = e^{p\Delta t}$.

Рассмотрим дискретную последовательность $\{u_k\} = u_0, u_1, \ldots$, содержащую отсчеты значений некоторого непрерывного сигнала $u(t)$. Тогда на основании (8.33) *прямое z-преобразование* определяется суммой ряда по отрицательным степеням комплексной переменной $z$:

$$\mathrm{U}(z) = u_0 + \frac{u_1}{z} + \frac{u_2}{z^2} + \ldots = \sum_{k=0}^{\infty} u_k z^{-k}. \tag{8.34}$$

Функция $\mathrm{U}(z)$ определена только для области переменной $z$, в которой степенной ряд (8.34) сходится.

**Пример 8.11.** Вычислить $z$-преобразование дискретной последовательности отсчетов сигнала $\{u_k\} = (1, 1, 0, 1, 0, 0, 0, \ldots)$.

Р е ш е н и е. Применяя формулу (8.34) прямого $z$-преобразования, находим

$$U(z) = 1 + \frac{1}{z} + \frac{0}{z^2} + \frac{1}{z^3} = \frac{z^3 + z^2 + 1}{z^3}.$$

Широкое применение в теоретической радиотехнике находят экспоненциальные импульсные сигналы. Пусть имеется непрерывный сигнал $u(t) = e^{-\alpha t}$, заданный своими отсчетами $u_k$ в точках $t = k\Delta t$. Тогда его $z$-преобразование

$$U(z) = \sum_{k=0}^{\infty} u_k z^{-k} = \sum_{k=0}^{\infty} e^{-\alpha k\Delta t} z^{-k} = \frac{z}{z - e^{-\alpha\Delta t}} \qquad (8.35)$$

является аналитической функцией при переменной $|z| > e^{-\alpha\Delta t}$.

По аналогии с преобразованием Лапласа для аналоговых сигналов (4.13) возможно отыскание оригинала, т. е. исходной дискретной функции $u_k$, по заданному изображению $U(z)$. Для этого используется *обратное z-преобразование*

$$u_k = \frac{1}{2\pi j} \oint z^{k-1} U(z) dz. \qquad (8.36)$$

В правой части этого соотношения находится контурный интеграл, определенный для $z$-плоскости по любому замкнутому контуру в области сходимости, охватывающему начало координат. Отметим, что обход контура интегрирования проводится в положительном направлении, т. е. против часовой стрелки.

**Пример 8.12.** По заданному $z$-преобразованию вида $U(z) = z^{-2}$ определить отсчеты дискретного сигнала $u_k$.

Р е ш е н и е. Обратившись к формуле (8.36), запишем общее выражение для отсчетов дискретного сигнала

$$u_k = \frac{1}{2\pi j} \oint z^{k-3} dz.$$

Воспользуемся известной в математике теоремой Коши о вычетах. Согласно основному положению этой теоремы, имеем контурный интеграл вида

$$\oint z^n dz = \begin{cases} 2\pi j, & n = -1; \\ 0, & n \neq -1, \end{cases}$$

откуда находим отсчетные значения сигнала:

$$u_0 = \frac{1}{2\pi j} \oint z^{-3} dz = 0; \quad u_1 = \frac{1}{2\pi j} \oint z^{-2} dz = 0; \quad u_2 = \frac{1}{2\pi j} \oint z^{-1} dz = 1; \quad u_3 = \frac{1}{2\pi j} \oint dz = 0.$$

Итак, искомый дискретный сигнал имеет вид: $\{u_k\} = (0, 0, 1, 0)$.

В цифровой технике важное значение имеет $z$-преобразование дискретной свертки двух разных дискретных сигналов $\{u_k\}$ и $\{x_m\}$. С помощью формулы (8.31) вычислим $z$-преобразование линейной дискретной свертки этих сигналов:

$$\mathbf{F}(z) = \sum_{m=0}^{\infty}\sum_{k=0}^{\infty} u_k x_{m-k} z^{-m} = \sum_{m=0}^{\infty}\sum_{k=0}^{\infty} u_k z^{-k} x_{m-k} z^{-(m-k)} = \tag{8.37}$$

$$= \sum_{k=0}^{\infty} u_k z^{-k} \sum_{n=0}^{\infty} x_n z^{-n} = \mathbf{U}(z)\mathbf{X}(z),$$

где коэффициент $n = m - k$.

Анализ формулы (8.37) показывает, что линейной дискретной свертке двух дискретных сигналов соответствует произведение их $z$-преобразований.

## 8.5. Основы теории цифровой фильтрации. Синтез цифровых фильтров

Предмет цифровой фильтрации сигналов является естественным введением в широкую и фундаментальную область цифровой обработки информации. Под фильтрацией понимают любое преобразование информации (сигналов, результатов измерений и т. д.), при котором в входной последовательности обрабатываемых данных целенаправленно изменяют определенные соотношения (динамические или частотные) между различными компонентами этих данных. К основным операциям фильтрации относят операции дифференцирования, интегрирования, сглаживания, прогнозирования и разделения сигналов, а также фильтрацию, при которой осуществляется селекция (выделение) полезных составляющих сигнала и подавление мешающих его компонент и шумов (помех).

### Цифровая фильтрация сигналов

Как и в аналоговых цепях, в зависимости от способа определения параметров фильтра — по импульсной характеристике $\{h_k\}$ или частотному коэффициенту передачи $\mathbf{K}_a(\omega)$ — возможны временно́й и частотный подходы. При временно́м подходе используют дискретную свертку (8.31), а при частотном — вычисляют ДПФ (БПФ) с последующим применением ОДПФ (ОБПФ).

Цифровые фильтры обладают целым рядом преимуществ по сравнению с аналоговыми. Прежде всего, это их стабильность и точность, а также возможность гибкой и оперативной перестройки структуры и параметров фильтров.

Эти свойства цифровых фильтров в сочетании с возможностями полупроводниковой, а теперь и нанотехнологии обусловливают целесообразность их применения при обработке аналоговых сигналов для решения различных прикладных задач. Выбор типа и параметров фильтра определяется целями, поставленными при обработке сигналов в конкретных условиях. Обычно требо-

вания к цифровому фильтру формируют в процессе анализа исходных данных о спектральном составе полезного сигнала, мешающих сигналов и шумов. Цифровой фильтр должен обеспечить максимально возможное подавление помех при минимально допустимых искажениях полезного сигнала.

Цифровые фильтры имеют ряд принципиальных отличий от аналоговых. Основное отличие заключается в том, что в аналоговых фильтрах входной и выходной сигналы $u(t)$ и $y(t)$ являются непрерывными функциями времени, а в цифровых — дискретными последовательностями цифровых отсчетов соответствующих сигналов $u_T(t) = u_k$, $y_T(t) = y_k$.

Графики входных и выходных сигналов аналогового (рис. 8.15, *а*, *б*) и цифрового (рис. 8.15, *в*, *г*) фильтров (эти рисунки логически связаны с рис. 4.46 и рис. 4.47) позволяют наглядно представить отмеченную особенность цифровой фильтрации. Совершенно очевидно, что дискретность цифровых сигналов в принципе исключает возможность полного совпадения характеристик цифрового фильтра и его аналогового прототипа.

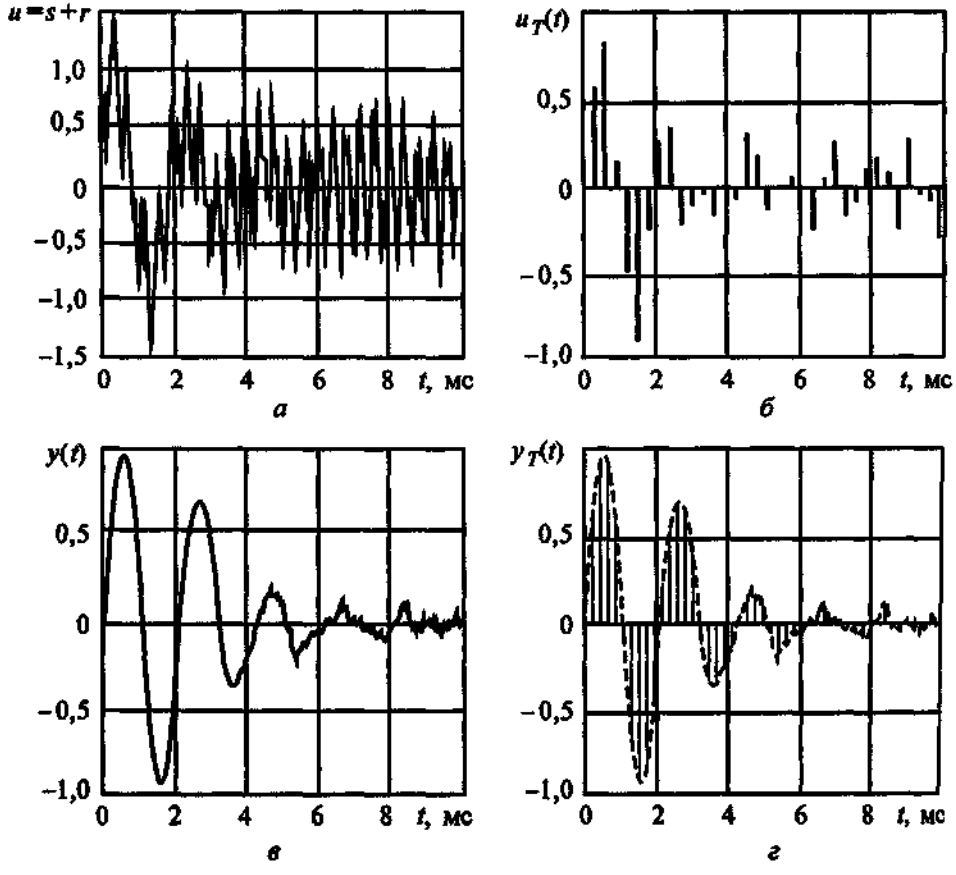



**Рис. 8.15.** К фильтрации сигналов:
*а* — исходный сигнал с шумом; *б* — сигнал на выходе аналогового фильтра;
*в* — дискретизированный исходный сигнал с шумом; *г* — сигнал на выходе цифрового фильтра

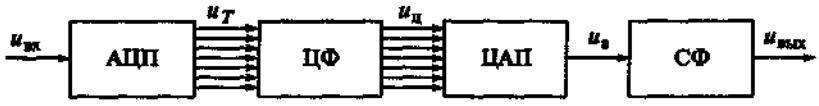



**Рис. 8.16.** Упрощенная структурная схема цифровой обработки сигналов

При цифровой фильтрации непрерывный во времени сигнал $u(t)$ с помощью аналого-цифрового преобразователя заменяется последовательностью дискретных отсчетов $u_Т(t)$, взятых, согласно теореме Котельникова, через интервал дискретизации $\Delta t$. Цифровая фильтрация заключается в цифровом преобразовании последовательности числовых отсчетов входного сигнала $\{u(k\Delta t)\} = \{u_k\}$ в последовательность числовых отсчетов выходного сигнала $\{y(k\Delta t)\} = \{y_k\}$.

*Системой цифровой обработки сигналов* (коротко системой) называют *объект*, выполняющий цифровое преобразование (обработку) входного сигнала в выходной. По умолчанию далее будем подразумевать цифровые системы *с одним входом и одним выходом*. Рассмотрим упрощенную структурную схему цифровой обработки сигналов (рис. 8.16). Аналоговый сигнал $u_{вх}(t) = u_{вх}$ (далее в схеме для удобства аргумент $t$ опущен) поступает на вход АЦП, на выходе которого создается цифровой код в виде двоичных чисел $\{u(k\Delta t)\} = \{u_Т\} = u_Т$ с фиксированным количеством разрядов, соответствующий дискретным отсчетам входного сигнала. Последовательность отсчетов $\{u_Т\}$, закодированных цифрами, поступает в *цифровой фильтр* (ЦФ), представляющий собой специализированный микрокомпьютер (ранее — спецвычислитель). В цифровом фильтре проводится цифровая обработка сигнала в соответствии с заданным алгоритмом, в результате чего на его выходе появляются новые цифровые коды $\{u_ц(k\Delta t)\} = \{u_ц\} = u_ц$, соответствующие профильтрованному входному сигналу.

Обычно обрабатываемый сигнал с цифрового фильтра поступает на ЦАП, в котором цифровая форма сигнала преобразуется в аналоговую $u_а(t) = u_а$. При включении ЦАП в схему обработки на ее выходе необходим *синтезирующий фильтр* (СФ) низкой частоты, который будет производить *сглаживание* полученного аналогового сигнала. В результате сглаживания форма кривой выходного сигнала $u_{вых}(t) = u_{вых}$ становится плавной, а не ступенчатой (по существу ЦАП интерполирует дискретный сигнал в непрерывный). Кстати, использование входных и выходных сигналов в аналоговой форме (и, следовательно, наличие АЦП и ЦАП) не всегда является необходимым. Так, при реализации цифрового генератора сигналов (или при поступлении цифровых сигналов по линиям связи) не нужен входной аналоговый сигнал, а ЦАП может отсутствовать, если конечный результат необходим только в цифровой форме.

Отметим важнейшее для цифровых фильтров обстоятельство. В общем случае все основные положения теории линейных аналоговых фильтров, оперирующих с аналоговыми сигналами, переносятся на математическую теорию цифровых фильтров, обрабатывающих дискретные и цифровые сигналы. Напомним, что, согласно формуле (4.21), выходной аналоговый сигнал $y(t)$ линей-

ного четырехполюсника (в данном случае фильтра) определяется сверткой его импульсной характеристики $h(t)$ и входного сигнала $u(t)$:

$$y(t) = \int_{-\infty}^{\infty} u(\tau)h(t-\tau)d\tau.$$

### Цифровые фильтры

*Цифровым фильтром* называют цифровое устройство, преобразующее исходную последовательность числовых отсчетов $\{u(k\Delta t)\} = \{u_T\} = \{u_k\}$ входного сигнала в заданную последовательность отсчетов $\{y(k\Delta t)\} = \{y_n\} = \{y_k\}$ выходного сигнала:

$$\{u_k\} \Rightarrow \{y_k\}. \tag{8.38}$$

Пусть последовательность чисел $\{u_k\}$ представляет собой отсчеты входного сигнала $u(t)$, взятые с интервалом дискретизации Котельникова $\Delta t$. Выходным сигналом цифрового фильтра будет последовательность цифровых отсчетов $\{y(k\Delta t)\} = \{y_k\} = y_k$. Для описания реакции цифрового фильтра на воздействие в виде последовательности чисел используют его *импульсную характеристику* $\{h(k\Delta t)\} = \{h_k\} = h_k$. Из теории цифровых фильтров известно, что упомянутая импульсная характеристика представляет собой дискретную последовательность чисел, являющуюся его реакцией на «единичный импульс» — *дискретизированную дельта-функцию* $\delta(k\Delta t) = \delta_k$ (в цифровом представлении это 1, 0, 0, 0, ...):

$$(1, 0, 0, 0, ..., ...) \Rightarrow \{h_0, h_1, h_2, ..., ..., h_k\} = h_k. \tag{8.39}$$

Дискретизированную дельта-функцию обозначают как

$$\delta[n] = \begin{cases} 1, & n = 0; \\ 0, & n \neq 0, \end{cases} \tag{8.40}$$

и называют *цифровой дельта-функцией*, или *функцией Кронекера* (рис. 8.17, *а*).

Очевидно, что любой дискретный сигнал можно разложить в сумму таких функций, сдвинутых во времени (рис. 8.17, *б*).

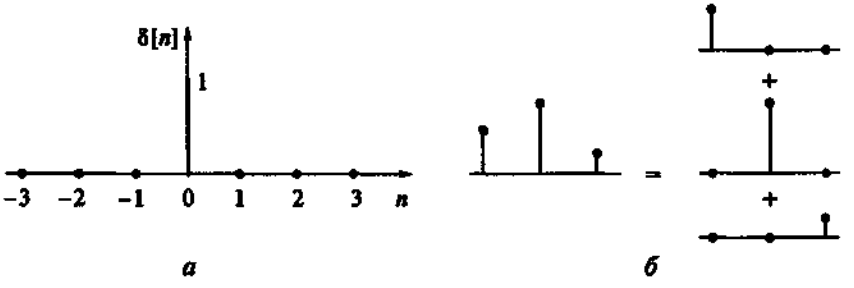



**Рис. 8.17.** Цифровая дельта-функция:
*а* — график; *б* — представление сигнала суммой цифровых дельта-функций

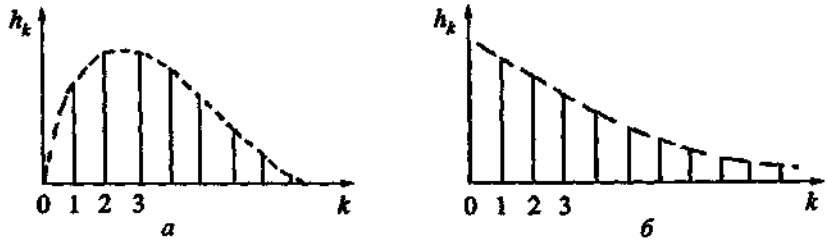



**Рис. 8.18.** Импульсные характеристики цифровых фильтров с числом отсчетов: *а* — конечным; *б* — бесконечным

Например, бесконечный сигнал $u\{n\}$ можно представить в виде

$$u\{n\} = u_n = \sum_{k=-\infty}^{\infty} u\{k\}\delta(n-k) = \sum_{k=-\infty}^{\infty} u_k \delta_{n-k}. \qquad (8.41)$$

Здесь дельта-функции — «базисные функции», а $u\{k\}$ — их коэффициенты в линейной комбинации. Если в последней формуле зафиксировать любое $n$, то получаем тождество $u\{n\} = u\{n\}\cdot 1$, поскольку все остальные члены суммы обратятся в 0, а дельта-функция отлична от нуля только в нуле.

Импульсную характеристику $h_k$ цифрового фильтра можно трактовать как результат дискретизации согласно теореме отсчетов непрерывной импульсной характеристики $h(t)$ соответствующего *аналогового линейного фильтра.* В цифровых фильтрах импульсные характеристики могут иметь конечное и бесконечное число отсчетов (рис. 8.18).

Оказывается, зная $h\{n\}$ (отклик системы на цифровую дельта-функцию), можно вычислить отклик системы на любой входной дискретный сигнал. Действительно, поскольку любой входной сигнал является линейной комбинацией сдвинутых во времени дельта-функций, то выходной сигнал будет той же самой линейной комбинацией сдвинутых во времени функций $h\{n\}$. Это следует из линейности системы и инвариантности к сдвигу во времени.

Реальные цифровые фильтры, обрабатывающие дискретные сигналы, должны отвечать принципу причинности, запрещающему выходному сигналу принимать отличные от нуля значения до прихода первого ненулевого входного отсчета. Такие фильтры называют *каузальными* (напомним, *causal* — физически осуществимый), их импульсная характеристика должна отвечать требованию

$$h_n = 0 \text{ при } n < 0,$$

т. е. реакция фильтра не может предшествовать приложенному воздействию.

Отметим, что требование каузальности, естественное для аналоговых фильтров, не имеет место, например, при рассмотрении дискретных сигналов, описывающих пространственные распределения какой-либо величины.

Рассмотрим принцип действия цифрового фильтра с простейшими алгоритмом фильтрации и импульсной характеристикой с ограниченным числом дискретных отсчетов (рис.8. 18, *а*).

Пусть входная дискретная последовательность $\{u_k\}$ непрерывного сигнала $u(t)$ подается на вход цифрового фильтра с импульсной характеристикой $\{h_m\}$. В результате некоторых заданных операций на выходе фильтра возникает последовательность отсчетов $\{y_k\}$. При воздействии первого отсчета сигнала $u_0$ на выходе фильтра формируется последовательность откликов, определяемых его импульсной характеристикой $y_0 = u_0 h_m$. Воздействие второго отсчета $u_1$ приведет к появлению на выходе фильтра последовательности еще $m$ откликов, которые можно выразить как $y_1 = u_1 h_{m-1}$. Подобная картина будет иметь место при отклике цифрового фильтра на действие остальных отсчетов входной последовательности. В результате прохождения всего входного сигнала из $k$ отсчетов на выходе цифрового фильтра с импульсной характеристикой $h_m$ создастся сумма откликов. Общее выражение для выходных отсчетов примет вид:

$$y_k = u_0 h_m + u_1 h_{m-1} + \ldots + u_m h_0 = \sum_{k=0}^{m} u_k h_{m-k} = \sum_{k=0}^{m} u_{m-k} h_k. \qquad (8.42)$$

Соотношение (8.42) и определяет алгоритм линейной цифровой фильтрации во временно́й области. Этот алгоритм показывает, что выходная последовательность представляет собой *дискретную свертку входного сигнала с импульсной характеристикой цифрового фильтра*. Как уже отмечалось импульсную характеристику цифрового фильтра $h_m$ часто называют *ядром свертки*, а число ее отсчетов $m$ — длиной ядра свертки.

Полное число выходных отсчетов при этом составляет $N = m + k - 1$. Отметим очевидное: для физически реализуемых цифровых фильтров коэффициенты $h_{-1}$, $h_{-2}$, ... импульсной характеристики обращаются в нуль.

**Системная (передаточная) функция цифрового фильтра**. С помощью метода $z$-преобразования удается определить важные характеристики цифрового фильтра. Положим, что дискретным входному $\{u_k\}$ и выходному $\{y_k\}$ сигналам и импульсной характеристике цифрового фильтра $\{h_k\}$ имеются соответствующие $z$-преобразования $\mathbf{U}(z)$, $\mathbf{Y}(z)$ и $\mathbf{H}(z)$. Поскольку выходной сигнал фильтра есть свертка входного сигнала с импульсной характеристикой, то, согласно формуле (8.37), $z$-преобразованию выходного сигнала отвечает функция

$$\mathbf{Y}(z) = \mathbf{U}(z)\mathbf{H}(z). \qquad (8.43)$$

*Системная функция* (*передаточная функция*; *transfer function*) дискретной системы представляет собой отношение $z$-преобразования выходного сигнала к $z$-преобразованию входного сигнала и связана с импульсной характеристикой:

$$\mathbf{H}(z) = \frac{\mathbf{Y}(z)}{\mathbf{U}(z)} = \sum_{k=0}^{\infty} h_k z^{-k}. \qquad (8.44)$$

Данная функция играет для цифрового фильтра такую же роль, что и передаточная функция для аналогового фильтра. Очевидным является и то, что задержка по времени входного сигнала на один отсчет приводит к умножению функции $\mathbf{H}(z)$ на $z^{-1}$.

### Структурные схемы линейных цифровых фильтров

Цифровые фильтры делятся на два больших класса: *нерекурсивные* и *рекурсивные*. Термин «рекурсивный» связан с известным математическим приемом — рекурсией — циклическим обращением к вычисленным данным, полученным на предыдущих этапах математических операций.

**Нерекурсивные цифровые фильтры.** В нерекурсивных (*nonrecursive*), или *трансверсальных* (от англ. *transversal — поперечный*, с точки зрения их графического построения), цифровых фильтрах отклик зависит только от значений входной последовательности, и для формирования $k$-го выходного отсчета используются лишь предыдущие значения входных отсчетов. Такие фильтры обрабатывают входной дискретный сигнал $\{u_k\}$ в соответствии с алгоритмом:

$$y_k = a_0 u_k + a_1 u_{k-1} + a_2 u_{k-2} + \ldots + a_m u_{k-m}, \qquad (8.45)$$

где $a_0$, $a_1$, ..., $a_m$ — действительные постоянные (*весовые*) коэффициенты; $m$ — *порядок* нерекурсивного фильтра, т. е. максимальное число запоминаемых чисел. Нетрудно заметить очевидную схожесть формул фильтрации (8.44) и (8.45).

Аналитическую сторону смысла алгоритма (8.45) наглядно характеризует структурная схема цифрового фильтра, представленная на рис. 8.19.

Основой любого цифрового фильтра являются элементы задержки входной цифровой последовательности $\{u_k\}$ на интервал дискретизации $Z^{-1}$ (согласно свойствам $z$-преобразования, задержка дискретной последовательности на один интервал $\Delta t$ соответствует умножению ее $z$-преобразования на $z^{-1}$, поэтому элементы памяти, осуществляющие такую задержку, обозначены на структурной схеме фильтра $Z^{-1}$), а масштабные (весовые) блоки $a_m$, выполняющие в цифровой форме операции умножения на соответствующие коэффициенты. Часто, и по существу, элементы задержки называют *ячейками памяти*. Сигналы с масштабных блоков поступают в сумматор +, на выходе которого образуется последовательность отсчетов выходного сигнала $\{y_k\}$.

Отметим, что формулы (8.42) и (8.45) тождественны и поэтому коэффициенты $a_0$, $a_1$, $a_2$, ..., $a_m$ совпадают с соответствующими отсчетами импульсной характеристики цифрового фильтра $h_0$, $h_1$, $h_2$, ..., $h_m$.

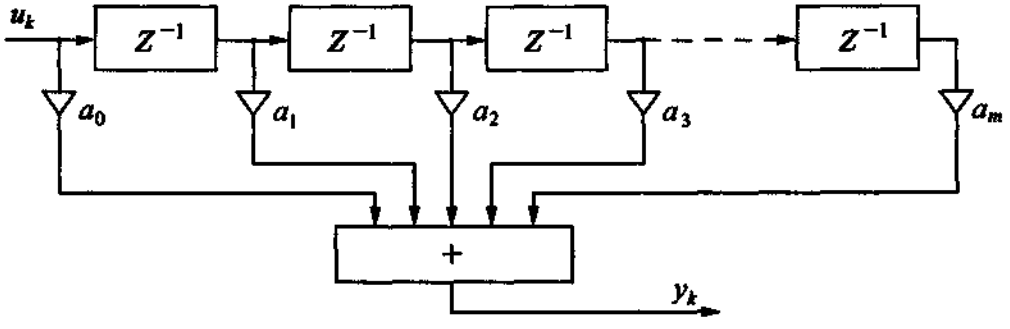



**Рис. 8.19.** Структурная схема нерекурсивного цифрового фильтра

**Пример 8.13.** Определить алгоритм работы нерекурсивного фильтра, у которого импульсная характеристика:

$$h_k = \begin{cases} 1, & k = 0; \\ 1, & k = 1; \\ 1, & k = 2; \\ 0, & k > 2. \end{cases}$$

Решение. Подставив в формулу (8.45) соответствующие отсчеты $h_k$, получим искомый алгоритм:

$$y_k = u_k \cdot 1 + u_{k-1} \cdot 1 + u_{k-2} \cdot 1 + u_{k-3} \cdot 0 = u_k + u_{k-1} + u_{k-2}.$$

Нерекурсивный цифровой фильтр может быть практически реализован, если заданная импульсная характеристика содержит ограниченное число отсчетов. Для получения же импульсной характеристики с большим числом членов требуется много ячеек памяти. В связи с этим нерекурсивные цифровые фильтры, как устройства, содержащие конечное число отсчетов импульсной характеристики, принято называть *фильтрами с конечными импульсными характеристиками* (КИХ-фильтрами). Примерный вид импульсной характеристики цифрового фильтра с конечным числом отсчетов показан на рис. 8.18, *а*.

Системную функцию нерекурсивного цифрового фильтра определим, применив $z$-преобразование к обеим частям уравнения (8.45):

$$\mathbf{Y}(z) = (a_0 + a_1 z^{-1} + a_2 z^{-2} + \ldots + a_m z^{-m})\mathbf{U}(z). \qquad (8.46)$$

Отсюда находим, что системная функция имеет вид:

$$\mathbf{H}(z) = a_0 + a_1 z^{-1} + a_2 z^{-2} + \ldots + a_m z^{-m} = \frac{a_0 z^m + a_1 z^{m-1} + a_2 z^{m-2} + \ldots + a_m}{z^m} \qquad (8.47)$$

и является дробно-рациональной функцией переменной $z$.

**Рекурсивные цифровые фильтры.** Возможности нерекурсивного цифрового фильтра существенно расширяются при введении в его схему обратных связей, которые позволяют формировать $k$-й выходной отсчет путем использования предыдущих значений как входного, так и выходного дискретных (выраженных в цифровой форме) сигналов

$$y_k = a_0 u_k + a_1 u_{k-1} + a_2 u_{k-2} + \ldots + a_m u_{k-m} + b_1 y_{k-1} + b_2 y_{k-2} + \ldots + b_n y_{k-n}. \qquad (8.48)$$

Так как при вычислениях используют предыдущие отсчеты выходного сигнала, в схеме присутствуют обратные связи. Поэтому такие фильтры называют *рекурсивными* (*recursive*), также фильтрами с обратной связью, положительной или отрицательной в зависимости от знака суммы коэффициентов $a_m$.

В формуле (8.48) коэффициенты $a_0$, $a_1$, $a_2$, ..., $a_m$, как и в (8.45), характеризуют нерекурсивную, а коэффициенты $b_1$, $b_2$, ..., $b_n$ — рекурсивную часть алгоритма цифровой фильтрации, причем последние не равны нулю одновременно.

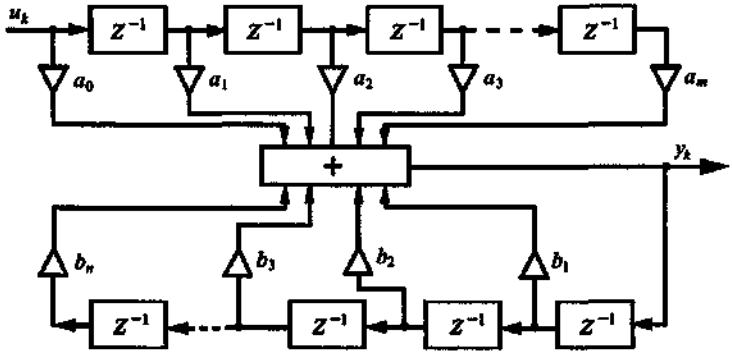



**Рис. 8.20.** Структурная схема рекурсивного цифрового фильтра

Порядок такого цифрового фильтра определяется коэффициентом $m$ нерекурсивной части алгоритма обработки. Структурная схема цифрового рекурсивного фильтра показана на рис. 8.20.

Системную функцию рекурсивного цифрового фильтра определим, применив $z$-преобразование к обеим частям выражения (8.48)

$$\mathbf{H}(z)=\frac{\mathbf{Y}(z)}{\mathbf{U}(z)}=\frac{a_0+a_1z^{-1}+\ldots+a_mz^{-m}}{1-b_1z^{-1}-\ldots-b_nz^{-n}}=\frac{a_0z^n+a_1z^{n-1}+\ldots+a_mz^{n-m}}{z^n-b_1z^{n-1}-\ldots-b_n}. \quad (8.49)$$

Таким образом, системная функция *физически реализуемого рекурсивного фильтра* может быть представлена в виде отношения полиномов по отрицательным степеням переменной $z$.

Основное достоинство рекурсивных фильтров — существенно меньшее количество элементов по сравнению с их числом в нерекурсивных фильтрах, выполняющих те же операции. Рекурсивные фильтры имеют определенную «память» по значениям предыдущих отсчетов, которая может быть бесконечной. Поэтому рекурсивные фильтры называют фильтрами с *бесконечной импульсной характеристикой* (БИХ-фильтр) (см. рис. 8.18, *б*). Реакция рекурсивного фильтра на сигнал с учетом «памяти» исключает возможность создания фильтров с четным импульсным откликом, и частотные характеристики рекурсивных фильтров всегда являются комплексными. Если хотя бы один из коэффициентов $a_m$ или $b_n$ зависит от переменной $k$, то фильтр является параметрическим, т. е. с переменными параметрами.

Далее будем рассматривать фильтры с постоянными коэффициентами (т. е. инвариантными по аргументу).

Структурную схему цифрового фильтра на рис. 8.20 принято называть *прямой формой* реализации рекурсивного фильтра (*direct form* I) и она не единственно возможная.

**Пример 8.14.** Вычислить импульсную характеристику рекурсивного цифрового фильтра, описываемого разностным уравнением 1-го порядка: $h_k = 0{,}5y_{k-1} + u_k$.

**Решение.** Пусть $y_{-1} = 0$, $u_k = \delta_k$ (первое равенство очевидно в силу условия физической реализуемости; второе — сигнал на входе при определении импульсной характеристики). Тогда выходной сигнал фильтра $y_k$ представляет его импульсную характеристику $h_k$, или $y_k = h_k = 0{,}5h_k + \delta_k$. Здесь $\delta_k = 1; 0; 0; \ldots$, ... Из этих выражений следует, что

$$h_0 = 0{,}5h_{-1} + \delta_0 = 1; \quad h_1 = 0{,}5h_0 + \delta_1 = 0{,}5; \quad h_2 = 0{,}5h_1 = 0{,}25.$$

Нетрудно заметить, что

$$h_k = (0{,}5)^k.$$

**Пример 8.15.** Определить структуру цифрового фильтра с системной функцией

$$\mathbf{H}(z) = \frac{1 - z^5}{1 - z}.$$

**Решение.** Прямым делением числителя на знаменатель получаем

$$\mathbf{H}(z) = 1 + z + z^2 + z^3 + z^4.$$

Методом обратного z-преобразования находим

$$h_k = \{1, 1, 1, 1, 1\}.$$

Данный цифровой фильтр является нерекурсивным (КИХ-фильтром).

**Пример 8.16.** Определить структуру цифрового фильтра с системной функцией

$$\mathbf{H}(z) = 1/(1 - 2z)$$

**Решение.** Методом обратного $z$-преобразования находим

$$h_k = 2^k.$$

Данный цифровой фильтр является рекурсивный (БИХ-фильтром).

**Пример 8.17.** Построить цифровой фильтр, соответствующий аналоговой цепи в виде колебательного контура, имеющего импульсную характеристику

$$h(t) = e^{-\alpha t}\cos\omega_0 t.$$

**Решение.** Импульсная характеристика цифрового фильтра будет представлена следующим образом:

$$h(k\Delta t) = h_k = e^{-\alpha k\Delta t}\cos(\omega_0 k\Delta t) = 0{,}5(e^{-\alpha k\Delta t + j\omega_0 k\Delta t} + e^{-\alpha k\Delta t - j\omega_0 k\Delta t}).$$

Системная функция, соответствующая этой импульсной характеристике:

$$\mathbf{H}(z) = \frac{1 - e^{-\alpha\Delta t}\cos\omega_0\Delta t \cdot z^{-1}}{1 - 2e^{-\alpha\Delta t}\cos\omega_0\Delta t \cdot z^{-1} + e^{-2\alpha\Delta t}z^{-2}} = \frac{a_0 + a_1 z^{-1}}{1 - b_1 z^{-1} - b_2 z^{-2}}.$$

Здесь $a_0 = 1$; $a_1 = e^{-\alpha\Delta t}\cos\omega_0\Delta t$; $b_1 = 2e^{-\alpha\Delta t}\cos\omega_0\Delta t$; $b_2 = -e^{-2\alpha\Delta t}$.

Данной системной функции отвечает уравнение

$$y(k\Delta t) = u(k\Delta t) + a_1 u(k-1)\Delta t + b_1 y(k-1)\Delta t + b_2 y(k-2)\Delta t.$$

Схема рекурсивного цифрового фильтра, соответствующая этому алгоритму, показана на рис. 8.21.

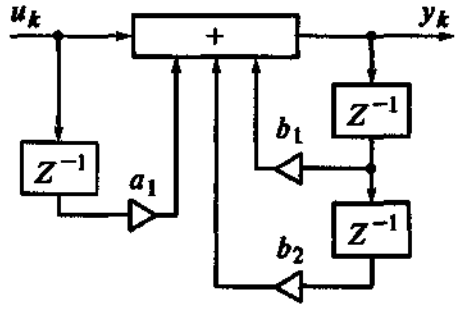



**Рис. 8.21.** Структурная схема цифрового фильтра к примеру 8.17

## Устойчивость цифровых фильтров

**Устойчивость нерекурсивных цифровых фильтров.** Поскольку аналоговыми прототипами нерекурсивых цифровых фильтров служат линейные цепи без ОС, то они заведомо относятся к устойчивым динамическим системам. Нерекурсивный цифровой фильтр всегда устойчив еще и потому, что имеет конечную (ограниченную во времени) импульсную характеристику.

**Устойчивость рекурсивных цифровых фильтров.** Рассмотрим вопрос об устойчивости рекурсивных цифровых фильтров, которые являются дискретными аналогами линейных систем с ОС, поскольку в ячейках памяти таких фильтров хранятся значения их предшествующих состояний. Если заданы некоторые начальные условия, т. е. совокупность значений $y_{i-1}$, $y_{i-2}$, ..., $y_{i-n}$, то при отключенном входном сигнале цифровой фильтр теоретически будет формировать элементы бесконечной последовательности отсчетов $y_i$, $y_{i+1}$, $y_{i+2}$, ..., $y_m$, которые можно считать его *свободными* (*собственными*) *колебаниями*.

С точки зрения импульсной характеристики устойчивость цифрового фильтра означает, что его реакция на ограниченное по размерам входное воздействие так же ограничена. Чтобы цифровой фильтр был устойчив, необходимо и достаточно выполнение условия

$$\sum_{n=0}^{\infty} \left| h(k\Delta t) \right| \le \infty. \qquad (8.50)$$

Анализ устойчивости цифровых фильтров может быть проведен по передаточной (системной) функции. В устойчивой системе значение $\mathbf{H}(z)$ должно быть конечным во всех точках $z$-плоскости, где $|z| \le 1$, а следовательно, передаточная функция не должна иметь особых точек (полюсов) на и внутри единичного круга на $z$-плоскости. Полюсы $\mathbf{H}(z)$ определяются корнями многочлена знаменателя передаточной функции (8.47) или (8.49).

**Пример 8.18.** Передаточная функция фильтра $\mathbf{H}(z) = b_0/(1 - a_1 z)$; $a_1 = 0{,}5$.

Решение. При $a_1 = 0{,}5$ полюс знаменателя $1 - a_1 z$: $z = 2$. Значит, $|z| > 1$ и фильтр устойчив.

**Пример 8.19.** Передаточная функция фильтра $H(z) = b_0/(1 + a_1 z)$; $a_1 = 1{,}1$.

**Решение.** При $a_1 = 1{,}1$ полюс знаменателя: $z = -0{,}909$. Значит, $|z| < 1$ и фильтр неустойчив.

Наиболее распространено другое понятие устойчивости цифрового фильтра, определяемое по характеру его свободного процесса. Рекурсивный цифровой фильтр является *устойчивым*, если возникающий в нем свободный процесс характеризует убывающую или постоянную последовательность элементов, т. е. значения $|y_n|$ при $n \to \infty$ не превышают некоторого положительного числа $M$ независимо от выбора начальных условий.

В рекурсивном цифровом фильтре на основании алгоритма (8.48) свободные колебания определяются в результате решения линейного разностного уравнения

$$y_k = b_1 y_{k-1} + b_2 y_{k-2} + \cdots + b_n y_{k-n}. \tag{8.51}$$

Решение уравнении (8.51) по аналогии с принципом решения линейных дифференциальных уравнений найдем в виде показательной функции

$$y_k = \alpha^k \tag{8.52}$$

с неизвестным пока значением $\alpha$. Подставив (8.52) в (8.51) и сократив на общий множитель, убеждаемся, что $\alpha$ является корнем характеристического уравнения:

$$\alpha^n - b_1\alpha^{n-1} - b_2\alpha^{n-2} - \cdots - b_n = 0. \tag{8.53}$$

Это уравнение в точности совпадает с уравнением, которому удовлетворяют полюсы системной функции (т. е. корни ее знаменателя) рекурсивного цифрового фильтра (8.49).

Положим, что корни $\alpha_1, \alpha_2, \ldots, \alpha_n$ уравнения (8.53) найдены. Тогда общее решение разностного уравнения (8.51) будет иметь следующий вид:

$$y_k = V_1\alpha_1^k + V_2\alpha_2^k + \cdots + V_n\alpha_n^k. \tag{8.54}$$

Значения постоянных коэффициентов $V_1, V_2, \ldots, V_n$ в (8.54) необходимо выбирать такими, чтобы удовлетворялись начальные условия.

Если все полюсы системной функции (8.49), т. е. модули постоянных вещественных чисел $z_1 = \alpha_1, z_2 = \alpha_2, \ldots, z_n = \alpha_n$ не превосходят единицы (по модулю меньше единицы), располагаясь внутри единичного круга с центром в точке $z = 0$, то на основании (8.54) любой свободный процесс в цифровом фильтре будет описываться членами убывающих геометрических прогрессий и фильтр является устойчивым. Говоря техническим языком, в данном случае система ни при каких входных воздействиях не пойдет «вразнос», т. е. не возбудится.

**Критерий устойчивости рекурсивных цифровых фильтров.** Вопрос об устойчивости рекурсивного фильтра произвольного порядка можно решить, связав эту задачу с определением расположения корней многочленов на комплекс-

ной плоскости. Напомним, что в соответствии с основной теоремой алгебры любой полином можно разложить на множители, соответствующие его корням.

Как уже отмечалось, преобразование

$$z = \frac{v+1}{v-1} \tag{8.55}$$

взаимно-однозначно отображает левую полуплоскость комплексной переменной $v$ на единичный круг в комплексной $z$-плоскости с центром координат в точке $z = 0$.

Действительно, точке $v = -1$ соответствует точка $z = 0$. При этом мнимая ось на $v$-плоскости, т. е. совокупность точек на плоскости с координатами $v = j\varepsilon$ ($\varepsilon$ — произвольное вещественное число), отображается в множество точек единичной окружности $z = -e^{j2\arctg\varepsilon}$.

Обратимся к характеристическому уравнению цифрового фильтра

$$z^n - b_1 z^{n-1} - b_2 z^{n-2} - \cdots - b_n = 0 \tag{8.56}$$

и подставим в него переменную $z$, связанную с переменной $v$ согласно (8.55):

$$\left(\frac{v+1}{v-1}\right)^n - b_1\left(\frac{v+1}{v-1}\right)^{n-1} - b_2\left(\frac{v+1}{v-1}\right)^{n-2} - \cdots - b_n = 0. \tag{8.57}$$

В соответствии с основной теоремой алгебры данный полином может быть приведен к общему знаменателю $(v-1)^n$. В результате получим характеристическое уравнение относительно переменной $v$:

$$(v+1)^n - b_1(v-1)(v+1)^{n-1} - b_2(v-1)^2(v+1)^{n-2} - \cdots - b_n(v-1)^n = 0. \tag{8.58}$$

Если левая часть данного многочлена по степеням $v$ имеет корни только в левой полуплоскости, то исходный характеристический многочлен (8.56), описывающий цифровой фильтр, имеет корни, располагающиеся лишь в единичном круге на $z$-плоскости. В этом случае рекурсивный фильтр является устойчивым.

Цифровые фильтры 1-го и 2-го порядков структурно проще, и их анализ проводится с помощью достаточно несложных операций. Кроме того, цифровые фильтры высоких порядков всегда можно представить в виде соединений цифровых фильтров 1-го 2-го порядков.

**Пример 8.20.** По описанному критерию исследовать устойчивость рекурсивного цифрового фильтра 3-го порядка с характеристическим уравнением

$$z^3 + 0{,}5z^2 - 0{,}4z + 1 = 0.$$

**Решение.** Воспользовавшись формулой (8.58), запишем преобразованное характеристическое уравнение

$$(v+1)^3 + 0{,}5(v-1)(v+1)^2 - 0{,}4(v-1)^2(v+1) + 1 = 1{,}1v^3 + 3{,}1v^2 + 2{,}9v + 1{,}9 = 0.$$

Здесь все коэффициенты положительны и в то же время $a_1a_2 - a_0a_3 = 6{,}9 > 0$. Данный многочлен отвечает условиям теоремы Гурвица (см. гл. 4, алгебраические критерии устойчивости). Значит, согласно критерию Рауса–Гурвица, анализируемый рекурсивный цифровой фильтр устойчив.

### Канонические схемы рекурсивных фильтров

Недостаток рекурсивных фильтров, реализуемых по алгоритму (8.48), это большое количество ячеек памяти, применяемых для рекурсивной и нерекурсивной частей. Уменьшить число ячеек позволяют *канонические* (по оптимальности реализации) схемы рекурсивных фильтров, в которых каждый элемент задержки используют как для нерекурсивной, так и рекурсивной связей. В канонических схемах число элементов задержки равно наибольшему из чисел $m$ и $n$.

Цифровой рекурсивный фильтр, соответствующий алгоритму обработки дискретного сигнала типа (8.49), может быть реализован в виде другой, эквивалентной схемы. Запишем уравнение (8.49) следующим образом:

$$\mathbf{Y}(z)=\frac{a_0+a_1z^{-1}+\cdots+a_mz^{-m}}{1-b_1z^{-1}-\cdots-b_nz^{-n}}\mathbf{U}(z). \tag{8.59}$$

Представим формулу (8.59) как

$$\mathbf{Y}(z)=(a_0+a_1z^{-1}+\cdots+a_mz^{-m})\mathbf{V}(z), \tag{8.60}$$

где

$$\mathbf{V}(z)=\frac{\mathbf{U}(z)}{1-b_1z^{-1}-\cdots-b_nz^{-n}}. \tag{8.61}$$

Алгоритм (8.60) определяет структуру построения нерекурсивного фильтра $m$-го порядка. Преобразование, соответствующее функции V(z) и заданное (8.61), осуществляют с помощью рекурсивного фильтра $n$-го порядка. Итак, общая схема рекурсивного фильтра включает в себя две части — нерекурсивную с коэффициентами $a_m$ и рекурсивную с коэффициентами $b_n$ (рис. 8.22).

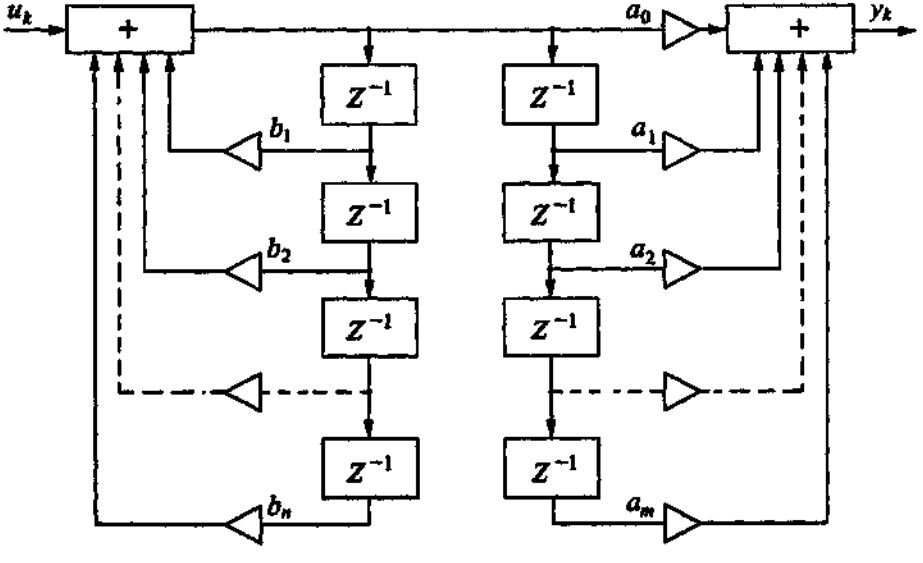



**Рис. 8.22.** Нерекурсивная и рекурсивная части цифрового фильтра

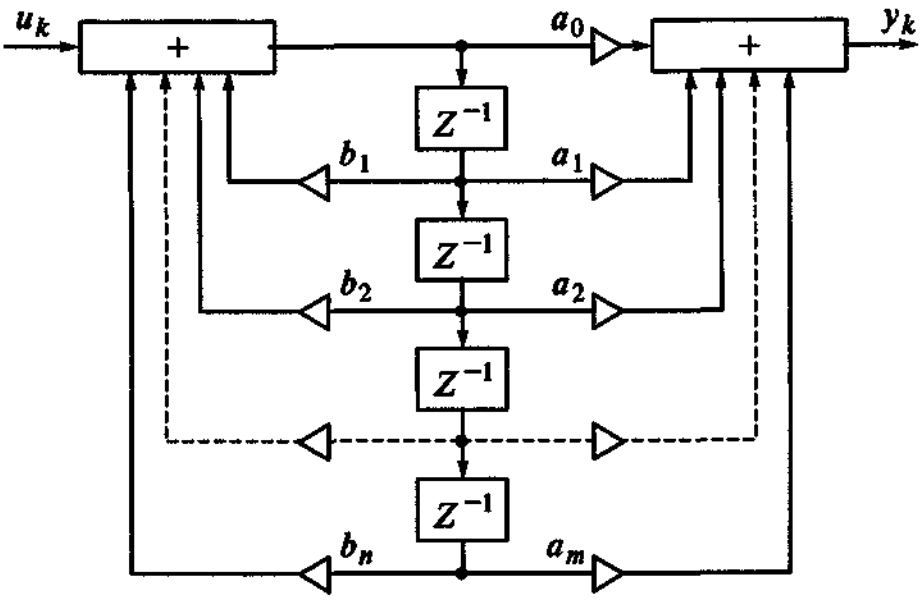



**Рис. 8.23.** Каноническая схема цифрового фильтра

Нетрудно заметить, что в обе части линий задержки подается один и тот же сигнал, поэтому они будут содержать одинаковые наборы отсчетов. Это позволяет объединить дублирующие линии задержки схемы рис. 8.22. Полученную схему фильтра (рис. 8.23) называют *канонической формой* (*canonic form*) реализации рекурсивного фильтра.

Заметим, что с теоретической точки зрения рекурсивная и каноническая схемы цифрового фильтра фактически эквивалентны. Однако при практической реализации цифровых фильтров необходимо обратить внимание на некоторые специфические особенности, присущие отмеченным схемам. В частности, при реализации канонической структуры цифрового фильтра для нерекурсивной и рекурсивной частей используют общую линию задержки, что существенно уменьшает число необходимых ячеек памяти. Но в этом случае абсолютные значения отсчетов, попадающих в линии задержки, могут существенно превосходить амплитуды отсчетов входного и выходного сигналов. Это приводит к необходимости увеличения разрядности представления чисел в линиях задержки по сравнению с разрядностью входного и выходного сигналов, что усложняет техническую реализацию фильтра.

При прямой реализации рекурсивных фильтров в линиях задержки хранят непосредственно отсчеты входного и выходного сигналов, т. е. разрядность линий задержки остается неизменной. Элементом, требующим повышения разрядности, в этом случае является сумматор, и это необходимо учитывать в архитектуре сигнальных микропроцессоров, специально предназначенных для обработки сигналов в реальном времени.

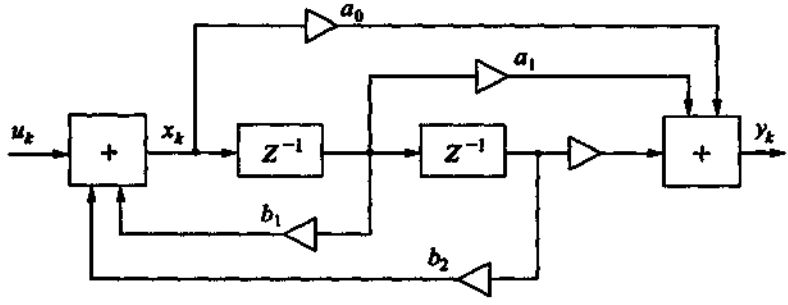



**Рис. 8.24.** Структурная схема канонического рекурсивного фильтра 2-го порядка

Рассмотрим реальную структурную схему канонического рекурсивного цифрового фильтра 2-го порядка (рис. 8.24), описываемого системной функцией

$$\mathbf{H}(z)=\frac{\mathbf{Y}(z)}{\mathbf{U}(z)}=\frac{a_0+a_1z^{-1}+\ldots+a_mz^{-m}}{1-b_1z^{-1}-\ldots-b_nz^{-n}}=\frac{a_0z^n+a_1z^{n-1}+\ldots+a_mz^{n-m}}{z^n-b_1z^{n-1}-\ldots-b_n}. \quad (8.62)$$

Для того чтобы убедиться в возможности реализации системной функции вида (8.62) заданной схемой канонического рекурсивного фильтра, введем вспомогательный дискретный сигнал $\{x_k\}$ на выходе первого сумматора и составим относительно него два следующих уравнения:

$$x_k = u_k + b_1x_{k-1} + b_2x_{k-2}; \quad (8.63)$$

$$y_k = a_0x_k + a_1x_{k-1} + a_2x_{k-2}. \quad (8.64)$$

Выполним два $z$-преобразования по уравнениям (8.63), (8.64) относительно дискретных сигналов $\{u_k\}$ и $\{y_k\}$:

$$\mathbf{U}(z) = \mathbf{X}(z)(1 - b_1z^{-1} - b_2z^{-2}); \quad (8.65)$$

$$\mathbf{Y}(z) = \mathbf{X}(z)(a_0 + a_1z^{-1} + a_2z^{-2}). \quad (8.66)$$

Воспользовавшись формулой (8.44) и поделив уравнение (8.66) на (8.65), находим, что полученная системная функция совпадает с заданной (8.62).

### Транспонированная форма цифровых фильтров

Поменяем в схеме нерекурсивного фильтра (см. рис. 8.19) последовательность выполнения операций умножения и задержки, использовав в каждой ветви отдельную линию задержки на требуемое количество дискретных отсчетов входного сигнала. Затем структурно разделим общий сумматор на несколько двухвходовых сумматоров с одним выходом в каждом. Получившаяся структура фильтра показана на рис. 8.25, *а*. Теперь, рассмотрев любую пару соседних сумматоров, можно заметить, что суммируемые ими дискретные сигналы претерпевают некоторую общую задержку. Это дает возможность поменять местами операции суммирования и задержки. Полученную схему (рис. 8.25, *б*) называют *транспонированной* (*transposed*) *реализацией* цифрового фильтра.

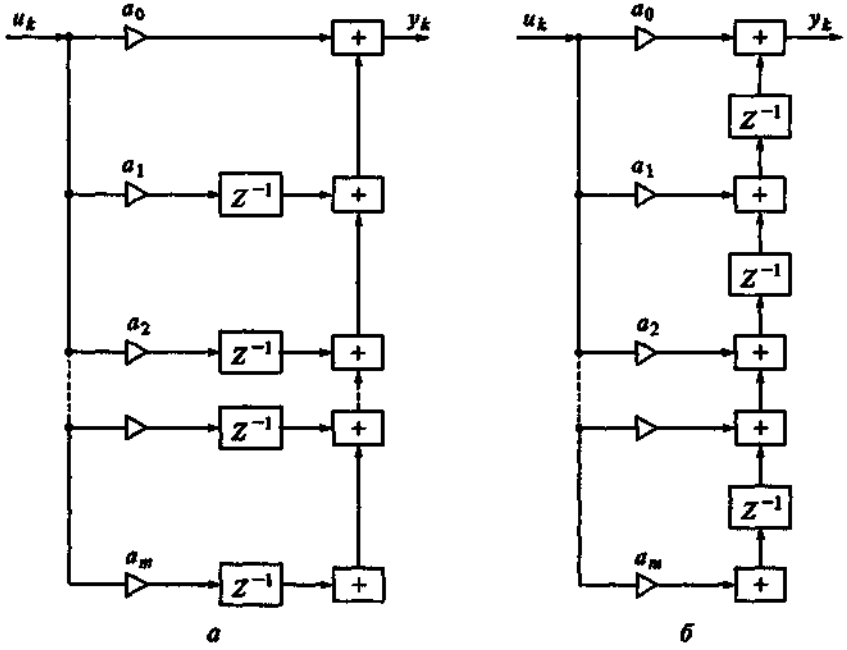



**Рис. 8.25.** Транспонированная форма цифрового фильтра:
*а* — изменение последовательности выполнения операций умножения и задержки;
*б* — транспонированная реализация нерекурсивного фильтра

Для упрощения преобразование в транспонированную форму проведено для нерекурсивного фильтра, однако его можно осуществить и для рекурсивного фильтра. Для этого в структурной схеме рис. 8.25 необходимо ввести все ветви с коэффициентами $b_m$, включая $b_0$ (рис. 8.26).

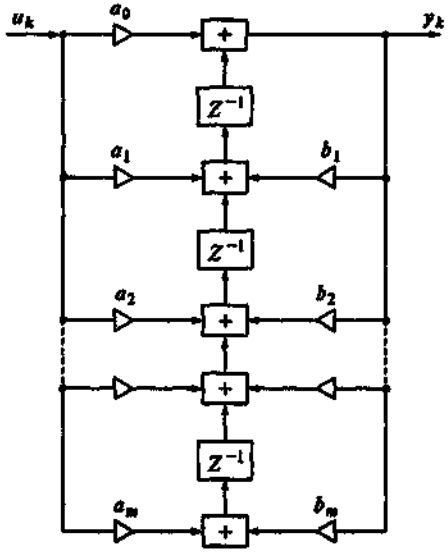



**Рис. 8.26.** Транспонированная реализация рекурсивного фильтра

Транспонированная форма цифрового фильтра позволяет эффективно распараллелить вычисления и потому применяется при реализации цифровых фильтров в виде специализированных интегральных микросхем, фактически сигнальных микропроцессоров. Действительно, при реализации фильтра в форме рис. 8.19 или 8.20 можно *одновременно* выполнять все операции *умножения*, но для получения выходного результата необходимо дождаться окончания выполнения *всех* операций сложения. В транспонированных же схемах фильтров, наряду с умножением, можно одновременно выполнять и все операции *сложения*, поскольку они являются независимыми, т. е. не используют в качестве суммируемых величин результаты других сложений.

Анализ схемы цифрового фильтра, приведенной на рис. 8.26, показывает, что для расчета значений выходного сигнала необходимо произвести *одно* умножение и *одно* сложение, а все остальные операции фактически выполняют подготовку промежуточных результатов для вычисления последующих отсчетов выходного сигнала.

**Цифровые фильтры на сплайн-функциях.** Вариантами рекурсивных цифровых фильтров являются фильтры, построенные на сплайн-функциях. Алгоритм данного фильтра основан на применении линейного преобразования на базе подвижного интервала времени, т. е. отфильтрованные значения вычисляют для средней точки интервала. Этот метод также характерен тем, что все отсчеты информационно связаны, но, как было показано, отсчеты вне интервала между отсчетами определенными по методу косинусоидальной аппроксимации с этим интервалом не связаны. Поэтому для фильтрации необходимо увеличивать число отсчетов. Следовательно, метод на основе сплайн-функций трансформируется просто к усреднению.

### Частотные характеристики цифровых фильтров

Наиболее важным и информативным показателем цифровых фильтров является *частотный коэффициент передачи*. Определение частотного коэффициента передачи цифрового фильтра становится достаточно простым и понятным, если условно входной тестовый сигнал представить в виде дискретной гармоники (косинусоиды) единичной амплитуды (рис. 8.27). Выборки непрерывного гармонического сигнала $u(t) = \cos\Delta t$, взятые с интервалом дискретизации $\Delta t$, описываются дискретной последовательностью

$$u_k = \cos(\omega k\Delta t) = e^{j\omega k\Delta t}. \qquad (8.67)$$

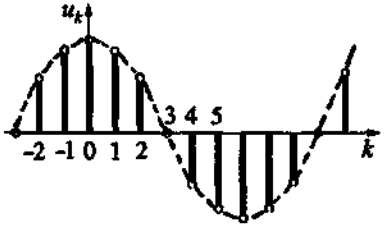



**Рис. 8.27.** Дискретная гармоника

Положим, что на вход линейного цифрового фильтра подается дискретная гармоническая последовательность (8.67), бесконечная во времени ($k = 0, \pm 1, \pm 2, \ldots$). Используя известную формулу дискретной свертки (8.42), запишем $m$-й отсчет выходного сигнала цифрового фильтра в следующем виде:

$$y_m = \sum_{k=-\infty}^{m} u_k h_{m-k} = \sum_{k=-\infty}^{m} e^{j\omega k\Delta t} h_{m-k} = e^{j\omega m\Delta t} \sum_{k=-\infty}^{m} e^{j\omega(k-m)\Delta t} h_{m-k}. \quad (8.68)$$

Просуммируем члены ряда по новому индексу $n = m - k$

$$y_m = e^{j\omega m\Delta t} \sum_{n=0}^{\infty} e^{-j\omega n\Delta t} h_n. \quad (8.69)$$

Из выражения (8.69) следует, что выходной сигнал цифрового фильтра имеет структуру дискретной гармонической последовательности с частотой колебаний, равной частоте входного сигнала ω.

С помощью (3.5) определим частотный коэффициент передачи цифрового фильтра. Поделив (8.69) на (8.67) и учитывая, что в формулах используются $m$-е отсчеты, находим комплексный частотный коэффициент передачи фильтра

$$\mathbf{K}_{\text{ц}}(\omega) = \sum_{n=0}^{\infty} h_n e^{-j\omega n\Delta t}. \quad (8.70)$$

Это соотношение позволяет сделать четыре очень важных вывода.

1. Частотные характеристики цифровых фильтров являются непрерывными функциями частоты.

2. Частотный коэффициент передачи цифрового фильтра $\mathbf{K}_{\text{ц}}(\omega)$, как и спектр дискретного сигнала $\mathbf{S}_T(\omega)$ (8.21), имеет периодическую структуру с периодом по оси частот, равным частоте дискретизации $\omega_1 = 2\pi/\Delta t$. Это фундаментальное положение для цифровых фильтров. Периодическая структура частотного коэффициента передачи цифрового фильтра позволяет либо выделить, либо подавить отдельные составляющие спектра дискретного входного сигнала. При этом первый низкочастотный период коэффициента передачи принято называть *главным частотным диапазоном*.

3. Функция $\mathbf{K}_{\text{ц}}(\omega)$ является дискретным преобразованием Фурье импульсной характеристики цифрового фильтра, аналитически представленной в виде последовательности дельта-функций:

$$h_\Delta(t) = h_0\delta(t) + h_1\delta(t - \Delta t) + h_2\delta(t - 2\Delta t) + \ldots$$

4. Для цифровых фильтров с вещественными коэффициентами импульсной реакции $h(n\Delta t)$ функция АЧХ является четной, а функция ФЧХ — нечетной. С учетом этого частотные характеристики фильтров обычно задаются только на интервале положительных частот 0 … $\omega_1$ главного частотного диапазона. Значения функций на интервале отрицательных частот являются комплексно-сопряженными со значениями на интервале положительных частот.

Представляет интерес связь между системной функцией и частотным коэффициентом передачи цифрового фильтра. Сравнив формулы (8.44) и (8.70), можно заметить, что для получения частотного коэффициента передачи достаточно в выражении для системной функции сделать подстановку $z = e^{j\omega\Delta t}$:

$$\mathbf{K}_{\text{ц}}(\omega) = \mathbf{H}(e^{j\omega\Delta t}). \quad (8.71)$$

Итак, системная функция, как и частотный коэффициент передачи, имеет периодическую структуру с периодом по оси частот, равным частоте дискретизации $\omega_1 = 2\pi/\Delta t$.

**Пример 8.21.** Алгоритм нерекурсивной цифровой фильтрации имеет следующий вид: $y_k = 10u_k + 5u_{k-1} + 2u_{k-2}$. Найти импульсную характеристику, системную функцию и частотный коэффициент передачи цифрового фильтра.

Решение. С помощью формул (8.42), (8.44) и (8.70) найдем соответственно $h_k$, $\mathbf{H}(z)$ и $\mathbf{K}_u(\omega)$.

1. Импульсная характеристика $\{h_k\} = (10, 5, 2)$.
2. Системная функция $\mathbf{H}(z) = 10 + 5z^{-1} + 2z^{-2}$.
3. Частотный коэффициент передачи найдем, заменив в формуле для $\mathbf{H}(z)$ переменную $z = e^{j\omega\Delta t}$:

$$\mathbf{K}_u(\omega) = 10 + 5e^{-j\omega\Delta t} + 2e^{-j2\omega\Delta t}.$$

Используя формулу Эйлера, окончательно запишем

$$\mathbf{K}_u(\omega) = 10 + 5\cos\omega\Delta t + 2\cos 2\omega\Delta t - j(5\sin\omega\Delta t + 2\sin 2\omega\Delta t).$$

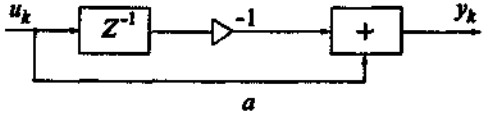
**Пример 8.22.** Рассчитать и построить АЧХ цифрового фильтра, структурная схема которого представлена на рис. 8.28, *а* для следующих значений фазовых углов $\omega\Delta t$ 0, 90, 180, 270 и 360°.

Решение. Алгоритм работы фильтра: $y_k = u_k - u_{k-1}$. Отсюда $z$-преобразование выходных отсчетов фильтра имеет вид

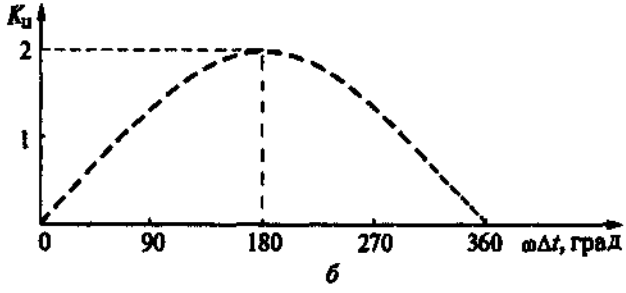
$$\mathbf{Y}(z) = \mathbf{U}(z) - \frac{\mathbf{U}(z)}{z} = \mathbf{U}(z)\frac{z-1}{z}.$$

Тогда системная функция фильтра

$$\mathbf{H}(z) = \frac{\mathbf{Y}(z)}{\mathbf{U}(z)} = \frac{z-1}{z}.$$




*а*




*б*

**Рис. 8.28. Цифровой фильтр:**
*а* — структурная схема; *б* — АЧХ

Заменив $z = e^{j\omega\Delta t}$, находим частотный коэффициент передачи:

$$K_ц(\omega) = 1 - e^{-j\omega\Delta t} = 1 - \cos\omega\Delta t + j\sin\omega\Delta t.$$

Как и в линейных аналоговых фильтрах, модуль частотного коэффициента передачи представляет АЧХ цифрового фильтра:

$$K_ц = |K_ц(\omega)| = \sqrt{1 - 2\cos\omega\Delta t + \cos 2\omega\Delta t}.$$

График АЧХ, построенный в соответствии с этой формулой, показан на рис. 8.28, *б*, где по оси абсцисс отложен фазовый угол $\omega\Delta t$.

**Пример 8.23.** Нерекурсивный цифровой фильтр второго порядка работает в соответствии с алгоритмом

$$y_k = u_k + u_{k-1} + u_{k-2}.$$

Определить и построить частотную характеристику цифрового фильтра. Проанализировать работу заданного фильтра при прохождении через него дискретной гармоники вида (8.67), имеющей интервал дискретизации $\Delta t = \pi/(3\omega)$.

Р е ш е н и е. Из алгоритма работы фильтра следует, что импульсная характеристика $\{h_k\} = 1, 1, 1$. Тогда системная функция фильтра

$$H(z) = 1 + z^{-1} + z^{-2}.$$

Заменив $z = e^{j\omega\Delta t}$, находим частотный коэффициент передачи

$$K_ц(\omega) = 1 + e^{-j\omega\Delta t} + e^{-j2\omega\Delta t} = 1 + \cos\omega\Delta t + \cos 2\omega\Delta t - j(\sin\omega\Delta t + \sin 2\omega\Delta t).$$

Так как модуль частотного коэффициента передачи представляет АЧХ цифрового фильтра, то

$$K_ц = |K_ц(\omega)| = \sqrt{3 + 4\cos\omega\Delta t + 2\cos 2\omega\Delta t}.$$

График АЧХ, соответствующий этому уравнению, показан на рис. 8.29, *а*, где по оси абсцисс отложен дискретный фазовый угол $\omega\Delta t$, значение которого соответствует интервалу дискретизации при текущем значении частоты.

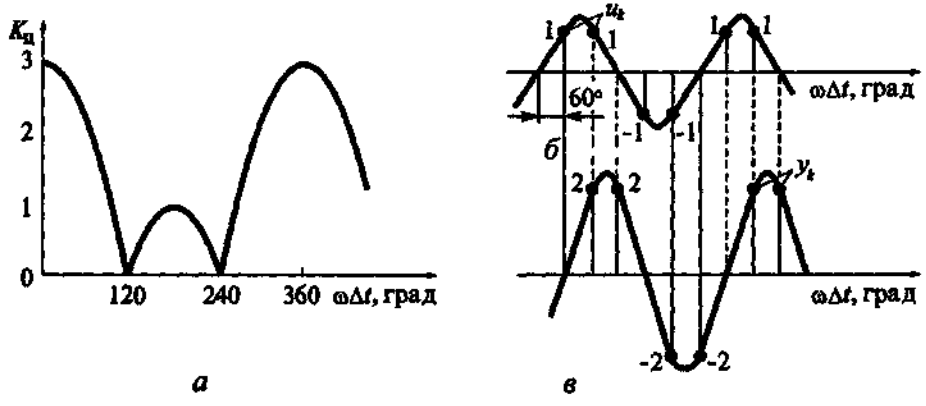



**Рис. 8.29.** К примеру 8.23:
*а* — АЧХ цифрового фильтра; *б* — сигнал на входе; *в* — сигнал на выходе

Заметим, что одному заданному интервалу дискретизации $\Delta t$ соответствует фазовый угол $\omega\Delta t = 60°$.

Фазочастотная характеристика анализируемого нерекурсивного цифрового фильтра определяется следующей формулой

$$\varphi_к(\omega) = -\text{arctg}\frac{\sin\omega\Delta t + \sin 2\omega\Delta t}{1 + \cos\omega\Delta t + \cos 2\omega\Delta t} = -\omega\Delta t$$

и имеет линейный вид.

Пусть на вход данного нерекурсивного фильтра подается дискретная гармоника единичной амплитуды (рис. 8.29, *б*). В этом случае цифровая входная последовательность будет иметь вид:

$$\ldots, 0, 1, 1, 0, -1, -1, 0, 1, 1, 0, \ldots$$

В соответствии с заданным алгоритмом выходной сигнал фильтра определится следующей последовательностью:

$$\ldots, 2, 2, 0, -2, -2, 0, 2, 2, \ldots$$

Как следует из значений выходной последовательности отсчетов фильтра, ей отвечает дискретная гармоника той же частоты, что и на входе. Амплитуда выходной гармоники равна удвоенной амплитуде входной гармоники, а начальная фаза смещена на 60° (т. е. на фазовый угол, соответствующий одному интервалу дискретизации) в сторону запаздывания (рис. 8.29, *в*).

Как правило, при частотном анализе цифровых фильтров значение интервала дискретизации $\Delta t$ принято принимать для упрощения за 1, что соответственно определяет задание частотных характеристик на интервале $(0, \pi)$ по круговой частоте $\omega$, или $(0, 1/2)$ по циклической частоте $f$. При использовании быстрых преобразований Фурье вычисления спектров осуществляют обычно в одностороннем варианте положительных частот в приведенном частотном интервале от 0 до $2\pi$ (0…1 Гц), где комплексно-сопряженная часть спектра главного диапазона $(-\pi\ldots 0)$ занимает интервал от $\pi$ до $2\pi$ (для ускорения вычислений используется принцип периодичности дискретных спектров). Заметим, что при выполнении БПФ количество точек спектра равно количеству точек входной функции, а следовательно, отсчет на частоте $2\pi$, комплексно-сопряженный с отсчетом на частоте 0, отсутствует. При нумерации точек входной функции от 0 до $N$ он принадлежит точке $N+1$ — начальной точке следующего периода, при этом шаг по частоте равен $2\pi/(N+1)$.

Современное программное обеспечение алгоритма БПФ допускает практически любое реальное число точек входного сигнала, причем для нечетного значения отсчетов $N$ частоте $\pi$ соответствует отсчет на точке $(N+1)/2$, не имеющий сопряженного отсчета, а при четном значении отсчетов $N$ отсутствует отчет на частоте $\pi$ (она располагается между отсчетами $k = N/2$ и $k = N/2 + 1$). При этом отсчетам с номерами $k$ главного диапазона БПФ (за исключением точки $k = 0$) соответствуют комплексно-сопряженные отсчеты $N + 1 - k$ (за исключением точки $k = (N+1)/2$ при нечетном $N$).

### Синтез цифровых фильтров

Разработка и практическое применение цифровых фильтров стали возможными во многом благодаря теории линейных аналоговых фильтров. Основные успехи были достигнуты при использовании некоторых важных соответствий между характеристиками цифровых и аналоговых фильтров.

*Задача синтеза цифрового фильтра* состоит в определении системной (передаточной) функции, расчете параметров цифрового фильтра, эквивалентного аналоговому прототипу, и разработке на основе полученных данных алгоритма фильтрации.

*Цель синтеза цифрового фильтра* заключается в том, чтобы цифровой тракт (аналого-цифровой преобразователь и цифровой фильтр) вырабатывал для заданного класса входных сигналов $u(t)$ последовательность выходных отсчетов $y_T(t) = y_k$, с гарантированной точностью совпадающих с дискретными значениями выходного сигнала гипотетического аналогового фильтра-прототипа $y(t)$.

Желаемые характеристики цифрового фильтра, задаваемые обычно в виде дробно-рациональной передаточной функции или частотных и временных характеристик, представляют исходные данные для задачи синтеза. Как отмечалось в главе 4, передаточную функцию аналогового фильтра записывают как

$$K(p) = \frac{R(p)}{Q(p)} = \frac{b_m p^m + b_{m-1} p^{m-1} + \cdots + b_1 p + b_0}{a_n p^n + a_{n-1} p^{n-1} + \cdots + a_1 p + a_0}. \qquad (8.72)$$

Аргумент $p$ в (8.72) представляет собой оператор дифференцирования $p = d/dt$, а сама передаточная функция определяет дифференциальное уравнение аналогового фильтра, которое можно представить в следующем виде:

$$\sum_{i=0}^{n} a_i \frac{d^i}{dt^i} u(t) = \sum_{j=0}^{m} b_j \frac{d^j}{dt^j} y(t), \qquad (8.73)$$

где $u(t)$ и $y(t)$ — входной и выходной аналоговые сигналы.

Для синтеза цифровых фильтров необходимо осуществить прежде всего дифференцирование передаточной функции. Однако дифференцировать можно только непрерывные функции, а для дискретных числовых последовательностей операция дифференцирования не определена. Следовательно, цифровые фильтры не могут иметь математическую модель в виде дифференциального уравнения. Смысл практически всех методов синтеза цифровых фильтров сводится к замене операции дифференцирования некоторым дискретным ее эквивалентом.

Таким образом, при синтезе выбирается гипотетический фильтр, называемый аналоговым прототипом синтезируемого цифрового фильтра, и производится *конвертация* (преобразование данных из одного формата в другой) аналоговых фильтров в эквивалентные цифровые.

Рассмотрим некоторые методы синтеза цифровых фильтров. Тремя наиболее распространенными методами конвертации аналоговых фильтров в эквивалентные цифровые являются метод инвариантного преобразования импульсной

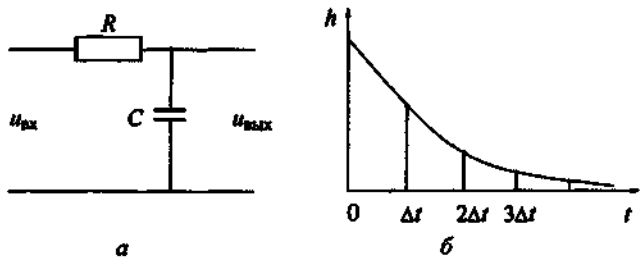



**Рис. 8.30.** Интегрирующая цепь:
*а* — схема; *б* — аналоговая и дискретная импульсные характеристики

характеристики, метод инвариантного преобразования частотной характеристики (метод билинейного $z$-преобразования) и метод дискретизации дифференциального уравнения аналоговой цепи-прототипа.

**Метод инвариантного преобразования (инвариантности) импульсной характеристики.** Этот метод синтеза основан на инвариантности (напомним, что инвариантность — подобие, или неизменность) импульсной характеристики аналогового фильтра-прототипа и полученной из нее путем дискретизации импульсной характеристики цифрового фильтра (рис. 8.30). Импульсная характеристика дискретного фильтра $\{h_k\}$ идентична характеристике аналогового фильтра $h(t)$ в дискретные моменты времени $t = n\Delta t$, $n = 0, 1, ...$ (т. е. в моменты взятия выборок сигнала). Именно поэтому метод называется *инвариантным преобразованием импульсной характеристики* (*impulse invariance*).

Учитывая, что синтезируется физически реализуемая цепь, для которой импульсная характеристика равна нулю при $t < 0$ [см. (4.20)], запишем выражение для импульсной характеристики цифрового фильтра

$$\{h_k\} = h(0), h(\Delta t), h(2\Delta t), ... = h_0, h_1, h_2, ... \qquad (8.74)$$

Анализируемый метод синтеза цифрового фильтра осуществляется путем применения прямого $z$-преобразования к импульсной характеристике $\{h_k\}$ вида (8.74) и вычисления системной функции $\mathbf{H}(z)$. Затем системную функцию фильтра $\mathbf{H}(z)$ сравнивают с общими выражениями (8.47) или (8.49) и определяют коэффициенты $a_m$ и $b_n$ соответственно нерекурсивной и рекурсивной частей алгоритма фильтрации.

Рассмотрим пример синтеза, основанный на дискретизации импульсной характеристики известного аналогового фильтра. Положим, что необходимо синтезировать нерекурсивной цифровой фильтр, подобный интегрирующей $RC$-цепи и имеющий заданную импульсную характеристику (см. рис. 8.30) (для упрощения принято, что в импульсной характеристике множитель $1/\tau_з = 1/(RC) = 1$):

$$h(t) = \begin{cases} 0, & t < 0; \\ e^{-\Delta t/\tau_з}, & t \geq 0. \end{cases} \qquad (8.75)$$

Чтобы синтезировать нерекурсивный фильтр, ограничим число отсчетов его импульсной характеристики до трех:

$$\{h_k\} = 1,\ e^{-\Delta t/\tau_a},\ e^{-2\Delta t/\tau_a}. \tag{8.76}$$

Системную функцию цифрового нерекурсивного фильтра определим, применив $z$-преобразование (8.44) к последовательности (8.76). Тогда

$$\mathbf{H}(z) = 1 + e^{-\Delta t/\tau_a} z^{-1} + e^{-2\Delta t/\tau_a} z^{-2}. \tag{8.77}$$

Сравнив выражения (8.47) и (8.77), находим, что системная функция определяет структуру цифрового нерекурсивного фильтра 2-го порядка.

Заменив в (8.74) $z = e^{j\omega\Delta t}$, запишем частотный коэффициент передачи

$$\mathbf{K}_u(\omega) = 1 + e^{-\Delta t/\tau_a} e^{-j\omega\Delta t} + e^{-2\Delta t/\tau_a} e^{-j2\omega\Delta t}. \tag{8.78}$$

Синтезируем рекурсивный цифровой фильтр для случая, когда импульсная характеристика (8.75) представляется бесконечной дискретной последовательностью отсчетов

$$\{h_k\} = 1,\ e^{-\Delta t/\tau_a},\ e^{-2\Delta t/\tau_a},\ e^{-3\Delta t/\tau_a},\ \ldots \tag{8.79}$$

Выполнив $z$-преобразование данной импульсной характеристики, найдем системную функцию фильтра

$$\mathbf{H}(z) = 1 + e^{-\Delta t/\tau_a} z^{-1} + e^{-2\Delta t/\tau_a} z^{-2} + \ldots = \sum_{k=0}^{\infty} e^{-k\Delta t/\tau_a} z^{-k}.$$

По аналогии с (8.36) эту формулу запишем в следующем виде

$$\mathbf{H}(z) = \frac{z}{z - e^{-\Delta t/\tau_a}} = \frac{1}{1 - e^{-\Delta t/\tau_a} z^{-1}}. \tag{8.80}$$

Таким образом, функция (8.80) определяет рекурсивный фильтр 1-го порядка, содержащий масштабный блок, элемент задержки и сумматор. Из этого соотношения путем подстановки $z = e^{j\omega\Delta t}$ находим частотный коэффициент передачи

$$\mathbf{K}_u(\omega) = \frac{1}{1 - e^{-\Delta t/\tau_a} e^{-j\omega\Delta t}}. \tag{8.81}$$

**Метод инвариантного преобразования (инвариантности) частотной характеристики.** В основах теории цифровых систем доказывается принципиальная невозможность создания физически реализуемого цифрового фильтра, частотный коэффициент передачи которого в точности повторяет частотный коэффициент передачи аналогового фильтра. Это связано с тем, что частотный коэффициент передачи цифрового фильтра представляет собой периодическую функцию частоты с периодом ее повторения, определяемым шагом дискретизации по Котельникову $\Delta t$.

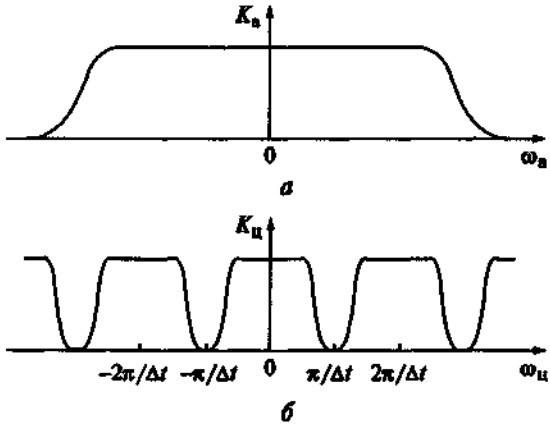



**Рис. 8.31.** Амплитудно-частотные характеристики фильтров:
*а* — аналогового; *б* — цифрового

Для сравнения на рис. 8.31 показаны АЧХ аналогового фильтра-прототипа и цифрового фильтров, откуда следует, что они фактически имеют идентичный вид. Однако характеристика цифрового фильтра, сохраняя масштаб по оси ординат, сжимается по оси абсцисс в определенное число раз. Поэтому при синтезе цифровых фильтров по методу инвариантности частотных характеристик используют условие, что весь интервал частот $-\omega_а \ldots \omega_а$, относящихся к аналоговому фильтру, преобразуется в полосу частот $-\omega_ц \ldots \omega_ц$ цифрового фильтра, удовлетворяющих неравенству

$$-\pi/\Delta t < \omega_ц < \pi/\Delta t. \tag{8.82}$$

Для осуществления синтеза по методу инвариантности частотных характеристик наиболее часто используют метод *билинейного* (*дробно-рационального*) преобразования. Такое преобразование представляет *конформное отображение* точек $p$-плоскости в точки $z$-плоскости. Положим, что $\mathbf{K}_а(p)$ — частотный коэффициент передачи аналогового фильтра, заданный по степеням комплексной частоты $p = j\omega$. Теперь рассмотрим билинейное преобразование

$$p = \frac{2}{\Delta t}\frac{z-1}{z+1}, \tag{8.83}$$

устанавливающее однозначное соответствие между точками единичной окружности в $z$-плоскости со всей мнимой осью $p$-плоскости. Заменив в выражении (8.83) переменную $z = \exp(j\omega_ц\Delta t)$, получим

$$j\omega_а = \frac{2}{\Delta t}\frac{e^{j\omega_ц\Delta t}-1}{e^{j\omega_ц\Delta t}+1}. \tag{8.84}$$

Воспользовавшись известной формулой Эйлера и проделав в (8.84) несложные преобразования, находим связь между соответствующими частотами аналогового фильтра-прототипа и цифровым фильтром

$$\omega_a = \frac{2}{\Delta t}\operatorname{tg}\frac{\omega_ц \Delta t}{2}. \qquad (8.85)$$

Согласно этому выражению, при выборе частоты дискретизации сравнительно большого значения ($\omega_ц \Delta t << 1$) имеем практически равенство частот: $\omega_a \approx \omega_ц$. Значит, в низкочастотной области частотные коэффициенты передачи аналогового и цифрового фильтров практически идентичны по структуре.

Итак, синтез цифрового фильтра по методу инвариантности частотных характеристик заключается в замене в известной функции $\mathbf{K}_a(p)$ переменной $p$ билинейным преобразованием по формуле (8.83). Полученная таким образом дробно-рациональная системная функция позволяет непосредственно определить структуру фильтра и записать алгоритмы цифровой фильтрации.

**Пример 8.24.** Аналитическая запись частотного коэффициента передачи аналогового фильтра-прототипа (рис. 8.31, *а*) на частоте среза $\omega_{са}$ имеет вид:

$$\mathbf{K}_a(p) = \frac{\omega_{са}^2}{p^2 + \sqrt{2}p\omega_{са} + \omega_{са}^2}. \qquad (8.86)$$

По методу инвариантного преобразования частотных характеристик цифровых фильтров осуществить синтез фильтра, частота среза которого $\omega_{сц} = 1{,}5 \cdot 103\ \text{с}^{-1}$ (рис. 8.31, *б*). Частота дискретизации сигнала $\omega_в = 104\ \text{с}^{-1}$.

Р е ш е н и е. По теореме Котельникова вычислим интервал дискретизации: $\Delta t = 2\pi\omega_в = 6{,}28 \cdot 10^{-4}$ с. Затем по формуле (8.85) определим частоту среза аналогового фильтра-прототипа

$$\omega_{са} = \frac{2}{\Delta t}\operatorname{tg}\frac{\omega_{са}\Delta t}{2} = 1620\,\text{с}^{-1}.$$

Выполнив с помощью соотношения (8.83) замену переменной в выражении (8.85), находим системную функцию синтезируемого цифрового фильтра

$$\mathbf{H}(z) = \frac{\omega_{са}^2}{\left(\frac{2}{\Delta t}\frac{z-1}{z+1}\right)^2 + \sqrt{2}\frac{2}{\Delta t}\frac{z-1}{z+1}\omega_{са} + \omega_{са}^2} =$$

$$= \frac{\omega_{са}^2 (z+1)^2}{\left[\left(\frac{2}{\Delta t}\right)^2 + \frac{2}{\Delta t}\sqrt{2}\omega_{са} + \omega_{са}^2\right]z^2 + 2\left[\omega_{са}^2 - \left(\frac{2}{\Delta t}\right)^2\right]z + \left[\left(\frac{2}{\Delta t}\right)^2 - \frac{2}{\Delta t}\sqrt{2}\omega_{са} + \omega_{са}^2\right]}$$

Подставляя в это выражение соответствующие числовые значения, получим формулу для системной функции цифрового фильтра

$$\mathbf{H}(z) \approx \frac{z^2 + 2z + 1}{7{,}6z^2 - 5{,}7z + 2} \approx \frac{0{,}13 + 0{,}26z^{-1} + 0{,}13z^{-2}}{1 - 0{,}75z^{-1} + 0{,}26z^{-2}}. \qquad (8.87)$$

Такой системной функции отвечает структура рекурсивного цифрового фильтра второго порядка, у которого постоянные весовые коэффициенты нерекурсивной части $a_0 = 0{,}13$, $a_1 = 0{,}26$, $a_2 = 0{,}13$, а рекурсивной — $b_1 = 0{,}75$, $b_2 = -0{,}26$.

**Синтез цифровых фильтров методом дискретизации дифференциального уравнения аналоговой цепи.** Цифровой фильтр, приближенно соответствующий аналоговой цепи, синтезируют, дискретизировав дифференциальное уравнение, описывающее аналоговый прототип. Рассмотрим синтез цифрового фильтра, прототипа колебательной системы 2-го порядка (резонансный контур), для которой связь между выходным колебанием $y(t)$ и входным колебанием $u(t)$ устанавливается известным дифференциальным уравнением

$$\frac{d^2y(t)}{dt^2} + 2\alpha\frac{dy(t)}{dt} + \omega_p^2 y(t) = u(t). \tag{8.88}$$

Установим шаг дискретизации входного сигнала равным $\Delta t$ и рассмотрим совокупности дискретных отсчетов выходного $\{y_k\}$ и входного $\{u_k\}$ сигналов. Если в (8.88) заменить производные их конечно-разностными выражениями, то дифференциальное уравнение превратится в разностное уравнение вида:

$$\frac{y_k - 2y_{k-1} + y_{k-2}}{\Delta t^2} + 2\alpha\frac{y_k - y_{k-1}}{\Delta t} + \omega_p^2 y_k = u_k. \tag{8.89}$$

После перегруппировки слагаемых в формуле (8.89) запишем

$$y_k = \frac{\Delta t^2 u_k + 2(1+\alpha\Delta t)y_{k-1} - y_{k-2}}{1 - 2\alpha\Delta t + \omega_p^2\Delta t^2}. \tag{8.90}$$

Полученное уравнение можно записать в виде

$$y_k = a_0 u_k + b_1 y_{k-1} + b_2 y_{k-2}, \tag{8.91}$$

где коэффициенты:

$$a_0 = \frac{\Delta t^2}{1 - 2\alpha\Delta t + \omega_p^2\Delta t^2}; \quad b_1 = \frac{2(1+\alpha\Delta t)}{1 - 2\alpha\Delta t + \omega_p^2\Delta t^2}; \quad b_2 = \frac{-1}{1 - 2\alpha\Delta t + \omega_p^2\Delta t^2}. \tag{8.92}$$

Разностное уравнение (8.90) определяет алгоритм цифрового рекурсивного фильтра 2-го порядка, прототипом которого является аналоговая колебательная система. Такой фильтр принято называть *цифровым резонатором*.

Системная функция цифрового резонатора

$$\mathbf{H}(z) = \frac{a_0}{1 - b_1 z^{-1} - b_2 z^{-2}}, \tag{8.93}$$

а амплитудно-частотная характеристика

$$\mathbf{K}_\text{ц}(\omega) = \frac{a_0}{1 - b_1 e^{-j\omega\Delta t} - b_2 e^{-j2\omega\Delta t}}. \tag{8.94}$$

При соответствующем выборе коэффициентов цифровой резонатор может выполнять роль частотно-избирательного фильтра, подобно колебательному контуру. Отметим, что при использовании метода инвариантного преобразования импульсной характеристики после оцифровки аналогового сигнала импульсная характеристика исходного аналогового фильтра сохраняется, а амплитудно-частотная характеристика — нет. Вследствие этого данный метод плохо подходит для разработки фильтров верхних частот или режекторных фильтров.

Метод инвариантного преобразования импульсной характеристики хорош для моделирования аналоговых систем с характеристиками фильтров нижних частот, но для частотно-избирательных БИХ-фильтров билинейный метод является наилучшим. Метод билинейного $z$-преобразования также дает возможность синтезировать весьма эффективные фильтры и прекрасно подходит для расчета коэффициентов частотно-избирательных (полосовых) фильтров. Он позволяет разрабатывать цифровые фильтры с такими широко известными классическими характеристиками, как Баттерворта, Чебышева и эллиптические. Цифровые фильтры, полученные с помощью метода билинейного $z$-преобразования, будут, в общем случае, сохранять специфические особенности амплитудной характеристики аналогового фильтра (например, граничные частоты, неравномерность в полосе пропускания и затухание в полосе подавления), но не свойства, связанные с временно́й областью.

Метод синтеза цифровых фильтров, основанный на дискретизации дифференциального уравнения, описывающего исходную аналоговую цепь, нагляден физически, однако эффективен не во всех задачах. При более сложных цепях синтеза данный метод становится громоздким.

Существует еще ряд методов синтеза цифровых фильтров, которые здесь не рассматриваются. Выбор того или иного метода синтеза определяется конкретной задачей, требуемыми параметрами и т. д.

## 8.6. Аналого-цифровые и цифроаналоговые преобразователи

При построении радиоэлектронных систем, связывающих потребителя информации с устройствами цифровой обработки, требуется преобразовать сигналы из аналоговой формы в цифровую и из цифровой в аналоговую.

Как уже отмечалось, такие операции с сигналами осуществляют аналого-цифровые и цифроаналоговые преобразователи. АЦП и ЦАП характеризуются погрешностью, быстродействием и динамическим диапазоном. Основные погрешности систем цифровой обработки сигналов связаны, главным образом, с квантованием сигнала в АЦП на конечное число уровней, определяемое разрядностью кода. При квантовании дискретного сигнала его отсчеты представляются в виде двоичных чисел, имеющих ограниченное число разрядов. Вследствие этого отсчеты могут принимать лишь конечное множество значений и, следовательно, при представлении сигнала неизбежно происходит его округление.

Напомним, что в процессе преобразования отсчетов сигнала в числа производится его *квантованием по уровню*, а возникающие при этом ошибки округления — *ошибками* (или *шумами*) *квантования*.

Увеличение числа разрядов повышает точность приема и позволяет расширить динамический диапазон передаваемых сигналов. Потерянная из-за недостатка разрядов АЦП информация невосстановима, и не существует простых критериев, подобных теореме Котельникова, показывающих, когда подобное восстановление могло бы иметь место. Существуют лишь оценки погрешности, например, через мощность шума, порожденного ошибкой в последнем разряде, однако не всегда они удовлетворительны. К примеру, слух человека логарифмичен (закон Вебера–Фехнера), и маломощная помеха в диапазоне, где мощность сигнала невысока, слишком хорошо различима. Для того чтобы оценить влияние помехи, используют отношение сигнал/шум.

Один из основных путей снижения шумов квантования — применение многоразрядных кодов. Обычно применяют 8-, 10-, 12-, 16-, 20- и 24-разрядные АЦП. Каждый дополнительный разряд улучшает отношение сигнал-шум на 6 дБ (вернее сказать, способен улучшить это отношение при оптимальной реализации всех прочих компонент системы сбора сигналов). Однако это автоматически приводит к уменьшению быстродействия цифровых фильтров вследствие увеличения времени обработки сигналов. При проектировании цифровых устройств специалисты стремятся снизить шум квантования до уровня, обеспечивающего необходимую точность восстановления непрерывного сигнала. Отсюда исходит и требование к необходимому количеству разрядов в кодах. Если уровни квантования (их число равно $N$) пронумеровать двоичными числами, то, согласно формулы (8.2), количество разрядов кодов

$$n \geq \log_2 N. \tag{8.95}$$

При этом относительная погрешность квантования

$$\delta = \Delta U / U_{\max}. \tag{8.96}$$

**Пример 8.25.** Определить число разрядов двоичного кода для измерения изменяющегося от $U_{\min} = 0$ В до $U_{\max} = 100$ В напряжения $U$ с относительной погрешностью $\delta$, не превышающей 0,5 В.

**Решение.** Шаг квантования $\Delta U = 2\delta = 2 \cdot 0{,}5 = 1$ В. Количество уровней квантования $N = U_{\max}/\delta = 100$. По формуле (8.95) находим разрядность двоичного кода $n > \log_2 N = 7$.

Процесс равномерного квантования непрерывных сигналов показан на рис. 8.32. Если на вход АЦП подано линейно нарастающее напряжение $u_1(t) = u_1$ (штриховая линия), то выходное напряжение $u_2(t) = u_2$ будет представлять собой линию ступенчатой формы (рис. 8.32, *а*). Эту линию называют *характеристикой квантования*. Разность напряжений $q(t) = q = u_2(t) - u_1(t)$ представляет собой шум (ошибку или погрешность) квантования.

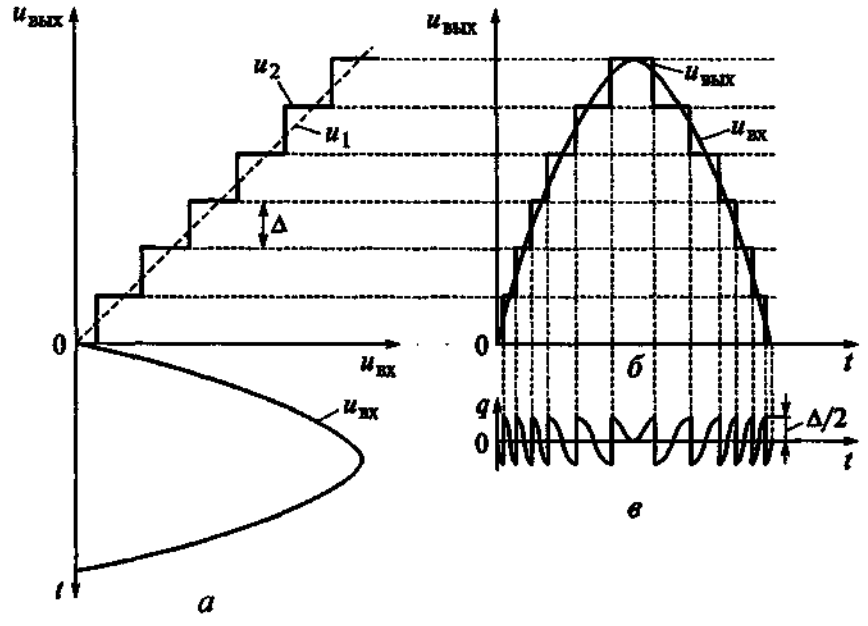



**Рис. 8.32.** Квантование гармонического сигнала:
*а* — характеристика квантования и сигнал на входе;
*б* — сигнал на выходе; *в* — шум квантования

Очевидно, что максимальное значение погрешности квантования не зависит от напряжения $u_1(t)$ и всегда равно половине шага квантования $\Delta/2$.

Пусть на вход АЦП поступает полупериод гармонического колебания $u_{вх}(t)$ (рис. 8.32, *а*). Выходной сигнал $u_{вых}(t) = u_{вых}$ приобретает ступенчатую форму (ступенчатая линия на рис. 8.32, *б*), отличающуюся от входной синусоиды $u_{вх}(t)$. Функция погрешности квантования (рис. 8.32, *в*) аналитически запишется в виде

$$q(t) = u_{вых}(t) - u_{вх}(t). \tag{8.97}$$

Отметим, что возникающая при этом абсолютная погрешность $q(t)$ представления является неустранимой, но контролируемой, поскольку не превышает половины шага квантования (рис. 8.32, *в*). Выбрав малый шаг квантования, можно обеспечить требуемую величину ошибки квантования. Возможные изменения амплитуды и частоты входного гармонического сигнала приводят лишь к изменению частоты следования зубцов функции $q(t)$, а их форма остается практически треугольной. Среднюю мощность (дисперсию) шума квантования легко вычислить из геометрических построений, приведенных на рис. 8.32, *в*:

$$\sigma^2 = \frac{1}{3}\left(\frac{\Delta}{2}\right)^2 = \frac{\Delta^2}{12}. \tag{8.98}$$

Из этой формулы следует, что шумы квантования снижаются с увеличением разрядности цифрового кода, т. е. с уменьшением шага квантования $\Delta$.

### Цифроаналоговые преобразователи

Проанализируем один из вариантов построения схем ЦАП, основанного на методе *суммирования фиксированных напряжений* нескольких источников, каждое из которых соответствует определенному разряду двоичного числа.

Принцип действия четырехразрядного ЦАП поясняет одна из простейших схем на операционном усилителе, представленная на рис. 8.33, *а*. Основу схемы составляет матрица резисторов с источником постоянного напряжения, соединенных с инвертирующим входом ОУ ключами, управляемыми двоичным кодом (например, выходным кодом счетчика). В зависимости от поступающего кода цифрового сигнала подключаются резисторы с различными номиналами сопротивлений. В схеме ключи замыкаются только при поступлении на них команд, соответствующих логической единице. Коэффициенты усиления инвертирующего усилителя по входам $2^0$, $2^1$, $2^2$ и $2^3$ соответственно равны:

$$K_0 = -R_0Q_0/R;\ K_1 = -2R_0Q_1/R;\ K_2 = -4R_0Q_2/R;\ K_3 = -8R_0Q_3/R. \qquad (8.99)$$

Здесь $Q_0$, $Q_1$, $Q_2$, $Q_3$ — кодовые числа, принимающие два значения: либо 1 (ключ замкнут), либо 0 (ключ разомкнут).

Напряжение на выходе ЦАП определяется выражением

$$u_{вых}(t) = u_{вых} = -E(K_0 + K_1 + K_2 + K_3) = \Delta(Q_0 + 2Q_1 + 4Q_2 + 8Q_3), \qquad (8.100)$$

где $\Delta = -ER_0/R$ — амплитудное значение соответствует младшему разряду двоичного числа, т. е. уровню квантования.

Из формулы (8.100) следует, что четырехразрядный двоичный код преобразуется в выходное напряжение, изменяющееся от 0 до 15Δ. Например, двоичному числу 1001 соответствует напряжение $u_{вых1} = \Delta(8\cdot1 + 4\cdot0 + 2\cdot0 + 1\cdot1) = 9\Delta$, а числу 1100 — $u_{вых2} = 12\Delta$. Поскольку на вход резистивной матрицы подается постоянное напряжение $E$, то выходное напряжение ЦАП изменяется скачками при переключении кода цифрового сигнала. Сглаживание сигнала на выходе осуществляется с помощью ФНЧ. Недостатки рассмотренной схемы — жесткие требования к высокой стабильности сопротивлений резисторов матрицы.

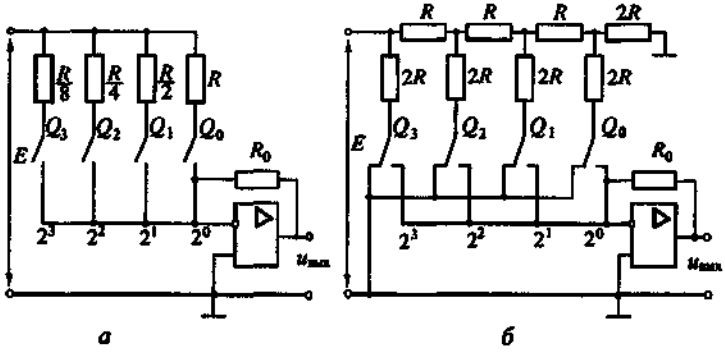



**Рис. 8.33.** Схемы четырехразрядных ЦАП:
*а* — простейшая; *б* — с резистивной матрицей *R*–2*R*

Во многом лишена указанных недостатков схема ЦАП с *резистивной матрицей R-2R* (рис. 8.33, *б*), содержащая резисторы с сопротивлениями всего двух номиналов. Резисторы соединены в матрицу таким образом, что ее входное сопротивление со стороны источника питания $E$ всегда равно $R$ и не зависит от положения ключей (кодовых чисел цифрового сигнала). При поступлении кодовых нулей ключи подсоединяют резисторы с сопротивлением $2R$ к нулевой точке, а при поступлении кодовых единиц — к входу операционного усилителя. При этом токи, протекающие через резисторы с сопротивлением $2R$, не изменяются. Для $k$-разрядного ЦАП выходное напряжение записывается в общем виде:

$$u_{вых} = (\Delta/2^k)(Q_0 + 2Q_1 + \cdots + 2kQ_k). \qquad (8.101)$$

Современные ЦАП в основном изготавливают в виде отдельных интегральных микросхем, позволяющих преобразовывать цифровые коды с 16-ю и более разрядов. При этом увеличение разрядности кодов не приводит к резкому уменьшению быстродействия.

### Аналого-цифровые преобразователи

По своей структуре аналого-цифровые преобразователи более сложны, чем ЦАП, причем последние часто являются основным узлом АЦП. В настоящее время существуют три принципиально различных метода построения схем АЦП: последовательный, параллельный и последовательно-параллельный.

*Последовательный метод* построения АЦП является наиболее распространенным в современной цифровой технике. Он основан на подсчете числа суммирований опорного напряжения младшего разряда, необходимого для получения напряжения, равного входному. При этом $k$-разрядный двоичный код одного отсчета определяется в схеме за $2^k$ тактовых интервалов дискретизации.

Проанализируем упрощенную структурную схему четырехразрядного АЦП последовательного счета (рис. 8.34). Начало преобразования входного сигнала определяется временем поступления импульса запуска, который через $RS$-триггер $T$ подключает счетчик Ст2 к выходу генератора тактовых (счетных) импульсов $M$. Схема ЦАП D/A, куда поступает цифровой код со счетчика, формирует выходное напряжение $u_{вых}$, которое сравнивается в компараторе $K$ с входным напряжением $u_{вх}$. При сравнении напряжений компаратор через логический элемент И (&) выдает сигнал прекращения подачи тактовых импульсов на счетчик Ст2.

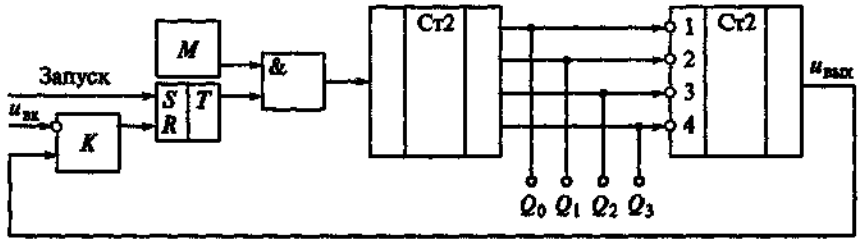



**Рис. 8.34.** Упрощенная структурная схема четырехразрядного АЦП последовательного счета

В результате осуществляется считывание с счетчика выходного четырехразрядного кода, представляющего в момент окончания преобразования цифровой эквивалент выходного напряжения преобразователя. В АЦП значения выходного цифрового кода в процессе преобразования многократно изменяются, поэтому этот преобразователь обладает сравнительно низким быстродействием. Максимальное время преобразования

$$t_{\text{пр max}} = (2^k - 1)\Delta t. \qquad (8.102)$$

Так как число разрядов АЦП задается, то время преобразования определяется минимальным периодом повторения тактовых импульсов, который зависит от длительности установления переходных процессов. Минимальное время преобразования аналогового сигнала

$$\Delta t_{\min} = t_{\text{сч}} + t_{\text{ЦАП}} + t_{\text{т}} + t_{\text{к}} + t_{\text{л}}, \qquad (8.103)$$

где $t_{\text{сч}}$ — длительность переходного процесса в счетчике; $t_{\text{ЦАП}}$ — время установления переходных процессов ЦАП при преобразовании; $t_{\text{т}}$, $t_{\text{к}}$, $t_{\text{л}}$ — длительность переключения соответственно триггера, компаратора и логического элемента.

**Пример 8.26.** Определить частоту следования счетных импульсов $f_{\text{в}} = 1/\Delta t$ и рассчитать максимальное время преобразования $t_{\text{пр max}}$ для шестиразрядного АЦП последовательного типа, в котором: $t_{\text{сч}} = 100$ нс; $t_{\text{ЦАП}} = 300$ нс; $t_{\text{т}} = t_{\text{л}} = 25$ нс; $t_{\text{к}} = 50$ нс.

Р е ш е н и е. Воспользовавшись формулой (8.103), находим, что $\Delta t_{\min} = 500$ нс; откуда частота следования счетных импульсов $f_{\text{в}} = 2$ МГц. В соответствии с выражением (8.102) максимальное время преобразования достигает 130 мкс, что позволяет данному АЦП осуществлять порядка 7 500 преобразований в 1 с.

Работа *параллельных k*-разрядных АЦП основана на использовании $2^k$–1 компараторов (рис. 8.35).

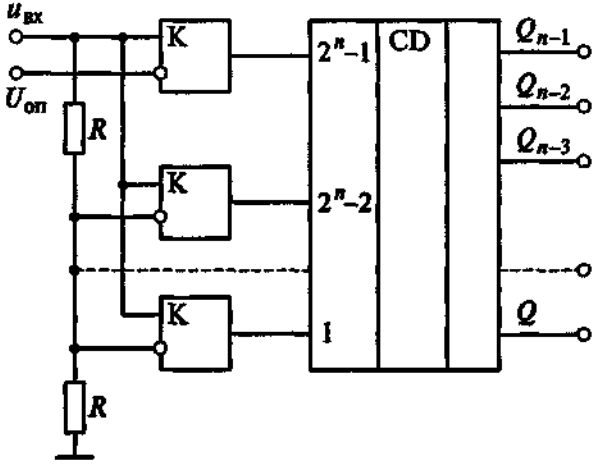



**Рис. 8.35.** Структурная схема параллельного АЦП

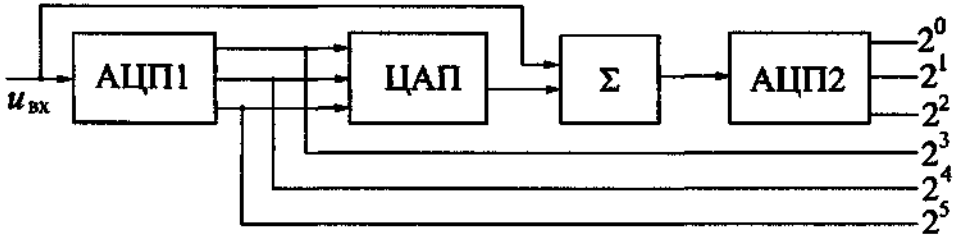



**Рис. 8.36.** Структурная схема последовательно-параллельного АЦП

Неинвертирующие входы ОУ компараторов объединены, и на них подается непрерывный сигнал, а к каждому инвертирующему входу подключено индивидуальное опорное напряжение, снимаемое с резистивного делителя. Разность между опорным напряжением двух соседних компараторов равна шагу квантования $\Delta = U_{оп}/2^k$.

Компараторы, у которых входное напряжение превысит соответствующее опорное напряжение, вырабатывают логическую 1, а остальные — логический 0. Информация с выходов компараторов поступает на шифратор CD, который преобразует ее в двоичный код.

Параллельные схемы обладают наибольшим быстродействием среди других типов АЦП. Однако для повышения точности и уменьшения мощности шумов квантования необходимо увеличение в схемах числа компараторов.

В *последовательно-параллельных схемах* АЦП используется сочетание методов последовательного и параллельного преобразований, что существенно увеличивает быстродействие последовательных преобразователей и уменьшает объем параллельных.

На рис. 8.36 показана структурная схема шестиразрядного аналого-цифрового преобразователя данного типа, в которой используются два трехразрядных параллельных АЦП, один трехразрядный ЦАП и сумматор Σ.

Аналого-цифровой преобразователь формирует из входного напряжения три старших разряда выходного кода, соответствующие значениям $2^3$, $2^4$ и $2^5$. Эти разряды поступают на вход трехразрядного ЦАП, в котором они вновь преобразуются в аналоговое напряжение, отличающееся от входного напряжения $u_{вх}$ на погрешность преобразования схемы АЦП1. Аналоговое напряжение с выхода ЦАП подается на сумматор Σ, где оно вычитается из входного напряжения $u_{вх}$. Полученное разностное напряжение подается на АЦП2, где оно преобразуется в три младших цифровых разряда $2^2$, $2^1$, $2^0$ выходного кода.

## Контрольные вопросы

1. Как осуществляется дискретизация непрерывного сигнала?
2. В чем состоит принцип квантования сигнала по уровню?
3. Сколько разрядов содержит цифровой код при квантовании сигнала на 256 уровней?
4. Сформулируйте теорему Котельникова.
5. В чем состоит сходство и отличие спектров дискретного и исходного непрерывного сигналов?
6. Как можно восстановить исходный непрерывный сигнал из дискретного?
7. На чем основан алгоритм дискретного прямого преобразования Фурье?
8. Запишите дискретные прямое и обратное преобразования Фурье.
9. Сколько гармонических составляющих можно найти с помощью ДПФ, если известно число отсчетов?
10. Назовите основные свойства ДПФ.
11. На каком принципе основан алгоритм БПФ?
12. Как строится сигнальный граф (бабочка) БПФ?
13. Как осуществляется $z$-преобразование дискретных сигналов?
14. В чем состоит основное свойство дискретной свертки?
15. Изобразите структурную схему цифровой обработки сигналов.
16. В чем состоит отличие аналоговых и цифровых фильтров?
17. Какими преимуществами обладает цифровой фильтр?
18. Что такое импульсная характеристика цифрового фильтра?
19. Назовите основные свойства импульсных характеристик цифровых фильтров.
20. Что определяет понятие системной функции цифрового фильтра?
21. Как связаны системная функция, импульсная характеристика и частотный коэффициент передачи цифрового фильтра?
22. Представьте алгоритм работы нерекурсивного фильтра.
23. Представьте алгоритмы работы рекурсивного фильтра.
24. Какие фильтры называют КИХ- и БИХ-фильтрами?
25. В чем основное преимущество цифровых фильтров, реализованных по каноническим схемам?
26. Почему нерекурсивные фильтры заведомо устойчивы?
27. Какие критерии устойчивости цифровых фильтров используются на практике?
28. Что собой представляют частотные характеристики цифровых фильтров?
29. Назовите основные методы синтеза цифровых фильтров.
30. От чего зависит мощность шума квантования?
31. На каком принципе основано действие ЦАП?
32. Представьте схемы ЦАП с резистивными матрицами.
33. На чем основан принцип действия АЦП?

# Глава 9. УСТРОЙСТВА ПЕРЕДАЧИ И ПОМЕХОУСТОЙЧИВОГО ПРИЕМА СООБЩЕНИЙ

Радиопередающее устройство (короче, радиопередатчик) является важнейшим элементом радиотехнических систем (и в частности систем связи) различного назначения. В большинстве случаев масса, габаритные размеры, энергопотребление, надежность и срок службы систем в первую очередь зависят от характеристик и параметров радиопередающих устройств.

Радиоприемные устройства (РПУ) — широкий класс радиотехнических средств, решающих задачу качественного приема и воспроизведения сообщения, переданного радиопередающим устройством. Радиосвязь без радиоприемного устройства невозможна. Одной из основных задач радиотехники является борьба с шумами и помехами. Повысить помехоустойчивость систем передачи информации можно разными методами. Один из методов связан с созданием таких устройств обработки, которые наилучшим образом выделяют сигнал из шумов и помех. Другой заключается в совершенствовании структуры передаваемых сигналов, применении помехоустойчивых способов их кодирования и модуляции.

## 9.1. Радиопередающие устройства

Радиопередающим называют устройство, служащее для генерирования, модуляции высокочастотных и сверхвысокочастотных колебаний и усиления мощности, подводимых к антенне и излучаемых в пространство (или передачи их по линиям связи к абоненту). Генерацию, модуляцию и усиление — объединяет общее понятие «формирование сигнала», под которым понимают колебание (радиосигнал), несущее информацию. Радиопередатчик часто применяют вместе с радиоприемником и питающим устройством; весь этот комплекс называют *радиостанцией*. Самостоятельно радиопередатчики используются в областях, где не нужен прием информации в месте ее передачи — сигналы точного времени, разнообразные навигационные радиомаяки для определения местоположения объектов, многопозиционная радиолокация, радиовещание и т. д.

**Общие сведения.** Структурно радиопередающие устройства состоят из собственно радиопередатчика (передатчика) и передающей антенны. Традиционно передающие (как и приемные) антенны изучают отдельно, в курсе антен-

но-фидерных устройств. Анализ антенных систем достаточно сложен, и поэтому в книге они рассматриваются лишь функционально. Антенна является необходимым элементом любого передатчика и приемника. Антенна передатчика (передающая антенна) предназначена для преобразования тока высокой частоты в энергию излучаемых электромагнитных волн. Антенна приемника (приемная антенна) предназначена для преобразования принятых ею электромагнитных волн в энергию тока высокой частоты. Как отмечалось в главе 1, характер процессов, происходящих в передающей и приемной антеннах, определяет их обратимость, т. е. одну и ту же антенну можно использовать и для передачи и приема. Однако при передаче антенна может эффективно излучать электромагнитные колебания только тогда, когда ее размеры соизмеримы с длиной волны.

Направленность излучения можно обеспечить, если антенное устройство по размерам существенно превышает длину волны. Направленность имеет большое значение в радиолокации, радионавигации. Большая мощность колебаний требуется на длинных волнах вследствие поглощения землей и в других диапазонах, например при сверхдальней космической связи. Освоение новых диапазонов требует новых технических средств, поэтому переход в коротковолновую область происходил постепенно по мере создания генерирующих устройств.

В технологическом плане радиопередающие устройства представляют собой устройства из микросхем, транзисторов, диодов, электровакуумных приборов, конденсаторов, трансформаторов и множества иных элементов, соединенных между собой согласно определенной электрической схеме. Наиболее совершенные конструкции полностью состоят из полупроводниковых интегральных и гибридных микросхем и транзисторных усилительных модулей.

**Передатчики** классифицируют по назначению, диапазону рабочих волн (частот), излучаемой мощности, виду модуляции сигналов, виду излучения и условиям эксплуатации.

Назначение передатчика определяется радиосистемой, в которой он используется, и что связано с видом передаваемой информации. По назначению передатчики бывают связными, вещательными, телевизионными, радиолокационными, телеметрическими, навигационными и т. д.

*Радиовещание* осуществляется в России в диапазонах километровых, гектометровых, декаметровых, метровых и дециметровых волн. В первых трех диапазонах традиционно используют амплитудную модуляцию с шагом сетки рабочих частот 10 кГц, а на двух последних — широкополосную частотную модуляцию с шагом сетки рабочих частот 250 кГц. Наиболее распространено вещание на метровых волнах в диапазонах 65,8...74,0 МГц (4,56...4,05 м) и 87,5...108,0 МГц (3,43...2,78 м) с применением методов частотной модуляции.

*Телевизионное вещание* ведется в России в диапазонах метровых, дециметровых и сантиметровых волн, причем в звуковом канале используют частотную модуляцию, а в канале изображения — амплитудную модуляцию с одной боковой полосой частот. Для телевизионного вещания отведено пять поддиапазонов в метровом и дециметровом диапазонах: I (48,5...66 МГц), II (76...100 МГц),

III (174...230 МГц), IV (470...622 МГц), V (622...958 МГц), на которых размещено более 70 каналов. Для кабельного телевидения выделены каналы СК1–СК8 и СК11–СК18, перекрывающие диапазоны (110...174 МГц) и (230...294 МГц).

В настоящее время в системах радиосвязи интенсивно развивается направление, использующее распространение радиоволн в СВЧ-диапазоне. Свойство этих волн пронизывать ионосферу используется в спутниковых системах телевидения и для связи с автоматическими станциями и космическими кораблями. Для всех спутниковых систем радиосвязи Международным комитетом по регистрации частот (МКРЧ) выделены следующие полосы частот в диапазонах, ГГц: *L* (1,452...1,500 и 1,610...1,710); *S* (1,930...2,700); *C* (3,400...5,250 и 5,725...7,075); *X* (7,250...7,750 и 7,900...8,400); *Ки* (10,700...12,750 и 12,750...14,800), *Ка* (14,400...26,500 и 27,000...50,200), *K* (84,000...86,000).

*По диапазону рабочих волн* современные передатчики делятся в соответствии с классификационной таблицей диапазонов радиоволн и частот (см. 1.1).

В соответствии с *классификацией волн различают передатчики* километровых, гектометровых, декаметровых, метровых, дециметровых и других волн: 10...1 км (30...300 кГц); 1000...100 м (300...3000 кГц); 100...10 м (3...30 МГц); 10...1 м (30...300 МГц); 1...0,1 м (300...3000 МГц) и др. С этим отличием связаны соответствующие особенности конструкций, так как в разных диапазонах различны колебательные контура и усилительные элементы. Далее будут рассмотрены в основном вещательные и связные передающие устройства.

*По средней излучаемой мощности* передаваемых радиосигналов различают передатчики очень малой (менее 3 Вт), малой (3 ...10 Вт), средней (10 ...500 Вт), большой (0,5 ...10 кВт) и сверхбольшой (более 10 кВт) мощности. Обычно мощность опередатчика определяют как максимальную мощность ВЧ-колебаний, поступающую в антенну, при отсутствии модуляции и непрерывном излучении.

*По виду модуляции сигнала* передатчики (и приемники) делятся на устройства с амплитудной, балансной и однополосной, частотной, фазовой, импульсной, квадратурной, импульсно-кодовой и другими видами модуляции.

*По виду излучения* различают передатчики, работающие в непрерывном и импульсном режимах. В первом случае при передаче сообщения сигнал излучается непрерывно, во втором — в виде радиоимпульсов.

*По условиям эксплуатации* бывают стационарные, бортовые (космические, корабельные, самолетные, автомобильные) и переносные (портативные) передатчики.

К основным параметрам передатчиков относятся коэффициент полезного действия (КПД), диапазон частот, шаг сетки рабочих частот, выделенная полоса частот излучения, нестабильность частоты несущего колебания, побочные и внеполосные излучения, коэффициент нелинейных искажений сигнала, электромагнитная совместимость и т. д.

Коэффициент полезного действия передатчика определяется формулой

$$\eta = P_A/P_0, \qquad (9.1)$$

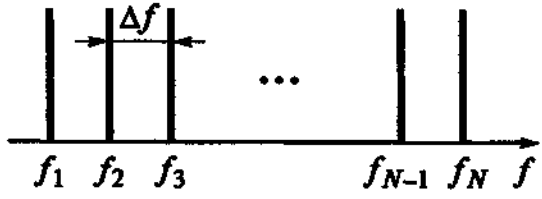



**Рис. 9.1.** Сетка рабочих частот передатчика

где $P_A$ — средняя мощность колебаний, отдаваемая в антенну; $P_0$ — мощность, потребляемая устройством от всех источников питания. КПД современных радиопередатчиков достигает 30...40 %, причем он растет с увеличением излучаемой мощности.

Диапазон частот несущих колебаний $f_1 \ldots f_N$, где $N$ — число частот внутри этого диапазона. В самом простом случае радиопередатчик может быть одночастотным и тогда $N = 1$.

Шаг сетки рабочих частот $\Delta f$ в заданном диапазоне определяется согласно выражению:

$$\Delta f = (f_N - f_1)/(N-1), \qquad N \geq 2. \tag{9.2}$$

Радиопередатчик может работать на любой из фиксированных частот внутри диапазона $f_1 - f_N$ (рис.9.1). Например, передатчик системы самолетной радиосвязи метрового диапазона работает в диапазоне частот 118...136 мГц при шаге $\Delta f = 25$ кГц, общее число частот $N = 721$.

*Выделенная полоса частот излучения* $\Delta f_{выд}$. При любом виде модуляции — амплитудной, частотной, фазовой или импульсной — спектр сигнала становится или линейчатым (рис. 9.2, *а*), или сплошным (рис. 9.2, *б*), занимая определенную полосу частот, от верхней $f_в$ до нижней $f_н$, т. е. $\Delta f_c = f_в - f_н$.

Для данного спектра выделяется определенная полоса частот $\Delta f_{выд}$. При этом следует соблюдать неравенство $\Delta f_c \leq \Delta f_{выд}$, т. е. спектр сигнала должен укладываться в выделенную для него полосу. В противном случае излучения одного передатчика могут мешать излучениям других передатчиков, проникая в выделенные для них полосы.

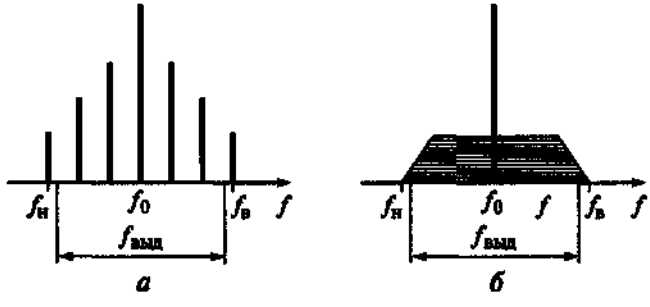



**Рис. 9.2.** Выделенная полоса частот излучения с видами спектров:
*а* — линейчатым; *б* — сплошным

**Нестабильность частоты несущих колебаний.** Напомним (см. гл. 6), что различают абсолютную и относительную нестабильность частоты колебаний, а также долговременную и кратковременную. Абсолютная нестабильность представляет собой разность $\Delta f$ между текущим $f$ и номинальным $f_р$ (или $f_н$) значениями частоты колебаний.

Например, номинальное значение частоты должно быть $f_н = f_0 = 125$ мГц, а фактически радиопередатчик формирует сигнал с частотой $f = 149{,}9995$ мГц. Следовательно, абсолютная нестабильность частоты составит

$$\Delta f = f_н - f = 125 \text{ МГц} - 124{,}995 \text{ МГц} = 0{,}005 \text{ МГц} = 5 \text{ кГц}.$$

Относительную нестабильность частоты определяют *коэффициентом нестабильности*, который представляет собой отношение абсолютной нестабильности частоты к ее номинальному значению

$$\Delta_f = \Delta f / f_н.$$

И тогда относительная нестабильность будет составлять

$$\Delta_f = 0{,}005/125 = 0{,}00004 = 40 \cdot 10^{-6} = 0{,}004\ \%.$$

В современных передатчиках относительная нестабильность частоты обычно не превышает $(2–3) \cdot 10^{-6}$. По существующим международным нормам отклонение от номинала частоты передатчика для радиосвязи на гектометровых волнах не должна превышать 0,005, для радиовещательных передатчиков отклонения частоты в этом диапазоне не должно превышать 10 Гц. На декаметровых волнах допустимая нестабильность частоты для передатчиков мощностью более 0,5 кВт равна $15 \cdot 10^{-6}$, что соответствует в диапазоне 3…30 МГц абсолютному отклонению частоты от 45 до 450 Гц. Некоторые системы радиосвязи по своему принципу требуют, чтобы стабильность частоты была значительно лучше, чем требуют указанные нормы. Например, системах радионавигации к этому параметру предъявляются еще более жесткие требования: в них следует иметь $\Delta_f \le 10^{-9}$. В настоящее время нестабильность частоты центральных государственных передатчиков близка к достигнутой нестабильности государственных эталонов ($10^{-8} – 10^{-9}$ — водородных, цезиевых).

*Побочные излучения передатчика.* В идеальном случае радиопередатчик любой системы передачи информации должен излучать только полезный сигнал на частоте несущей и его спектр должен укладываться в выделенную полосу частот (рис. 9.3, *а*). Однако нелинейный характер процессов в трактах передатчика приводит к появлению в полосе спектра полезного сигнала побочных (паразитных, в том числе и интермодуляционных, искажений) составляющих $f_п$ (рис. 9.3, *б*). Побочные излучения, лежащие за пределами, но вблизи полезной полосы частот, называются *внеполосными*.

Помимо внеполосных, передатчик может излучать гармоники номинальной частоты $f_н = f_0$ — сигналы с частотами $2f_0$, $3f_0$ и т. д., а также субгармоники (*subharmonic*) — сигналы с частотами ниже номинальной: $f_0/n$ ($n = 1, 2, \ldots$).

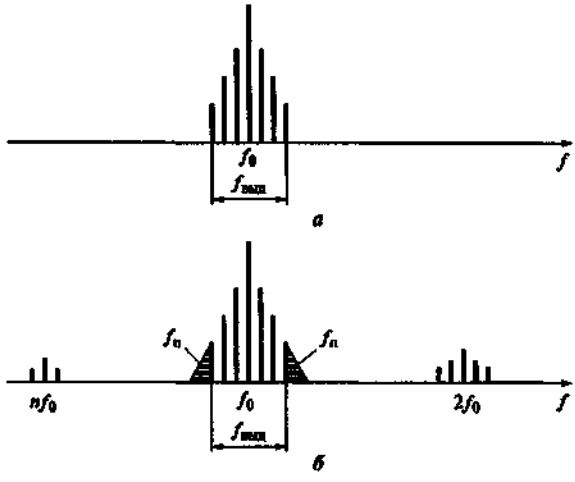



**Рис. 9.3.** Излучения передатчика:
*а* — без побочных составляющих; *б* — наличие побочных составляющих

Кроме того, возможно излучение так называемых *паразитных колебаний*, причиной возникновения которых является самовозбуждение в усилительных каскадах радиопередатчика. Возникновение таких колебаний, и тем более излучение, должно быть практически исключено.

Поскольку полностью исключить побочные излучения нельзя, особенно в мощных передатчиках, то устанавливают норму на их значение или в абсолютных единицах (т. е. указывается, что мощность такого-то побочного излучения не должна быть больше $P$, мВт), или в относительных единицах к мощности полезного излучения. Обычно уровень внеполосных излучений должен превышать –60 дБ (т. е. мощность сигнала должна превышать мощность побочных излучений в $10^6$ раз). На некоторых частотах норма может составлять –100 дБ и более.

Электромагнитная совместимость. В мире работают огромное количество радиопередатчиков, создающих вокруг Земли электромагнитное поле. Радиосигналы приходят на Землю и из космоса: на околоземных орбитах находятся тысячи спутников и на каждом из них установлено по несколько передатчиков. Только систем дальней спутнико-космической радиосвязи в мире более 150. При одновременной работе множества радиосистем помехи радиоприему неизбежны. Интенсивность помех определяется числом действующих излучателей, их мощностью, расположением в пространстве, формой диаграммы направленности антенн, условиями распространения радиоволн и т. д. Поэтому среди многих проблем передающих устройств особо выделяется задача обеспечения электромагнитной совместимости радиосистем.

Способность радиотехнических систем одновременно функционировать в реальных условиях эксплуатации с требуемым качеством при воздействии на них непреднамеренных электромагнитных помех и не создавать недопустимых таких же помех другим радиосистемам называют *электромагнитной совмести-*

*мостью* (ЭМС; *electromagnetic compatibility — EMC*). При этом следует рассматривать электромагнитную совместимость радиотехнической аппаратуры, работающей в пределах одного объекта, и на обширной территории.

Передатчику каждой радиотехнической системы отводится строго определенная полоса частот, в которой допускается радиоизлучение. Однако фактически любой радиопередатчик помимо полезного сигнала излучает и побочные колебания, которые по отношению к другой радиотехнической системе являются помехами. Рассмотрим диаграммы на рис. 9.3, *б*. Пусть номинальная частота передатчика одной радиосистемы равна $f_0$. Но помимо нее антенна излучает и радиосигнал на частоте $2f_0$, пусть и малой мощности. На эту частоту может быть настроен приемник соседней радиотехнической системы. По отношению к ней сигнал с частотой $2f_0$ будет являться помехой, причем весьма сильной, если передатчик первой системы расположен близко от приемника второй системы.

Параметры передаваемого сообщения. Таким сообщением может быть речевая, факсимильная, телевизионная, телеметрическая и другая разнообразная информация, в том числе и считываемая с компьютера. Сообщение может передаваться в форме аналогового или цифрового сигнала. При аналоговом сообщении основным характеризующим его параметром является полоса частот спектра сигнала, при цифровом — число бит в секунду.

Наличие в передатчике нелинейных и параметрических цепей вызывает появление *нелинейных искажений* передаваемых радиосигналов. В частности, в процессе модуляции несущее колебание претерпевает определенные изменения, т. е. искажается. Возникающие при этом высшие гармоники сигнала излучаются передающей антенной в пространство. Побочные излучения в виде высших гармоник попадают в частотный диапазон других радиосистем и создают им помехи в работе. Например, при передаче сообщения в виде гармонического сигнала таким параметром служит коэффициент нелинейных искажений, определяющий появление в исходном сигнале 2-й, 3-й и т.д. гармоник. При передаче импульсных сигналов искажения характеризуют по изменению его формы. Кроме нелинейных искажений, в передатчике возникают и *линейные* (частотные) искажения, связанные с прохождением радиосигналов через частотно-избирательные цепи с неидеальными АЧХ и нестрого линейными ФЧХ. Линейные искажения влияют на качество передачи только в своем канале.

Нормы на надежность и долговечность, массу и габаритные размеры радиопередатчика устанавливаются в соответствии с общими нормами для радиотехнической аппаратуры. В радиопередатчиках повышенной мощности устанавливаются специальные нормы, связанные с *техникой безопасности*.

Конструкции, габаритные размеры и масса передатчиков в основном определяются средней излучаемой мощностью. *Стабильность и устойчивость* работы передатчика оценивают по его способности сохранять свои электрические характеристики в допустимых пределах при воздействии окружающей среды (температуры, влажности, атмосферного давления, механических нагрузок, климатических воздействий) и изменении параметров источника питания.

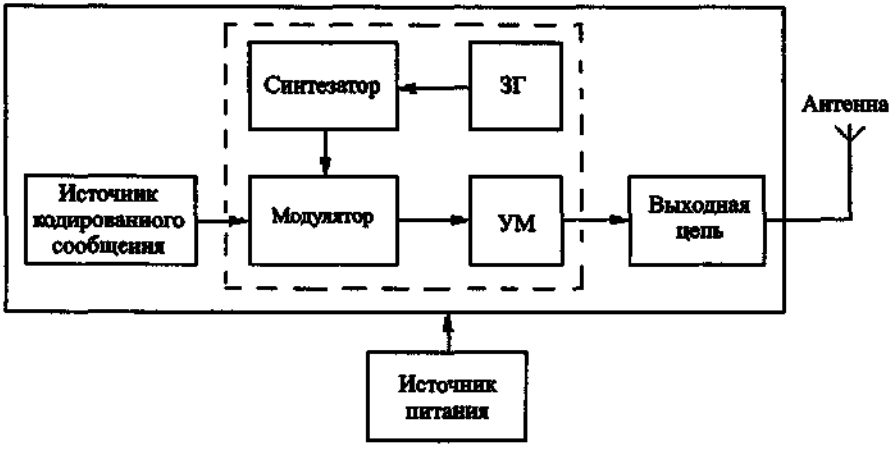



**Рис. 9.4.** Обобщенная структурная схема современного передатчика

Структурная схема современного передатчика (рис. 9.4) состоит:
- из источника кодированного сообщения, которое требуется передать;
- задающего генератора частоты (ЗГ), создающего высокостабильное гармоническое сравнительно низкой частоты колебание;
- синтезатора сетки несущих частот;
- модулятора;
- усилителя мощности (УМ), который увеличивает мощность промодулированной несущей за счет внешнего источника энергии;
- выходной (согласующей) цепи и антенны.

### Передатчик с амплитудной модуляцией

Простейшая схема передатчика с амплитудной модуляцией несущего колебания (рис. 9.5) содержит возбудитель, каскады умножения частоты (УЧ), усиления мощности (УМ), усилитель низкой частоты (УНЧ), на который подается передаваемый сигнал $u_{вх}$ и амплитудный модулятор (АМ).

*Возбудитель* представляет собой маломощный задающий автогенератор, стабилизированный кварцем. Малая мощность задающего генератора позволяет использовать при его разработке более высокочастотные полупроводниковые приборы, обладающие меньшей инерционностью, обеспечивает облегченный

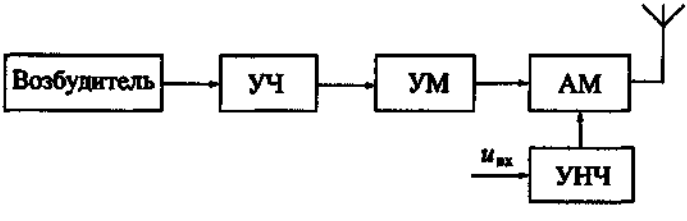



**Рис. 9.5.** Структурная схема передатчика с амплитудной модуляцией

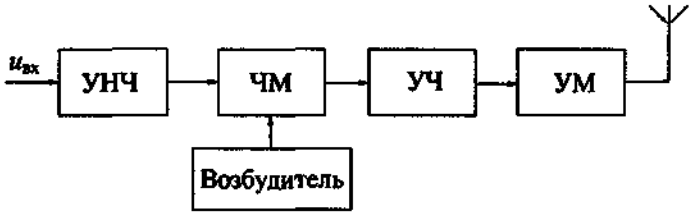



**Рис. 9.6.** Структурная схема передатчика с частотной модуляцией

тепловой режим работы усилительного прибора и кварцевого резонатора, что повышает стабильность частоты. Кварцевые генераторы работают на сравнительно невысоких (до сотен мегагерц на гармониках кварца) частотах. Поэтому после задающего генератора включают каскады *умножителей частоты*, повышающие частоту колебаний до значения несущей. Часто в умножителях частоты осуществляют и увеличение мощности колебаний. Для создания требуемой мощности на выходе передатчика применяют *усилители мощности*. Как правило, усилители мощности включены между каскадами умножителей частоты, и тракт называют *усилительно-умножительной цепочкой*. Выходной усилитель мощности передатчика нагружен на волновод (кабель и т. п.), соединенный с антенной. Амплитудную модуляцию осуществляют в выходном усилителе мощности. Часто такой усилитель является оконечным каскадом радиопередатчика.

### Передатчик с частотной модуляцией

В диапазонах метровых и дециметровых волн в вещательных и связных передатчиках применяют частотную модуляцию. В простейших, одно-, двухканальных передатчиках (рис. 9.6) частотную модуляцию осуществляют путем воздействия передаваемым сигналом $u_{вх}$ (он предварительно усиливается в УНЧ) на частоту колебаний ЧМ-генератора (ЧМ). При этом неизменная частота несущего колебания $f_0$ формируется возбудителем. Частотно-модулированные колебания с ЧМ-генератора поступают на усилительно-умножительную цепочку (УЧ–УМ), где частота и мощность колебаний многократно увеличиваются. Высокую стабильность несущей частоты передатчика поддерживают специальными методами, в частности с помощью автоматической подстройки частоты.

Кварцевые автогенераторы обеспечивают формирование высокостабильных сигналов только в узкой полосе частот или на нескольких частотах. Поэтому в многочастотных широкодиапазонных ЧМ-передатчиках в качестве возбудителя включают *синтезаторы частот*. При необходимости иметь большой набор генерируемых частот (*сетку* частот) используют кварцевые синтезаторы частот, а частотная модуляция сигнала выполняется в модуляторе ЧМ-сигналов.

### Синтезаторы частоты

В современных многоканальных радиопередатчиках (это относится и к радиоприемникам) требование высокой стабильности несущей частоты и ее быстрой перестройки трудно совместимы.

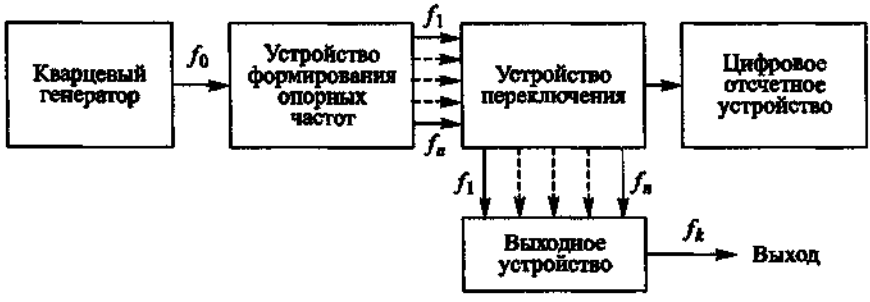



**Рис. 9.7.** Обобщенная структурная схема синтезатора частот

Поэтому при разработке синтезаторов частоты переходят к дискретному перекрытию частотного диапазона, при котором допускается генерирование сигналов на любой из множества частот, следующих друг за другом с фиксированным интервалом, называемым *шагом дискретной сетки*.

Кварцевые синтезаторы частоты — многочастотные генераторы гармонических колебаний с дискретной перестройкой частоты. Синтезаторы обеспечивают синусоидальную форму колебаний, высокую спектральную «чистоту», большую точность установки и возможность программной перестройки частоты. Они позволяют получать напряжения фиксированных частот с дискретностью до сотых долей герц. По точности установки и стабильности частоты синтезаторы превосходят обычные генераторы с плавной перестройкой частоты.

Обобщенная структурная схема синтезатора частот содержит кварцевый генератор частоты $f_0$, устройство формирования опорных частот $f_1,\ldots,f_n$, устройство переключения, подключающее сигнал нужной частоты, цифровое отсчетное и выходное устройства (рис. 9.7).

**Аналоговые синтезаторы частот**. На рис. 9.8 представлена структурная схема простейшего аналогового синтезатора частоты с цифровым управлением. Синтезатор содержит опорный кварцевый генератор (ОКГ), управляемый делитель частоты (УДЧ), управляемый генератор (УГ), фазовый детектор (ФД) с системой фазовой автоматической подстройки частоты и программируемое цифровое устройство (ПЦУ). На фазовый детектор подается два колебания: первое со стабильной частотой $f_{оп}$ — от опорного кварцевого генератора; второе с частотой $f/N \approx f_{оп}$ через управляемый делитель частоты с коэффициентом деления $N$ — от управляемого генератора.

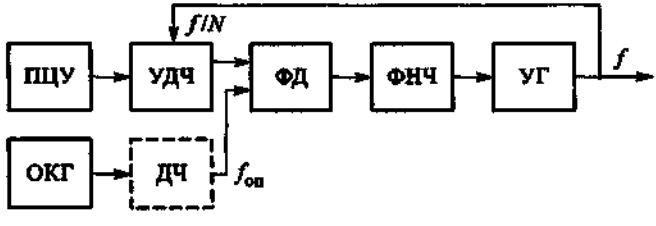



**Рис. 9.8.** Структурная схема аналогового синтезатора частот

Напряжение с выхода фазового детектора через фильтр нижних частот (ФНЧ) воздействует на управляемый генератор и подстраивает его до обеспечения равенства частот $f/N$ и $f_{оп}$. Изменяя с помощью ПЦУ коэффициент деления $N$, можно получить требуемую сетку частот с шагом, равным $f_{оп}$. Поскольку выходная частота синтезатора связана с частотой опорного кварцевого генератора как $f = Nf_{оп}$, то относительные нестабильности этих частот равны. Если в таком синтезаторе требуется сформировать очень низкую частоту, то между опорным кварцевым генератором и фазовым детектором необходимо дополнительно ввести делитель частоты (ДЧ).

Представленный простейший вариант построения синтезатора частоты имеет серьезные недостатки. Первый из них связан с конечностью ширины полосы синхронизации управляемого генератора, которая зависит от управляющих элементов генератора и коэффициентов передачи ФД и ФНЧ. Поэтому для получения широкой сетки синтезируемых частот приходится изменять собственную частоту $f$ управляемого генератора. Второй недостаток обусловлен узкими возможностями УДЧ, построенного, как правило, на основе счетчика импульсов. Введением обратной связи в делителе частоты можно изменять его коэффициент деления, который будет принимать любые целочисленные значения, допустимые разрядностью счетчика.

Цифровые синтезаторы. Прогресс в области микропроцессорной техники обусловил появление измерительных генераторов с новыми принципами формирования сигналов. Достоинством цифровых методов синтеза является малое время установления частоты колебаний при перестройке, что важно для функционирования быстродействующих автоматизированных систем, а также отсутствие разрыва фазы при смене частот. В синтезаторах частоты часто требуется использовать дробные значения коэффициента деления частоты. Метод дробного преобразования частоты используется в новейших разработках цифровых синтезаторов, реализуемых по базовой схеме (рис. 9.9).

В синтезаторе частоты коэффициент деления программно-управляемого делителя частоты (ПУДЧ) изменяется во времени, образуя последовательность временны́х циклов определенной длительности. Полученный цикл делится еще на несколько подциклов, в течение каждого из которых коэффициент деления остается постоянным. Изменение коэффициента деления производят в момент перехода от одного подцикла к другому так, чтобы средний за цикл коэффициент деления был равен заданному. В схеме цифрового синтезатора частоты используются цифровой фазовый детектор (ЦФД), ЦАП и микропроцессор (МП).

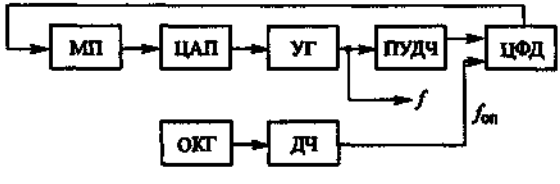



**Рис. 9.9.** Структурная схема цифрового синтезатора частоты

Подстройку выходной частоты производят в конце каждого цикла. Для этого используют управляемый генератор, напряжение подстройки частоты на который подается с ЦАП. Сигнал управления (ошибки рассогласования) вырабатывается цифровым фазовым детектором и его уровень соответствует значению средней за время цикла разности фаз колебаний, получаемых от опорного кварцевого генератора и управляемого генератора. Затем сигнал управления с фазового детектора подается на микропроцессор, который через ЦАП по заданному коду требуемой частоты осуществляет программное управление схемой ПУДЧ.

В целом современные синтезаторы частот строят на основе одной большой микросхемы, в которую объединяют практически все звенья схемы на рис. 9.9.

### Тенденции развития радиопередающих устройств

Большинство каскадов современных передатчиков выполняют только на цифровых и аналоговых микросхемах. Электронные (мощные усилительные лампы) и дискретные полупроводниковые приборы (в основном полевые транзисторы) используют лишь в выходных каскадах усилителей передатчиков большой и сверхбольшой мощности.

Если передатчик работает на одной фиксированной частоте, задающий генератор содержит соединенные последовательно маломощный высокостабильный кварцевый автогенератор и нескольких усилительных каскадов. Когда число рабочих частот передатчика не более десяти, в тракте возбудителя используют несколько кварцевых автогенераторов или один автогенератор с переключающимися кварцевыми резонаторами. В настоящее время в качестве задающих генераторов возбудителя в основном применяют цифровые синтезаторы частот. Высокостабильные задающие генераторы на основе синтезаторов частот могут работать в диапазоне 100...200 МГц. Изготовление же передатчиков с кварцами на более высокие частоты встречает серьезные технологические проблемы.

Применение умножителей частоты в трактах радиопередатчиков позволяет и в диапазоне СВЧ на частотах 1...100 ГГц получать колебания, стабильность которых теоретически определяется кварцевым резонатором задающего генератора. В передатчиках низкочастотного диапазона обычно используются транзисторные умножители частоты, однако с увеличением несущей частоты и переходом в область СВЧ параметры транзисторных умножителей частоты и усилителей мощности заметно ухудшаются. Поэтому в выходных каскадах передатчиков выгодно применять варакторные умножители частоты. Наиболее важными показателями умножителей частоты, применяемых в радиопередающих устройствах, являются коэффициент умножения, выходная колебательная мощность, коэффициент гармоник и КПД.

Требуемые уровни выходной мощности передатчиков современных систем передачи информации в некоторых случаях на три-пять порядков превышает максимальную мощность, генерируемую электронными приборами. Этот разрыв между мощностью радиопередатчика и мощностью единичного генератора стал особенно ощутим при переходе к полупроводниковым приборам.

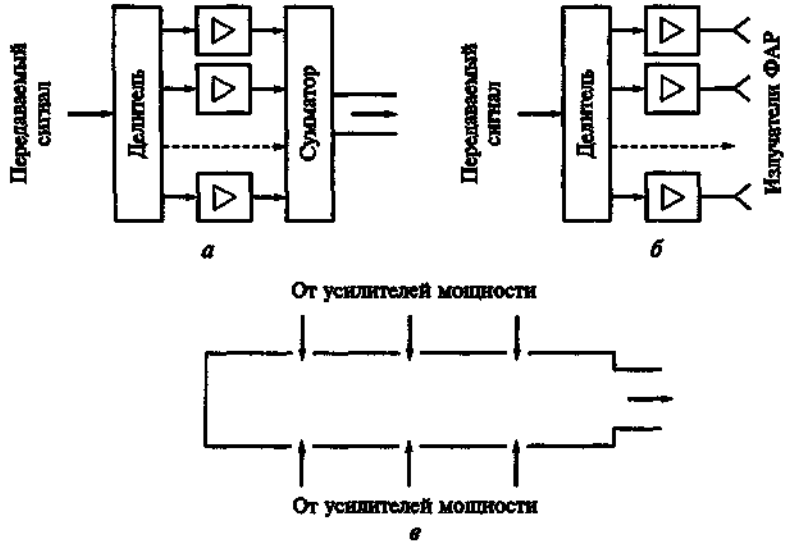



**Рис. 9.10.** Методы суммирования мощностей в передатчиках радиосистем:
*а* — с помощью многополюсных сумматоров;
*б* — путем сложения сигналов с помощью ФАР; *в* — в общем резонаторе

Вместе с тем необходимая мощность СВЧ-передатчиков в непрерывном режиме работы достигает нескольких киловатт. Но даже и в СВЧ-передатчиках мощностью в десятки ватт мощность полупроводникового прибора во многих случаях оказывается меньше в несколько раз. В этих случаях требуемые уровни выходной мощности передатчиков достигают методами *сложения мощностей* нескольких идентичных полупроводниковых узлов выходных каскадов.

Известны три основных способа суммирования мощностей однотипных генераторов (рис. 9.10, *а*):

- с помощью многополюсных сумматоров;
- путем сложения сигналов в пространстве с помощью ФАР;
- в общем резонаторе.

При первом способе к специальному многополюсному сумматору параллельно подключают большое число однотипных синхронизированных через делитель мощности входного сигнала генераторов (или усилителей мощности), мощность сигналов которых поступает в общий выходной сигнал, связанный с нагрузкой (рис. 9.10, *а*). Однако при этом ухудшается устойчивость усилительных каскадов, а выходная мощность передатчика оказывается заметно ниже их суммарной мощности. Повышение выходной мощности передатчика в нагрузке и взаимную развязку генераторов (усилителей) обеспечивают мостовыми схемами сложения мощностей. В таких сумматорах каждый активный прибор работает самостоятельно на оптимальную для него нагрузку, а режимы работы всех каскадов не зависят друг от друга. При этом повышается надежность работы передатчика, поскольку выход из строя одного из нескольких усилительных каскадов лишь снижает мощность в передающей антенне.

При втором способе сложение мощностей колебаний происходит в пространстве с помощью ФАР, состоящей из большого числа определенным образом ориентированных элементарных излучателей, каждый из которых возбуждается от самостоятельного усилителя мощности или генератора (рис. 9.10, *б*). Все сигналы, подводимые к излучателям, идентичны, за исключением значений начальных фаз, связанных между собой определенным законом. Излучатели располагают так, чтобы электромагнитная связь между ними была слабой. Если сигналы излучателей сформированы соответствующим образом, то мощность, излучаемая ФАР, складывается в пространстве и практически равна сумме мощностей всех усилителей мощности. При этом возникает задача по стабилизации и синхронному управлению фронтом фаз ряда сигналов одинаковой структуры.

При третьем способе суммирования сигналы усилителей мощности (генераторов) подводятся к общей колебательной системе (в СВЧ-диапазоне это объемный резонатор), в котором и происходит их сложение (рис.9.10, *в*).

Первый способ сложения мощностей позволяет увеличить мощность радиопередатчика по отношению к мощности одного усилительного прибора на 15...20 дБ; второй — на 30…40 дБ; третий — на 10…15 дБ. Все способы позволяют существенно повысить надежность радиопередатчика, поскольку отказ одного из усилителей мощности приводит только к снижению суммарной выходной мощности, и устойчивость работы усилительного тракта, так как сумматоры улучшают развязку между каскадами. При суммировании мощностей улучшаются условия охлаждения полупроводниковых приборов, располагаемых на большой поверхности.

## 9.2. Радиоприемные устройства

Во всех рассмотренных классах радиосистем радиоприемное устройство является одной из основных составных частей каждой конкретной системы, а в пассивных системах — единственным радиотехническим устройством.

**Общие сведения.** Радиоприемные (приемные) устройства предназначены для приема радиосигналов и преобразования их к виду, позволяющему использовать содержащуюся в них полезную информацию. Любое радиоприемное устройство состоит из *приемной антенны* и собственно *радиоприемника* (проще, *приемника*).

*Радиоприемным устройством* в широком смысле называют систему узлов и блоков, которые осуществляют следующие операции:

- преобразование с помощью приемной антенны электромагнитного поля сигнала (помех) в радиосигнал и обеспечение пространственной и поляризационной избирательности полезного радиосигнала;
- выделение (фильтрация по частоте) полезных радиосигналов из совокупности других (мешающих) сигналов и помех, действующих на выходе приемной антенны и не совпадающих по частоте с полезным сигналом;
- усиление радиосигнала;

• преобразование и усиление принимаемых сигналов для обеспечения качественной работы детектора, декодера, схем защиты приемника от помех;
• демодуляция (детектирование) принятого сигнала для выделения информации (модулирующей функции), содержащейся в полезном радиосигнале;
• усиление демодулированного сигнала;
• декодирование принятого сигнала;
• обработка принимаемых сигналов с целью ослабления мешающего воздействия помех искусственного и естественного происхождения.

Последняя операция предусматривает введение в приемник средств помехозащиты и эффективную обработку сигналов и помех, при которой достигается наилучшее обнаружение сигналов или оценка принятой информации (сообщения) по какому-либо критерию оптимальности приемника в соответствии с целевым содержанием решаемой практической задачи.

Во всех рассмотренных радиотехнических системах полезная информация заложена в параметрах радиосигнала на входе радиоприемного устройства. Поэтому все основные операции, связанные с пространственной и частотной избирательностью радиосигнала, его усилением, демодуляцией и обработкой, могут решаться на основе единой теории анализа и синтеза трактов приемника и одинаковых принципов их схемотехнической реализации.

**Приемники** классифицируют по назначению, диапазону принимаемых волн (частот), виду модуляции передаваемых сигналов и условиям эксплуатации. Качественные показатели радиоприемных устройств определяются электрическими, конструктивно-эксплуатационными и экономическими характеристиками.

*По назначению* приемники делят на профессиональные и вещательные (бытовые). К профессиональным относятся приемники связные, радиолокационные, радионавигационные, радио-, телеуправления и др. Бытовые приемники обеспечивают прием программ звукового и телевизионного вещания.

*Диапазон рабочих частот*, т. е. область частот настройки, в пределах которой обеспечиваются все другие электрические характеристики приемника. Современные радиоприемные устройства обеспечивают уверенный прием радиосигналов в большом частотном диапазоне, где возможна работа радиосистем.

*По виду модуляции сигнала* радиоприемники, как и радиопередатчики, делятся на устройства с амплитудной, амплитудной балансной и однополосной, частотной, фазовой, импульсной, импульсно-кодовой (цифровой) и другими видами модуляции.

*По условиям эксплуатации* приемники бывают стационарными, бортовыми (космическими, корабельными, самолетными, автомобильными) и переносными (портативными) или мобильными.

К основным характеристикам приемника относятся чувствительность, избирательность, помехоустойчивость и динамический диапазон.

Чувствительность приемника — его способность обеспечивать прием очень слабых полезных сигналов. Ее оценивают мощностью входного радиосигнала, необходимой для получения номинальной (требуемой) мощности

на выходе приемника при заданном отношении сигнал/шум. На практике антенны принимают сигналы мощностью $10^{-10}...10^{-15}$ Вт (50...300 мкВ; около $10^{-6}$ В) или напряженностью поля 0,3...5 мВ/м. На выходе приемника для надежной регистрации сигнала требуется мощность порядка единиц ватт, т. е. необходимо усиление по мощности до $10^{10}...10^{15}$, по напряжению — до $10^{7}$. Это достигают с помощью многокаскадных УВЧ, УПЧ и УНЧ.

Какова же может быть предельная чувствительность современного приемника? Теоретически можно разработать приемник, чувствительный к сколь угодно слабым сигналам. Для этого требуется лишь увеличить число его усилительных каскадов. Однако в реальных устройствах, наряду с полезным (информационным) сигналом, всегда имеются шумы, образующиеся за счет внешних помех и собственных шумов приемника. Уровень собственных шумов в основном определяется шумами первых каскадов приемника. Поэтому при создании высокочувствительных приемников одной из важнейших задач является разработка малошумящих активных и пассивных элементов и усилительных каскадов. В ряде систем передачи информации для повышения чувствительности приемника применяют охлаждение его входных устройств. Установленное определение чувствительности справедливо тогда, когда внешними помехами приему можно пренебречь. Однако в реальных условиях эксплуатации пренебрежение внешними помехами приему часто недопустимо, особенно для приемников специальных радиосистем. Поэтому вводят понятие эффективной чувствительности по отношению к уровню помех, как внутренних, так и внешних.

Эффективная чувствительность приемника — способность принимать слабые сигналы с заданным качеством (отношением сигнал/шум) и вероятностью приема в условиях воздействия всего ансамбля помех.

Избирательность, или селективность, приемника — способность выделить полезный сигнал из множества других сигналов и помех, принятых антенной. Это понятие избирательности определяется только частотной фильтрацией полезного сигнала от мешающих сигналов в высокочастотном тракте. Реальная же избирательность приемника в целом зависит также от нелинейных явлений в его каскадах. Поэтому используют *эффективную частотную избирательность*, под которой понимают способность приемника различать полезный сигнал (на частоты которого он настроен) и помехи (с частотами за пределами полосы пропускания), уровни которых таковы, что они создают нелинейные эффекты при одновременном действии полезного и мешающих сигналов. Нелинейные эффекты в усилительных и преобразовательных каскадах приемника обусловлены в основном нелинейностью ВАХ усилительных приборов при больших уровнях сигнала или помех. Не анализируя эти эффекты подробно, отметим, что они вызывают следующие явления в приемниках:

- сжатие амплитуды сигнала, т. е. нарушение линейной зависимости между амплитудами сигнала на выходе и входе каскада — это явление происходит в режиме большого полезного сигнала и обусловлено уменьшением средней крутизны усилительных приборов за период под его воздействием;

• блокирование полезного сигнала, выражающееся в изменении коэффициента передачи приемника при действии помех, частоты которых отличаются от частот основного и побочного каналов; часто коэффициент передачи приемника уменьшается, что приводит к снижению чувствительности;

• появление перекрестных искажений радиосигналов, отражающиеся в изменениях структуры спектра выходного сигнала при одновременном действии на входе приемника сигнала и модулированной помехи, частота которой не совпадает с основным и побочными каналами приема;

• интермодуляцию между передаваемым сигналом и внеполосными сигналами (сигналами других передатчиков), а также между всеми этими сигналами и шумом. Интермодуляция заключается в том, что при воздействии на нелинейный элемент в приемнике двух или более внеполосных сигналов и помех на выходе элемента возникает сложный спектр интермодуляционных колебаний, которые практически не возможно отфильтровать.

При проектировании приемников задают значения избирательности по соседнему, зеркальному и другим побочным каналам приема.

*Временну́ю избирательность* используют при приеме импульсных радиосигналов (обычно в приемниках РЛС), когда достаточно точно известен момент их появления. При этом применяют метод временно́го стробирования, при котором приемник открывается только на время ожидаемого прихода импульсного сигнала. Остальное время приемник закрыт, что уменьшает действие помех.

*Пространственную избирательность* осуществляют с помощью остронаправленных приемных антенн, а в настоящее время и путем управления диаграммой направленности фазированных антенных решеток. Если передатчик сигнала и источник помехи разнесены по угловым направлениям, то можно существенно ослабить уровень внешней помехи на входе приемника. Для этого формируют максимум диаграммы направленности приемной антенны в направлении на передатчик сигнала, а в направлении источника помехи — нули (провалы) в диаграмме направленности.

*Поляризационная избирательность* осуществима, если имеются различия в поляризациях электромагнитных волн полезного сигнала и помехи. Она производится приемной антенной, которую настраивают на вид поляризации сигнала.

Пространственную и поляризационную избирательности в основном характеризуют приемные (и передающие) антенны, которые характеризуют их способность различать направление прихода и поляризацию электромагнитных волн.

К о э ф ф и ц и е н т  ш у м а — общепринятая оценка шумовых свойств приемников (и усилителей) при условии, что на входе и внутри них шум *белый.*

Коэффициентом шума называют отношение мощностей сигнала и шума на входе $(P_с/P_ш)_{вх}$, отнесенное к такому же отношению мощностей на выходе линейной части (на входе детектора) приемника $(P_с/P_ш)_{вых}$:

$$K_ш = \frac{(P_с/P_ш)_{вх}}{(P_с/P_ш)_{вых}}. \qquad (9.3)$$

Физически коэффициент шума показывает, во сколько раз уменьшается отношение мощности сигнала к мощности шума (помехи) при включении в тракт передачи сигнала данного приемника. Можно предложить и такую трактовку: коэффициент шума показывает, во сколько раз шумы на выходе линейной части приемника увеличиваются за счет шумов, возникающих в самом приемнике.

Если (теоретически это возможно при абсолютном нуле температуры) приемник сам по себе не шумит, то его входное и выходное отношения сигнал/шум равны и коэффициент шума $K_ш = 1$ (это 0 дБ).

Помехоустойчивость приемника (обычно под этим подразумевают систему связи) — способность обеспечивать прием переданной или извлеченной информации с заданной достоверностью при выбранных видах сигналов (в том числе видов модуляции или кодирования) и наличии помех в радиоканале. Поскольку действие помех проявляется в том, что принятое сообщение отличается от переданного, то качественно помехоустойчивость при заданной помехе можно характеризовать *степенью соответствия принятого сообщения переданному*. Эту величину в теории информации называют термином «верность». Помехоустойчивость повышается всеми видами избирательности, а также созданием оптимальных (чаще, квазиоптимальных; квазиоптимальное — реальное, стремится к идеальному) структур приемников и специальными мерами борьбы с помехами при обработке принимаемых сигналов.

Известны общие методы борьбы с помехами, к которым относятся:

- увеличение мощности сигнала за счет увеличения мощности передатчика;
- устранение причин возникновения помех;
- охлаждение входных устройств приемника;
- применение направленных антенн, в том числе и ФАР;
- изменение несущей частоты сигнала;
- накопление сигналов;
- совершенствование устройств, выделяющих сигналы на фоне помех и извлекающих из них информацию.

Однако этим путем не всегда можно добиться успеха в борьбе с помехами. Принципиально неустранимы внутренние, атмосферные и космические шумы. В связи с этим возникает задача наделения передаваемого информационного сигнала свойством противостоять воздействию помех, т. е. свойством помехоустойчивости к тому или иному их виду. Для решения подобной задачи необходимо из всех возможных видов модуляции и способов кодирования выбрать такие, которые в данных условиях и при заданных ограничениях имели бы необходимую помехоустойчивость.

Такие из перечисленных показателей, как реальная чувствительность, избирательность и помехоустойчивость, в значительной мере определяют характеристику электромагнитной совместимости радиоприемного устройства. Указанные показатели характеризуют прежде всего восприимчивость радиоприемника к мешающим электромагнитным воздействиям через антенну и по цепям питания, управления и коммутации. Однако следует отметить, что содержащиеся в ра-

диоприемном устройстве генераторные, цифровые и иные устройства создают узкополосные или широкополосные электромагнитные излучения, которые могут быть помехами для других радиоэлектронных средств.

Допустимые искажения принимаемого сигнала в отсутствие помех. Искажения делят на линейные (амплитудно-частотные и фазочастотные) и нелинейные. Амплитудно-частотные искажения приводят к изменению соотношений между амплитудами составляющих полезного сигнала на выходе приемника и на его входе. Их оценивают допустимой нелинейностью АЧХ приемника в заданном диапазоне модулирующих частот. Фазочастотные искажения проявляются в том, что разные составляющие спектра сигнала при прохождении через приемник сдвигаются во времени не на одинаковое значение и оцениваются допустимой нелинейностью ФЧХ приемника. Нелинейные искажения передаваемого сигнала проявляются в виде дополнительных частот (гармоник и комбинационных) на выходе приемника, не содержащихся в передаваемом сообщении. Их оценивают допустимым коэффициентом нелинейных искажений (коэффициентом гармоник) при заданном коэффициенте модуляции. Искажения импульсных сигналов оценивают допустимыми длительностями фронта и среза, неравномерностью вершины, выбросами на вершине и в паузе.

Динамический диапазон — отношение граничных уровней входных воздействий (максимальных и минимальных), характерных для приема данного вида излучения, в пределах которых обеспечивается нормальное качество приема. Упрощенно, динамический диапазон — область уровней принимаемого сигнала, в пределах которой приемо-передающий тракт практически линеен. Максимальный уровень сигнала ограничен допустимыми нелинейными искажениями в усилительных и преобразовательных каскадах приемника из-за нелинейных характеристик электронных приборов при сильных сигналах. Минимальный уровень входного сигнала ограничивается уровнем собственных шумов, т. е. чувствительностью приемника. Динамический диапазон определяет основные параметры приемника — эффективные чувствительность и избирательность, а также такие параметры систем, в состав которых входит приемник, как точность воспроизведения сигналов, разрешающую и пропускную способности, ЭМС и т. д. Относительное изменение уровней полезных сигналов и помех на входе приемников в обычных условиях работы может составлять 90...100 дБ. С помощью специальных устройств (автоматической регулировки усиления приемника) динамический диапазон по основному каналу может быть доведен до 100...120 дБ, т. е. он может перекрывать диапазон входных воздействий.

Обобщенная структурная схема приемника, отражающая основные функции радиоприемного устройства, показана на рис. 9.11. Она содержит шесть функциональных узлов-трактов: антенну (А); усилительно-преобразовательный тракт (УТ); информационный тракт (ИТ); тракт адаптации, управления и контроля (ТАУК), гетеродинный тракт (ГТ) и опорный генератор (ОГ). Электромагнитные волны (ЭВ) принимаются антенной и в виде радиосигнала поступают в усилительно-преобразовательный тракт.

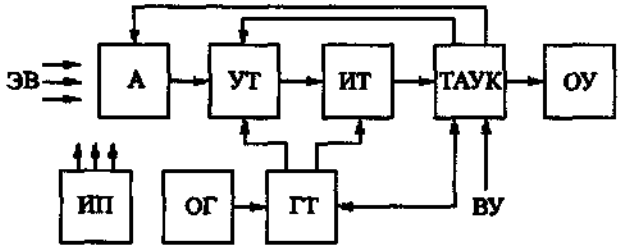



**Рис. 9.11.** Обобщенная структурная схема приемника

УТ производит выделение полезного сигнала из совокупности поступающих от антенны сигналов и помех, не совпадающих с ним по частоте, и усиление этого сигнала до уровня, необходимого для нормальной работы следующих каскадов. Хотя в УТ с сигналом производятся определенные нелинейные операции (усиление по высокой частоте, смещение спектра, ограничение амплитуды и др.), в принимаемую информацию указанный тракт существенных искажений не вносит и в этом смысле его можно считать практически линейным.

В информационном тракте осуществляется основная обработка принятого радиосигнала с целью выделения содержащейся в нем информации (демодуляция и ослабление мешающего воздействия помех). При этом задачей важнейшей приемника является выделение информации с максимальной достоверностью — так называемый *оптимальный прием*. Для этого в составе ИТ включают оптимальный фильтр, цепи последетекторной обработки, следящие системы частотной и фазовой автоматической подстройки частоты, используемые для детектирования сигнала, а также для его поиска и сопровождения по частоте, фазе и временно́й задержке.

Гетеродинный тракт преобразует частоту собственного или внешнего ОГ и формирует сетку частот, необходимую для работы преобразователей частоты в УТ, следящих систем и устройств обработки сигнала в ИТ. Обычно это синтезатор частот, обеспечивающий работу всего приемника.

С помощью ТАУК осуществляют ручное, дистанционное и автоматизированное управление режимом работы приемного устройства (включение и выключение, поиск и выбор сигнала, адаптация к изменяющимся условиям работы и т. д.). При этом качество его работы и отражается на соответствующих индикаторах. Для внешнего управления (ВУ) приемником к ТАУК подводятся соответствующие сигналы управления.

Дальнейшее преобразование сигнала зависит от конкретных особенностей применения радиоприемника. Если, например, приемник предназначен для многоканальной радиосвязи, то продетектированный и усиленный сигнал подводится к оконечному устройству (ОУ), в котором происходит разделение сигналов по отдельным каналам и, если требуется, дополнительная их обработка. В оконечном устройстве энергия выделяемого сигнала используется для получения требуемого выходного эффекта — акустического (телефон, громкоговоритель), оптического (кинескоп, дисплей), механического (печатающее устройство).

Источник питания (ИП) осуществляет питание постоянным током всех узлов приемника. В ИП может осуществляться преобразование (повышение или понижение) напряжения, выпрямление, фильтрация, стабилизация.

Данная структурная схема является наиболее общей, однако в конкретных радиоприемных устройствах отдельные связи между трактами и даже некоторые тракты могут отсутствовать или выполнять более ограниченный набор функций, при этом упрощение структурной схемы приемного устройства и ограничение некоторых функций отдельных трактов снижают полноту реализации возможностей радиоприема.

По особенностям построения схем радиоприемники можно разделить на две основные группы: приемники прямого усиления и супергетеродинные приемники. Первые делят на приемники прямого усиления без регенерации (от лат. *regeneratio* — возобновление, воспроизведение) и приемники прямого усиления с регенерацией. Иногда создают сверхрегенеративные приемники.

### Приемник прямого усиления

Структурная схема простейшего приемника прямого усиления без регенерации (рис. 9.12) включает в себя *входную цепь*, *усилитель высокой* (*радио*) *частоты* (УВЧ, УРЧ), *детектор* (Д) и *усилитель низкой* (*звуковой*) *частоты* (УНЧ, УЗЧ). Иногда в таких приемниках перед УВЧ включают *малошумящий усилитель* (МШУ). Кстати, первая схема лампового радиоприемника прямого усиления была создана в 1907 г. Луи де Форестом.

Входная цепь и УВЧ составляют высокочастотный тракт приемника и содержат системы резонансных контуров, которые служат для получения максимальной мощности сигнала из антенны, а также выделяют требуемый сигнал из множества других сигналов и помех. МШУ (используемые при необходимости) предназначены для снижения уровня собственных шумов приемника и определяют чувствительность приемников. В некоторых случаях при достаточной мощности принимаемого сигнала УВЧ может отсутствовать.

Выделенная детектором (демодулятором) из радиосигнала модулирующая функция, содержащая полезную информацию, усиливается и фильтруется от помех и других комбинационных частот в УНЧ. Его усиление определяет напряжение (мощность), которую надо подвести к оконечному устройству.

При необходимости получения большого усиления УВЧ может содержать несколько каскадов, что сопряжено со снижением устойчивости его усилительных каскадов и общей избирательности приемника, затрудняет техническую реализацию перестройки по частоте.

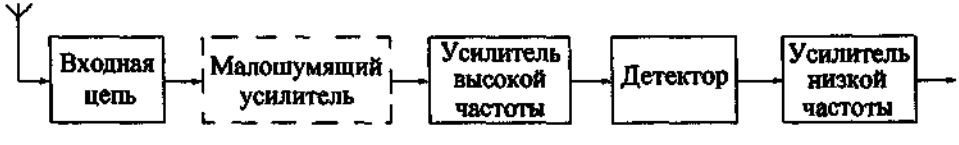



**Рис. 9.12.** Структурная схема приемника прямого усиления

Вместе с тем многокаскадные УВЧ позволяют в принципе не использовать регенеративные и сверхрегенеративные усилители, обеспечивающие большее усиление на каскад. Однако такие усилители обладают повышенными искажениями, относительно низкой устойчивостью по отношению к дестабилизирующим факторам, повышенной вероятностью паразитного излучения. Настройка приемника на полезный сигнал осуществляют перестройкой по частоте входной цепи, МШУ и УВЧ. Синхронная перестройка по частоте всех этих блоков является непростой задачей. В диапазоне СВЧ технически трудно согласовать полосы пропускания приемника с шириной спектра полезного сигнала для фильтрации последнего от помех, несовпадающих по частоте с сигналом. Отмеченные факторы являются недостатком приемников прямого усиления.

Приемники прямого усиления применялись в начале 20-го столетия и назывались *детекторными*, поскольку в них не осуществлялось усиление ни на каких частотах, а принятый антенной сигнал поступал непосредственно на амплитудный детектор. В настоящее время детекторные приемники используются в измерительной или регистрирующей технике СВЧ-диапазона. Термин «приемник прямого усиления» подчеркивает ту его особенность, что селекция и усиление производятся на несущей частоте принимаемого радиосигнала. Приемник прямого усиления имеет ряд существенных недостатков. В частности, для обеспечения высокой избирательности приходится увеличивать число высокодобротных резонансных контуров, что усложняет перестройку приемника по диапазону. Поэтому приемники прямого усиления находят ограниченное применение.

**Приемник прямого усиления с регенерацией** содержит регенеративный каскад, включаемый между антенной и оконечным устройством и выполняющий одновременно три функции: усиление высокой частоты, детектирование и усиление низкой частоты. В узком смысле под регенерацией в радиотехнике понимают компенсацию потерь в колебательной цепи с помощью положительной ОС. Явление регенерации используется в приемниках для усиления колебаний. По существу схема регенеративного усилителя не отличается от схемы автогенератора. Однако в регенеративном усилителе уровень обратной связи не доводится до значения, соответствующего порогу генерации колебаний. При этом осуществляется лишь частичная компенсация потерь в колебательной системе усилителя и действие ОС сводится к повышению добротности контура. Не анализируя их подробно, отметим, что приемники с регенерацией обладают низкой стабильностью и высоким уровнем искажений.

В схему сверхрегенеративного приемника дополнительно вводится источник вспомогательного гармонического напряжения, служащий для периодического возбуждения регенеративного усилителя. Сверхрегенеративные приемники имеют несколько лучшую стабильность работы, но очень большой уровень нелинейных искажений. Их достоинство — малая мощность источников питания при минимальных размерах и массе. Поэтому подобную структуру используют для портативных (обычно связных) радиоприемников, допускающих большой уровень искажений.

### Супергетеродинный приемник



Э. Армстронг

Существенное улучшение большинства показателей приемных устройств достигается на основе принципа преобразования частоты принимаемого сигнала — переноса в частотную область, где он может быть обработан с наибольшей эффективностью. Самое широкое распространение во всех диапазонах получила построенная на этом принципе схема *супергетеродинного приемника*.

Приемник супергетеродинного типа (в 1918 г. изобрел американский инженер Эдвин Армстронг — *Edwin Armstrong*; 1890–1954) обеспечивает очень высокую и практически одинаковую избирательность во всех диапазонах волн, а также более равномерное усиление в высокочастотном тракте. Это достигается введением в главный тракт приемника (рис. 9.13) *преобразователя частоты*, состоящего из *смесителя* (СМ), локального маломощного генератора высокой частоты — *гетеродина* (Г), *фильтра сосредоточенной селекции* (ФСС; он по существу является оптимальным фильтром для принимаемых сигналов) и *усилителя промежуточной частоты* (УПЧ). Часть приемника, включающая входную цепь, МШУ и УВЧ (*преселектор*), подобен структуре приемника прямого усиления и обеспечивает чувствительность и предварительную селекцию на частоте. Радиосигнал из антенны подается на вход МШУ и УВЧ, а затем на вход *смесителя* — специального элемента с двумя входами, осуществляющего операцию преобразования сигнала по частоте. На второй вход смесителя подается сигнал с гетеродина. Колебательный контур гетеродина перестраивается одновременно с входным контуром смесителя. Таким образом, на выходе смесителя преобразователя частоты образуются сигналы с частотой, равной сумме и разности частот гетеродина и принимаемой радиостанции. Разностный сигнал постоянной промежуточной частоты (обычно более низкой) выделяется с помощью ФСС и усиливается одним или несколькими каскадами УПЧ, затем поступает на детектор, восстанавливающий сигнал низкой (звуковой) частоты.

В супергетеродинном приемнике спектр принимаемого сигнала с помощью преобразователя частоты переносится с несущей частоты $f_0$ на промежуточную

$$f_{\text{пч}} = |f_0 - f_\text{г}| = |f_\text{г} - f_0|, \tag{9.4}$$

где $f_\text{г}$ — частота гетеродина.

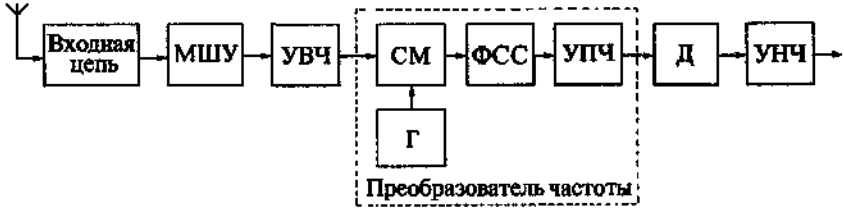



**Рис. 9.13.** Структурная схема супергетеродинного приемника

ФСС и УПЧ не перестраиваются по частоте, что позволяет получить в супергетеродинном приемнике высокую частотную избирательность при неизменной полосе пропускания, а также реализовать оптимальную фильтрацию сигнала от помех, применяя согласованные фильтры на промежуточной частоте. Таким образом, в супергетеродинном приемнике устраняются основные недостатки, присущие приемнику прямого усиления. Поскольку сигнал несет в себе полезную информацию, то в процессе преобразования частоты эту информацию необходимо сохранять, т. е. преобразователь частоты для полезных сигналов должен быть линейным устройством.

Вместе с тем, обладая принципиальными достоинствами, супергетеродинный приемник не лишен ряда недостатков. Один из них — наличие дополнительных паразитных каналов приема. Наиболее значительным недостатком является наличие *зеркального канала приема* (*the mirror channel*) — мешающего входного сигнала, который не может быть подавлен фильтрами ПЧ. Поскольку промежуточная частота в соответствии с (9.4) равна абсолютной разности частот принимаемого сигнала и гетеродина, то приемник может принимать сигналы двух передающих радиостанций с разными несущими частотами $f_0$, $f_з$ и соответственно спектрами $S(f_0)$, $S(f_з)$, расположенными симметрично (зеркально) относительно частоты гетеродина $f_г$ (рис. 9.14). Если, например, частота $f_0$ одного сигнала меньше, а другого $f_з$ — больше частоты гетеродина $f_г$ на $f_{пч}$, то на выходе УПЧ одновременно будут присутствовать два преобразованных сигнала с равными частотами — *основной* $f_{пч} = f_г - f_0$ и *зеркальной* $f_{зк} = f_{пч} = f_з - f_г$. Как нетрудно заметить, $f_з - f_0 = 2f_{пч}$, т. е. зеркальный канал отстоит от основного (полезного, принимаемого) на удвоенное значение промежуточной частоты. Таким образом, если частоты двух передатчиков различаются между собой на удвоенную величину промежуточной частоты и не принято никаких мер, чтобы ослабить зеркальный, то они будут почти одинаково хорошо слышны.

Поскольку зеркальный канал расположен «по другую сторону» спектра относительно частоты гетеродина на частоту $f_{пч}$, то и методы борьбы с ним в основном фильтровые. Подавление частот зеркального канала осуществляется с помощью специальных полосовых фильтров, включенных во входную цепь и УВЧ приемника. Они содержат резонансные контуры, которые настраиваются на частоту принимаемого сигнала $f_0$. Эти же фильтры ослабляют сигнал частоты зеркального канала $f_з$, по отношению к которой резонансные контуры расстроены (см. АЧХ $K(f)$ на рис. 9.14). Для уменьшения помех от зеркального канала применяют метод двойного (или даже тройного) преобразования частоты. Говоря о способах подавления зеркального канала, нужно отметить, что помимо самого распространенного фильтрового существует и фазовый метод подавле-

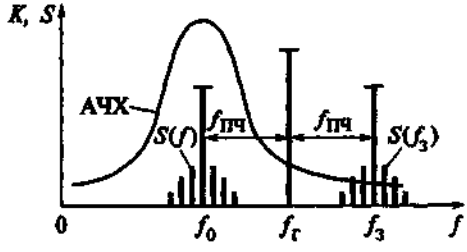



**Рис. 9.14.** Ослабление зеркального канала полосовыми фильтрами

ния. В частности, конструкции некоторых смесителей позволяют ослабить зеркальный канал в самом смесителе до 30 дБ и более.

*Соседний канал приема* — канал, частота которого меньше или больше на частоту $f_{пч}$ принимаемого сигнала. Из-за недостаточной избирательности он не отфильтровывается преселектором и образует в преобразователе частоты сигнал $f_{пр} = |f_г - f_{сн}| \approx f_{пч}$, попадающий в полосы пропускания ФСС (или фильтра на ПАВ) и поэтому усиливаемый и обрабатываемый наравне с полезным сигналом. Основная борьба с соседними каналами — повышение избирательности ФСС.

*Интермодуляционный канал приема* возникает при прохождении через преселектор вместе с полезным сигналом частоты $f_c$ двух (и более) внеполосных сигналов и помех на частотах $f_{п1}$, $f_{п2}$, которые в смесителе образуют колебания с комбинационными частотами $nf_c \pm mf_{п1} \pm pf_{п2} \pm \ldots$, где $n$, $m$, $p$ — целые числа. Если какая-либо одна или несколько из комбинационных частот попадает в полосу пропускания ФСС и УПЧ, то создается интермодуляционный побочный канал приема. Интермодуляционные составляющие усиливаются в тракте наравне с полезным сигналом, накладываются на него, снижают отношение сигнал/шум и искажают принимаемое сообщение. Особенно опасны ИМИ в системах связи с многостанционным доступом, когда интермодуляция возникает между полезными сигналами. Для уменьшения влияния последнего следует повышать частотную избирательность преселектора и уменьшать усиление в УВЧ.

Следует отметить, что подавление сигналов побочных каналов приема (в том числе и зеркального) улучшается при повышении промежуточной частоты $f_{пч}$. Однако при этом ухудшается избирательность УПЧ приемника.

При изменении частоты настройки высокочастотного тракта приемника необходимо также одновременно перестраивать частоту гетеродина. Это достигается путем сопряжения органов настроек высокочастотного тракта и гетеродина, для чего в приемнике применяется единственная ручка (обычно электронный элемент) настройки. С ее помощью одновременно изменяют резонансную частоту высокочастотного тракта и частоту колебаний гетеродина, сохраняя между ними фиксированную промежуточную частоту.

К недостаткам супергетеродинного приемника следует отнести и возможность приема помехи с частотой, близкой или равной промежуточной частоте. Такая помеха может пройти весь тракт без преобразования в преобразователе частоты и усилиться в УПЧ. Ее прием аналогичен приему по схеме прямого усиления независимо от частоты гетеродина. Для подавления подобных помех во входную цепь супергетеродинного приемника вводят режекторный фильтр, настроенный на промежуточную частоту.

В супергетеродинном приемнике возможно также появление так называемых *комбинационных каналов приема*, вызывающих *комбинационные свисты*. На некоторых частотах работы приемника в результате сложных нелинейных процессов в преобразователе частоты получается сигнал с частотой $f'_{пч}$, близкой к промежуточной частоте. Тогда полезный и паразитный сигналы одинаково усиливаются в УПЧ, но при этом возникают *биения* (нелинейные эффекты) их

несущих частот. В результате биений появляется низкочастотная огибающая полезного сигнала с разностной частотой $f_{св} = |f_{пч} - f'_{пч}|$, которая выделяется амплитудным детектором и после усиления в УПЧ прослушивается в виде свиста.

При проектировании супергетеродинного приемника побочные каналы приема могут быть практически устранены рациональным выбором промежуточной частоты, режима работы преобразователя частоты и необходимой частотной избирательности преселектора и усилителя промежуточной частоты. Наиболее эффективным является использование схем смесителя, выполняющих близкое к идеальному перемножение напряжений, а также электронной развязки цепей гетеродина и высокочастотного тракта. В настоящее время супергетеродинная структура приемника стала фактически стандартом в профессиональной и любительской радиосвязи.

**Автоматические устройства управления и регулировки приемника**

Современный приемник работает в часто меняющихся условиях. Могут различаться уровни принимаемых сигналов от разных передающих станций, возможна нестабильность уровня сигнала от одного передатчика из-за изменения условий распространения радиоволн. Часто возникает необходимость обеспечить перекрытие широкого частотного диапазона и высокую точность настройки на определенную несущую частоту. Для выполнения таких требований в приемник специально и вводят устройства, позволяющие вручную и автоматически, непосредственно или дистанционно изменять и регулировать коэффициент усиления, частоту настройки и подбирать оптимальный режим работы.

Из специальных устройств, обеспечивающих стабильную работу приемника, выделим системы *автоматической регулировки усиления* (АРУ; *Automatic gain control — AGC*), *автоматической подстройки частоты* (АПЧ; *Automatic frequency trim — AFT*) и *фазовой автоподстройки частоты* (ФАПЧ; *Phase-locked loop — PLL*). Отметим и *временну́ю автоматическую регулировку усиления* (ВАРУ; *Time automatic gain control — TAGC*), применяемую в приемниках РЛС. Все эти системы могут выполнятся как в аналоговом, так и цифровом виде.

**Автоматическая регулировка усиления**. Такая регулировка обеспечивает на выходе приемника практически неизменный уровень полезного сигнала при больших (50...100 дБ) колебаниях амплитуд входного сигнала. Структурная схема супергетеродинного приемника с АРУ и АПЧ, применяемая в большинстве аналоговых систем радиосвязи с амплитудно-модулированными колебаниями, приведена на рис. 9.15.

Действие АРУ основано на автоматическом изменении коэффициентов усиления отдельных каскадов усилителей приемника при колебаниях уровня входного сигнала. Главный элемент АРУ — амплитудный детектор АРУ (Д АРУ), который подключен к выходу УПЧ. Вырабатываемое Д АРУ управляющее напряжение $u_y$ должно быть пропорционально среднему уровню принимаемого сигнала и не зависеть от глубины амплитудной модуляции. Поэтому на выходе детектора АРУ включают ФНЧ (на рис. 9.15 для упрощения не показан).

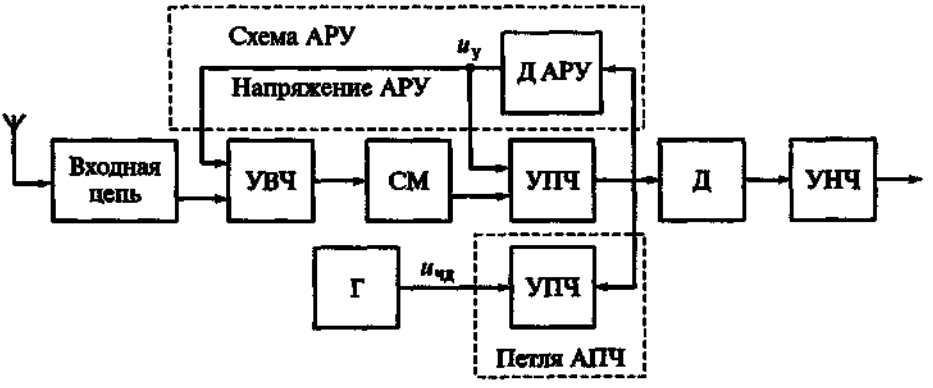



**Рис. 9.15.** Структурная схема супергетеродинного приемника с АРУ и АПЧ

Постоянное управляющее напряжение $u_у$ поступает на усилительные каскады и соответствующим образом изменяет режим работы активных элементов по постоянному току. Рассмотренная схема АРУ является наиболее простой и поясняет лишь принцип действия таких устройств. Основной недостаток данной схемы — уменьшение коэффициента усиления приемника даже при малых уровнях сигнала на выходе, что может привести к его полной потере. Поэтому в настоящее время используются более сложные схемы АРУ, позволяющие обеспечить оптимальные режимы работы усилительных каскадов приемника при любых приемлемых уровнях принимаемого сигнала.

**Автоматическая подстройка частоты.** Такую подстройку применяют для предотвращения самопроизвольной расстройки приемника с приема выбранной радиостанции, обусловленной нестабильностью частот передатчика и гетеродина. Нестабильность частоты настройки приемника проявляется в нестабильности промежуточной частоты. Чтобы эта частота изменялась в допустимых пределах или была стабильной, применяют АПЧ гетеродина (рис. 9.15).

Основным элементом любой системы АПЧ является частотный детектор (ЧД), подключенный к выходу УПЧ (петля АПЧ). Если промежуточная частота точно совпадает с номинальным значением, то напряжение на выходе частотного детектора равно нулю.

При отклонении значения промежуточной частоты от номинального на выходе частотного детектора появляется постоянное напряжение $u_{чд}$, значение которого пропорционально расстройке приемника, а полярность соответствует знаку расстройки. Это постоянное напряжение поступает на вход гетеродина и подстраивает его таким образом, чтобы на выходе УПЧ было номинальное значение промежуточной частоты. Система АПЧ отслеживает частоту принимаемой станции только в том случае, когда полезный сигнал имеет достаточный уровень, и на входе приемника не действует более мощный, близкий по частоте, мешающий сигнал. Если же такой мешающий сигнал поступает на вход приемного устройства, то система АПЧ может настроить приемник не на полезный, а на мешающий сигнал.

**Цифровая система АРУ.** По выполняемым функциям, принципу действия и структурной схеме система *цифровой автоматической регулировки усиления* (ЦАРУ) мало в чем отличается от аналоговой АРУ: специфика ЦАРУ состоит лишь в цифровой реализации узлов известной аналоговой структурной схемы. Одна из проблем, которую необходимо решить при разработке системы ЦАРУ, заключается в ее сопряжении с регулируемыми каскадами (усилителями с переменными коэффициентами усиления, управляемыми аттенюаторами и пр.). Если в регулируемом каскаде приемника имеется цифровой управляющий вход, то сопряжение не представляет никаких затруднений. При аналоговом же управлении регулируемым каскадом необходимо ввести в схему приемника дополнительный цифроаналоговый преобразователь.

Рассмотрим структурную схему системы цифровой АРУ, которая осуществляет регулировку коэффициента усиления каскадов УПЧ приемника (рис. 9.16). В схеме использованы усилительные каскады с нерегулируемым коэффициентом усиления, однако перед ними включен автоматический электронный аттенюатор в виде магазина сопротивлений, переключаемый цифровым кодом.

Для определения уровня регулируемого сигнала к выходу УПЧ подключен аналоговый *детектор* АРУ (Д АРУ). Сигнал на его выходе $U_{\text{Д АРУ}}$ подвергается квантованию в АЦП. Проанализируем принцип действия системы ЦАРУ при простейшем, бинарном квантовании сигнала в АЦП. Алгоритм работы схемы квантователя АЦП при этом следующий:

- если $U_{\text{Д АРУ}}$ меньше напряжения задержки $E_3$ (уровень выходного напряжения УПЧ $u_{\text{вых}}(t)$, при котором происходит срабатывание ЦАРУ), то вырабатывается сигнал ошибки $Z = -1$;
- если $U_{\text{Д АРУ}}$ больше напряжения задержки $E_3$, то $Z = +1$.

Данный сигнал ошибки $Z$ поступает в *цифровой интегратор* — усредняющий реверсивный счетчик (РС1) с коэффициентом счета $R_1$. При переполнении счетчика РС1 на его выходе появится импульс, увеличивающий или уменьшающий на единицу (в зависимости от знака переполнения) число во втором реверсивном счетчике (РС2). Состояние счетчика РС2, т. е. записанный в нем цифровой код, может изменяться в пределах от 0 до $R_{2\max}$. Этот счетчик не дол-

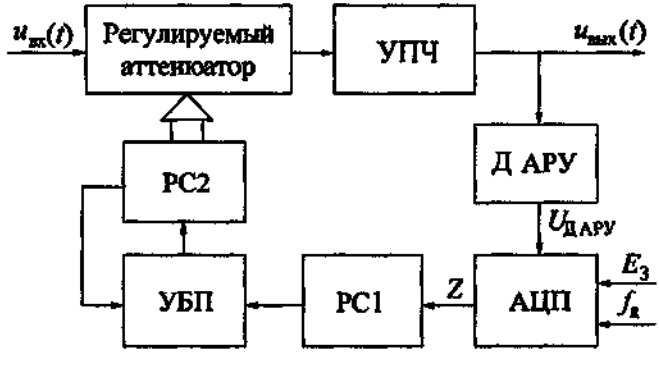



**Рис. 9.16.** Структурная схема системы цифровой АРУ

жен переполняться: когда код в РС2 достигнет значения 0 или $R_{2max}$, происходит блокировка поступления импульсов соответствующего знака с помощью устройства блокировки переполнения (УБП). Цифровой код счетчика РС2 дискретно регулирует коэффициент передачи управляемого аттенюатора, который, в свою очередь, изменяет коэффициент усиления УПЧ.

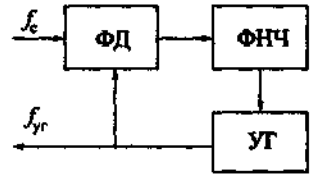



**Рис. 9.17.** Структурная схема системы ФАПЧ

Частоту $f_д$ дискретизирующей последовательности импульсов выбирают, по крайней мере, в 10...15 раз больше верхней частоты спектра сигнала на входе АЦП. Последняя определяется шириной полосы пропускания ФНЧ, включенного на выходе детектора АРУ. Желательно выбирать постоянную времени ФНЧ на порядок меньше минимально допустимой длительности переходных процессов в основных узлах схемы ЦАРУ, тогда эта длительность будет определяться цифровой, а не аналоговой частью устройства автоматического регулирования.

**Цифровая система ФАПЧ.** Структурная схема системы ФАПЧ приведена на рис. 9.17. В системе производится сравнение на фазовом детекторе фаз некоторого эталонного сигнала частоты $f_c$ и колебания частоты $f_{уг}$, вырабатываемого управляемым генератором (УГ). Затем осуществляется подстройка частоты $f_{уг}$ управляемого генератора под эталонную частоту $f_c$. Схема характеризует как аналоговую, так и цифровую систему ФАПЧ (ЦФАПЧ). В последнем случае все или некоторые блоки схемы на рис. 9.17 выполняют цифровыми. При этом возможны различные варианты реализации отдельных блоков. Так, ФД может быть построен по любой из схем, описанных в главе 7. При этом ФНЧ может быть в аналоговом или цифровом исполнении, что зависит от вида сигнала, управляющего УГ. Если УГ имеет управляющий вход аналогового вида, то предпочтительнее использование аналогового фильтра, а если же он управляется цифровым кодом, то предпочтительнее использовать ЦФ.

Структурная схема ЦФАПЧ показана на рис. 9.18, *а*. В качестве УГ в ней использована отдельная интегральная схема (например, 531ГГ1, обозначена Г; входы $Q$ и $E1$); ФД реализован на одном логическом элементе «Исключающее ИЛИ» (=1), а ФНЧ — простейший, аналоговый в виде интегрирующей $RC$-цепи.

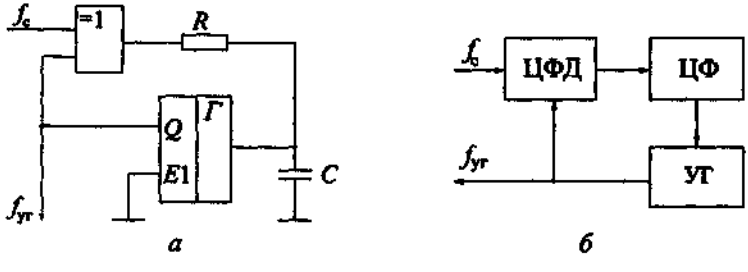



**Рис. 9.18.** Структурные схемы цифровых ФАПЧ:
*а* — упрощенная с аналоговым ФНЧ; *б* — полностью цифровая

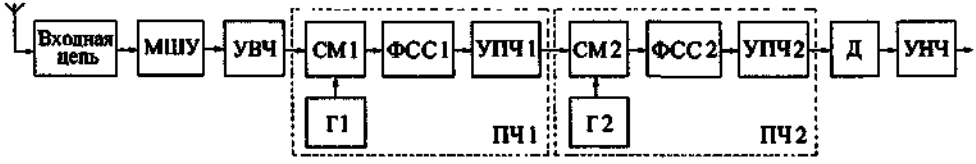



**Рис. 9.19.** Структурная схема приемника с двойным преобразованием частоты

Наиболее современной является полностью цифровая система ЦФАПЧ приемника, в которой фильтр нижних частот выполнен по схеме цифрового фильтра (ЦФ), в качестве фазового детектора использована схема цифрового ФД (ЦФД), а УГ имеет цифровое управление (рис. 9.18, *б*).

Использование высокостабильных гетеродинов позволяет принципиально решить задачу настройки приемника на частоту принимаемого сигнала. Для получения стабильных частот в неперестраиваемых (узкодиапазонных) приемниках в схему гетеродина вводят автогенератор с кварцевой стабилизацией частоты. Если же станционный передатчик работает на ряде фиксированных частот, то в приемниках в качестве гетеродинов используются синтезаторы частоты.

**Двойное преобразование частоты в приемниках.** Избирательность по зеркальному каналу повышают путем увеличения промежуточной частоты, что достигается ее двойным преобразованием. Рассмотрим структурную схему приемника с двойным преобразованием частоты (рис. 9.19), позволяющего перекрывать ряд частотных диапазонов и существенно снижать помехи зеркального канала.

В случаях, когда частоту несущей необходимо уменьшить в сотни раз, ее понижают вначале первым преобразователем ПЧ1, состоящим из *смесителя* (СМ1), *гетеродина* (Г1) и *усилителя промежуточной частоты* (УПЧ1), до первой промежуточной частоты. Обычно эта частота ниже частоты сигнала в 10...20 раз. Поскольку частота зеркального канала на 10...20 % отличается от несущей частоты, то первое преобразование позволяет существенно ослабить зеркальные помехи. Затем преобразованный сигнал подается на второй преобразователь ПЧ2, где с помощью *смесителя* (СМ2), *гетеродина* (Г2) и *усилителя промежуточной частоты* (УПЧ2) частота понижается до номинального значения.

### Тенденции развития радиоприемных устройств

В современной технике радиоприема достигнут серьезный прогресс, обусловленный интенсивным внедрением аналоговой и цифровой микросхемотехники. Имеющиеся сейчас микросхемы дают возможность разрабатывать приемники с высокой чувствительностью, лучшей избирательностью по зеркальному каналу, меньшими частотными и нелинейными искажениями, а также позволяют решить ряд проблем принципиально новыми путями. В частности, микропроцессоры обеспечивают оптимальное качество приема в условиях помех, управление автопоиском, электронную память десятков радиостанций, коммутацию программ, работу таймера, включающего и выключающего приемник по заданной программе, т. е. осуществление программного управления. Использу-

ется цифровая настройка (нажатием цифровых кнопок) и обзорная (приемник «просматривает» диапазон до получения команды «стоп», после чего фиксирует настройку и следит за станцией, обеспечивая в случае необходимости нужную подстройку усиления и частоты с помощью АРУ и АПЧ, переключение тракта УПЧ на более узкую полосу при появлении помехи по соседнему каналу, переключает регулятор тембра). Для дистанционного управления приемниками в пределах одного помещения применяют ультразвуковые и инфракрасные линии связи (инфракрасные линии более удобны). Сигналы управления с пульта дистанционного управления поступают на кодирующее устройство, в котором генерируется последовательность импульсов, поступающая на фотодиод, где осуществляется импульсно-кодовая модуляция инфракрасного излучения. Промодулированное излучение поступает на приемник (фототранзистор), затем на усилитель и декодирующее устройство и, наконец, на устройство управления.

Несомненные достоинства сулит использование цифрового радиовещания. Цифровая система передачи звука уже работает в системах спутниковой связи и спутникового радиовещания, а также используется для цифровой звукозаписи. Цифровое вещание может обеспечить практически неискаженное воспроизведение звука: полосу воспроизводимых частот 5...20 000 Гц, коэффициент нелинейных искажений менее 90 дБ, практически полное отсутствие внешних помех, а также осуществить стереофоническое вещание. Недостатком линий цифрового вещания является широкая полоса частот порядка 8 МГц, занимаемая одной радиостанцией, что определяет диапазоны несущих частот цифрового вещания. Цифровое радиовещание позволяет просто реализовать вывод информации на дисплей, режим повтора, запоминание сообщений и т. д.

В последнее время приемники с цифровой обработкой принимаемого сигнала в основном строились по смешанной схеме, в которой на цифровых микросхемах реализованы лишь отдельные узлы: детекторы, фильтры, синтезаторы частоты и пр. В новейших и перспективных разработках приемников используют в основном цифровую технику. Упрощенная структурная схема современного цифрового приемника показана на рис. 9.20. *Усилительный тракт* (УТ) приемника выполнен на аналоговых элементах и производит предварительную частотную фильтрацию принятого сигнала, усиление и преобразование его частоты. АЦП преобразует аналоговый сигнал в цифровой код, который подается на собственно цифровой приемник.

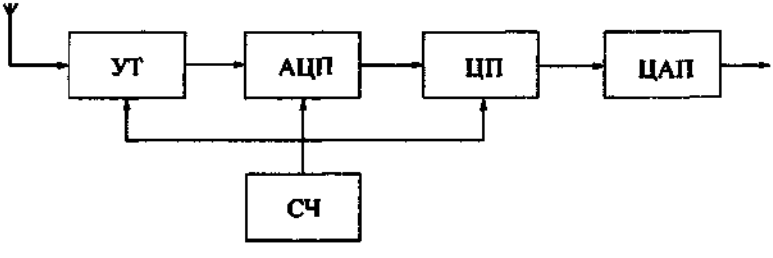



**Рис. 9.20.** Упрощенная структурная схема современного цифрового приемника

Последний представляет собой *цифровой процессор* (ЦП), осуществляющий цифровую обработку принятого сигнала по заданному алгоритму. Такой алгоритм включает задачу поиска сигнала по диапазону, дополнительного преобразования частоты, фильтрацию, детектирование и т. д. Если необходим сигнал в аналоговой форме, то на выходе приемника вводится ЦАП. Перестройка приемника по различным каналам производится с помощью *синтезатора частоты* (СЧ). Управление работой и регулировки новейших приемников выполняют с помощью микропроцессоров.

Сейчас все большее внимание уделяется использованию в бытовой радиоаппаратуре систем управления и оповещения человеческим голосом (с помощью речевых команд). Использование команд оператора подтверждается синтезированным человеческим голосом. Сигнал управления превращается в цифровую форму и поступает в микропроцессор управления. Системы распознавания голоса станут частью приемников, которые будут выполнять команды определенного человека. После исполнения команды микропроцессор вырабатывает сигнал ответа, который поступает в синтезатор человеческой речи, и громкоговоритель воспроизводит ответ. Решение же многих проблем ближайшего периода развития приемников будет зависеть в основном от разработки новых интегральных схем, а также от внедрения цифровой техники во все узлы радиосистем.

## 9.3. Оптимальная линейная фильтрация сигналов

Передача сигналов сопровождается шумами, которые искажают передаваемую информацию. Поэтому на протяжении всего развития радиотехники (в частности, приемных устройств) центральной проблемой была и остается борьба с помехами и шумами (далее, обобщенно, шумами). В случаях, когда мощность полезного сигнала соизмерима со средней мощностью шума, трудно не только выделить, но и обнаружить сигнал. Одним из путей повышения помехоустойчивости являются различия между собой сигналов, соответствующих различным сообщениям. Затем нужно выбрать такой метод приема, который наилучшим образом реализует различие между сигналами. При этом важнейшей задачей является выделение информации с максимальной достоверностью — оптимальный (согласованный) прием. Для этого в состав приемника включают оптимальный фильтр, цепи последетекторной обработки, следящие схемы АПЧ и ФАПЧ.

### Оптимальный (согласованный) линейный фильтр

Уменьшение влияния шумов достигают различными способами, в том числе выбором наилучших характеристик цепей, через которые проходит смесь сигнала и шума. Основой большинства практических методов выделения сигнала из *аддитивной* смеси сигнала и шума в приемниках является *оптимальная линейная фильтрация*, использующая линейные частотные фильтры. Удобнее всего описывают оптимальные фильтры с помощью импульсной или частотной (коэффициент передачи) характеристики.

**Критерии оптимального приема радиосигналов.** В зависимости от назначения системы передачи информации и характера принимаемого сигнала на фоне действующих помех принимают различные критерии оптимального приема. В одних случаях критерием является обнаружение полезного сигнала, в других — разрешение сигналов, в третьем — измерение параметров этого сигнала.

Под *обнаружением* понимают сам факт приема радиосигнала. Такой случай характерен для радиолокации, одна из задач которой состоит в обнаружении предполагаемой цели. Поступление на вход радиоприемного устройства сигнала, отраженного от цели, свидетельствует об ее обнаружении. В системах радиоэлектронной борьбы обнаружение состоит в определении факта выхода в эфир радиопередатчика предполагаемого противника.

Под *разрешением* сигналов понимают, какое именно из нескольких возможных переданных сообщений поступило на вход радиоприемного устройства. Например, при передаче цифровых сообщений двоичным кодом необходимо определить, какой бит, 1 или 0 передан в данный момент по радиоканалу.

Измерение параметров сигналов позволяет извлечь необходимую информацию об объекте, с которого она поступила. В частности, измерение частоты принятого сигнала позволяет на основе эффекта Доплера судить о скорости движения объекта. В другом случае по времени задержки прихода сигнала можно определить расстояние до объекта.

Качество принятого сообщения в зависимости от его характера оценивается по-разному. Так, в цифровых системах передачи битовой информации это качество определяется вероятностью ошибки принятого символа. Например, если эта вероятность составляет $10^{-3}$, то это означает, что из тысячи переданных бит один может быть ошибочным. При передаче речи качество принятого сообщения оценивается по его разборчивости, т. е. по количеству правильно понятых слов, смысл которых не искажен. При передаче телевизионного сигнала вводится несколько критериев, по которым оценивается качество принятого изображения на экране телевизора. Отмеченные разнородные критерии при передаче аналоговых сообщений являются функцией отношения мощности сигнала к мощности помехи (шума) на выходе радиоприемного устройства.

При *обнаружении* сигнала в шумах наиболее эффективен *критерий максимума отношения сигнал/шум* по мощности на выходе фильтра. Линейный фильтр, для которого это отношение максимально, называют *оптимальным* (подразумевая наилучшим), или *согласованным* (*matched*), фильтром, а также *коррелятором* (*correlator*).

**Отношение сигнал/шум.** Используем в приемнике линейный фильтр с таким частотным коэффициентом передачи $K(\omega)$, что значения его модуля $|K(\omega)|$ велики в частотном диапазоне, где сконцентрирована основная доля мощности полезного сигнала, и малы в частотных областях, где существенна спектральная плотность мощности шума. Следует ожидать, что при подаче на вход оптимального фильтра аддитивной суммы полезного сигнала и шума на его выходе можно получить заметное увеличение отношения сигнал/шум.

Оценим количественно данное положение. Пусть на входе линейного фильтра приемника присутствует входное колебание, являющееся суммой полезного сигнала $u(t)$ и шума $r(t)$, т. е.

$$z(t) = u(t) + r(t). \tag{9.5}$$

На практике и сигнал и шум являются узкополосными с одинаковыми центральными частотами $\omega_0$. Кроме того, они некоррелированны, т. е. среднее значение их произведения на некотором интервале $T$:

$$\overline{u\,r} = \lim \frac{1}{T}\int_0^T u(t)r(t)dt = 0. \tag{9.6}$$

Также предположим стационарность сигнала и шумов на неограниченно протяженном интервале времени.

Интенсивность колебаний на входе фильтра можно характеризовать значением среднего квадрата (средней мощности) входного сигнала, которая в силу равенства (9.6) есть сумма средних квадратов полезного сигнала и шума:

$$\overline{z^2} = \overline{u^2} + \overline{r^2} = \overline{u^2} + \overline{\sigma^2}, \tag{9.7}$$

где $\overline{\sigma^2}$ — дисперсия (мощность) входного шума.

Для описания относительного уровня полезного сигнала вводят так называемое *отношение сигнал/шум* на входе фильтра по формуле

$$Q_{\text{вх}} = \overline{u^2} / \overline{\sigma^2}. \tag{9.8}$$

Отметим, что безразмерное число $Q_{\text{вх}}$ характеризует уровень сигнала по отношению к уровню шума весьма приближенно и неполно. Пользоваться этим отношением целесообразно лишь тогда, когда заранее известно, что временны́е колебания сигнала и реализации шума в определенном смысле «схожи» между собой. Так, входной шум обычно хорошо описывается моделью нормального узкополосного случайного процесса. Отдельные реализации данного шума представляют собой квазигармонические колебания. Естественно, что в этом случае можно пользоваться формулой (9.8) для оценки уровня полезных модулированных сигналов с амплитудной или частотной модуляцией.

**Пример 9.1.** Пусть на вход линейного фильтра воздействуют однотональный АМ-сигнал $u_{\text{АМ}}(t) = U_{\text{н}}[1 + M\cos(\Omega t + \theta_0)]\cos\omega_0 t$ и гауссов шум $r(t)$, односторонний спектр мощности которого

$$F_r(\omega) = \begin{cases} F_0, & \omega_0 - 2\Omega < \omega < \omega_0 - 2\Omega; \\ 0 & \text{при иных значениях } \omega. \end{cases}$$

Найти отношение сигнал/шум на входе фильтра.

Р е ш е н и е. Средняя мощность АМ-сигнала согласно (2.122)

$$P_{\text{ср}} = 0{,}5U_{\text{н}}^2 + 0{,}25U_{\text{н}}^2M^2.$$

Здесь первое слагаемое $0{,}5U_н^2$ соответствует средней мощности несущего колебания, которое не содержит информации о передаваемом сообщении. Поэтому при расчетах помехоустойчивости принято опускать эту составляющую и считать

$$P_{ср} = 0{,}25U_н^2M^2.$$

Дисперсия шума на входе фильтра

$$\sigma^2 = \int_0^\infty F_r(\omega)d\omega = \int_{\omega_0-2\Omega}^{\omega_0+2\Omega} F_0\,d\omega = 4F_0\Omega.$$

Отношение сигнал/шум

$$Q_{вх} = U_н^2M^2/(16F_0\Omega).$$

оказывается прямо пропорциональным квадрату коэффициента модуляции и обратно пропорциональным частоте модуляции.

Одним из основных параметров фильтров приемника является коэффициент передачи. Определим *коэффициент передачи оптимального фильтра* приемника при условии, что сигнал принимается на фоне белого шума с двусторонней спектральной плотностью мощности $W_0$ (напомним, что часто белый шум удобнее задавать односторонней, т. е. в области физических частот, спектральной плотностью мощности $N_0 = 2W_0$).

Для удобства анализа представим коэффициент передачи оптимального фильтра в виде

$$\mathbf{K}(\omega) = K(\omega)e^{j\varphi_k(\omega)}, \tag{9.9}$$

где $K(\omega)$ — АЧХ; $\varphi_k(\omega)$ — ФЧХ фильтра.

Пусть *входной сигнал* $u(t)$ имеет спектральную плотность

$$\mathbf{S}(\omega) = S(\omega)e^{j\varphi_c(\omega)}. \tag{9.10}$$

Здесь $S(\omega)$ и $\varphi_c(\omega)$ — соответственно амплитудный и фазовый спектры принимаемого сигнала.

Отметим некоторый, пока неизвестный, момент времени $t = t_0$, при котором отношение сигнал/шум на выходе фильтра будет максимальным. В соответствии с формулой (4.9) сигнал на выходе фильтра (линейного четырехполюсника)

$$u_{вых}(t_0) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathbf{S}(\omega)\mathbf{K}(\omega)^{j\omega t_0}\,d\omega. \tag{9.11}$$

Поскольку $\mathbf{S}_{вых}(\omega) = \mathbf{S}_{вх}(\omega)\mathbf{K}(\omega)$, то с помощью соотношения (3.57) находим *среднюю мощность (дисперсию) белого шума на выходе фильтра*

$$\sigma^2 = \frac{1}{2\pi}\int_{-\infty}^{\infty}W_0\mathbf{K}^2(\omega)d\omega = \frac{W_0}{2\pi}\int_{-\infty}^{\infty}\mathbf{K}^2(\omega)d\omega. \tag{9.12}$$



О. Коши В. Буняковский К. Шварц

Используя (9.11) и (9.12), найдем отношение мощностей сигнала и шума

$$Q_{\text{вых}} = \frac{|u_{\text{вых}}(t_0)|^2}{\sigma^2} = \frac{\left|\frac{1}{2\pi}\int_{-\infty}^{\infty}\mathbf{S}(\omega)\mathbf{K}(\omega)e^{j\omega t_0}d\omega\right|^2}{\frac{W_0}{2\pi}\int_{-\infty}^{\infty}\mathbf{K}^2(\omega)d\omega}. \qquad (9.13)$$

Для удобства вычислений введем эквивалентный коэффициент передачи линейного фильтра

$$\mathbf{K}_э(\omega) = \mathbf{K}(\omega)e^{j\omega t_0} = K(\omega)e^{j[\varphi_k(\omega)+\omega\ t_0]}. \qquad (9.14)$$

Оптимальный коэффициент передачи анализируемого фильтра максимизирует правую часть выражения (9.13).

Задачу нахождения оптимального коэффициента передачи $\mathbf{K}(\omega)$ решают на основе уже упоминавшегося неравенства Буняковского–Коши–Шварца (Виктор Яковлевич Буняковский — русский математик; 1804–1889; Огюстен Коши — Augustin Cauchi — французский математик; 1789–1857; Карл Шварц — Karl Schwarz — немецкий математик; 1843–1921), для этого случая имеющего вид

$$\left|\int_{-\infty}^{\infty}\mathbf{S}(\omega)\mathbf{K}_э(\omega)d\omega\right|^2 \le \int_{-\infty}^{\infty}|\mathbf{S}(\omega)|^2 d\omega \int_{-\infty}^{\infty}|\mathbf{K}_э(\omega)|^2 d\omega. \qquad (9.15)$$

Прямая подстановка показывает, неравенство обращается в равенство, если

$$\mathbf{K}_э(\omega) = A\mathbf{S}^*(\omega), \qquad (9.16)$$

где $A$ — произвольный постоянный коэффициент; $\mathbf{S}^*(\omega)$ — функция комплексно-сопряженная с $\mathbf{S}(\omega)$.

Представим эквивалентный коэффициент передачи (9.16) в виде произведения с фазовым множителем

$$\mathbf{K}_э(\omega) = \mathbf{K}(\omega)e^{j\omega t_0} = A\mathbf{S}^*(\omega) = AS(\omega)e^{-j\varphi_c(\omega)}.$$

Отсюда находим коэффициент передачи фильтра

$$\mathbf{K}(\omega) = AS^*(\omega)e^{-j\omega t_0} = AS(\omega)e^{-j[\varphi_c(\omega)+\omega t_0]}. \qquad (9.17)$$

Формула (9.17) полностью определяет коэффициент передачи оптимального фильтра, максимизирующего отношение сигнал/шум. Отсюда же следуют требования к АЧХ и ФЧХ оптимального фильтра:

$$K(\omega) = AS(\omega); \qquad (9.18)$$

$$\varphi_k(\omega) = -[\varphi_c(\omega) + \omega t_0]. \qquad (9.19)$$

По определению частотный коэффициент передачи — безразмерная величина, поэтому *постоянный коэффициент A должен иметь размерность, обратную размерности амплитудного спектра входного сигнала* $S(\omega)$.

Суть метода обработки принимаемого сигнала оптимальным фильтром приемника поясняет рис. 9.21, где соответственно показаны и обозначены: спектры входных сигнала $S(\omega)$ и белого шума $W_0$; спектр выходного сигнала $S_{вых}(\omega)$ и АЧХ фильтра $\mathbf{K}(\omega)$; энергетический спектр выходного шума $\sigma_x^2(\omega)$.

Полученные результаты имеют глубокий физический смысл. Соотношение (9.18) устанавливает, что АЧХ фильтра $K(\omega)$ должна с точностью до масштабного множителя $A$ совпадать по форме с амплитудным спектром $S(\omega)$ входного сигнала. Благодаря этому, подавляющая часть спектральных составляющих входного сигнала, имеющих наибольшие амплитуды, проходит на выход оптимального фильтра почти без ослабления и вносит основной вклад в образование его пикового значения. Из множества же спектральных компонентов входного белого шума, располагающихся в бесконечной полосе частот, на выход фильтра проходят и не ослабляются только те, которые находятся под кривой его АЧХ, т. е. в ограниченной полосе частот.

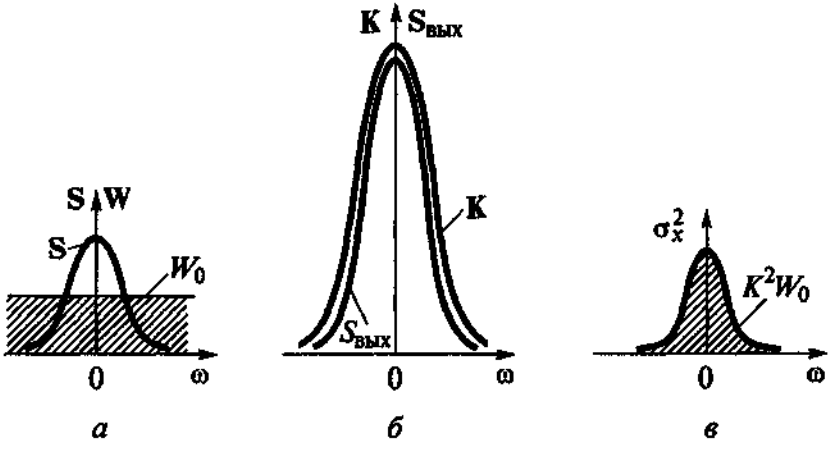



**Рис. 9.21.** Оптимальная фильтрация:
*а* — спектры входных сигнала и шума; *б* — спектр выходного сигнала и АЧХ фильтра;

Это приводит к ослаблению средней мощности шума $\sigma_x^2$ на выходе фильтра по сравнению со спектральной плотностью мощности белого шума $W_0$ на входе. В результате такого действия отношение сигнал/шум на выходе оптимального фильтра увеличивается.

Формулу (9.19), описывающую фазочастотную характеристику оптимального фильтра, можно трактовать как *условие компенсации начальных фаз* всех гармонических составляющих спектра выходного сигнала. Согласно этому условию, оптимальный фильтр должен иметь такую ФЧХ, чтобы получаемый в нем фазовый сдвиг каждой гармоники $-\varphi_c(\omega)$ был равен по значению и противоположен по знаку начальной фазе соответствующей составляющей спектральной плотности $S(\omega)$ входного сигнала. Оптимальный фильтр проводит *компенсацию* («*обнуление*») начальных фаз всех спектральных составляющих входного сигнала $u(t)$, в результате чего и образуется пик (максимум) выходного сигнала. Составляющая ФЧХ $-\omega t_0$ указывает на то, что пик выходного сигнала задержан относительно начала действия входного сигнала на интервал времени $t_0$. Связь между фазовой характеристикой $\varphi_c(\omega)$ входного сигнала, компенсирующей ее фазовой характеристикой $-\varphi_c(\omega)$ и ФЧХ фильтра поясняется рис. 9.22. Фазовая характеристика выходного сигнала, определяемая формулой

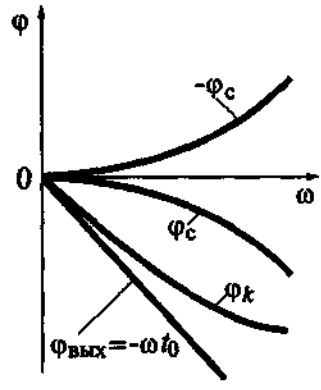



**Рис. 9.22.** Связь между фазовой характеристикой сигнала и ФЧХ фильтра

$$\varphi_{вых}(\omega) = \varphi_c(\omega) + \varphi_k(\omega) = \varphi_c(\omega) + [-\varphi_c(\omega) - \omega t_0] = -\omega t_0, \qquad (9.20)$$

показана на рис. 9.22 прямой линией.

Таким образом, коэффициент передачи фильтра, описываемый соотношением (9.9), *согласован* с амплитудным и фазовым (или фазовой характеристикой) спектрами входного сигнала. Поэтому рассмотренный оптимальный линейный фильтр часто называют *согласованным*.

Вернемся вновь к формуле (9.13) и рассмотрим энергетические соотношения между принимаемым сигналом и шумом на выходе исследуемого оптимального фильтра. Поскольку квадрат модуля комплексного числа равен квадрату его действительной части, то после несложных преобразований получим следующее выражение:

$$Q_{вых} = \frac{|u_{вых}(t_0)|^2}{\sigma_x^2} = \frac{\left|\frac{1}{2\pi}\int_{-\infty}^{\infty} S(\omega)K(\omega)e^{j\omega t_0}d\omega\right|^2}{\frac{W_0}{2\pi}\int_{-\infty}^{\infty} K^2(\omega)d\omega} = \frac{\frac{1}{2\pi}\int_{-\infty}^{\infty} S^2(\omega)d\omega}{W_0}. \qquad (9.21)$$

Числитель в (9.21) в соответствии с равенством Парсеваля представляет собой энергию входного сигнала Э. Тогда последнее соотношение примет вид:

$$Q_{\text{вых}} = \frac{|u_{\text{вых}}(t_0)|^2}{\sigma_x^2} = \frac{\text{Э}}{W_0} = \frac{2\text{Э}}{N_0}. \qquad (9.22)$$

Согласно выражению (9.22), *оптимальный фильтр максимизирует отношения сигнал/шум* (а не согласует отношение сигнал/шум, выражение «согласованный по сигнал/шум» — бессмысленно), *которое зависит только от энергии входного сигнала и спектральной плотности мощности белого шума и совершенно не связано с формой входного сигнала.*

**Пример 9.2.** Полезный сигнал, поступающий на вход оптимального фильтра, представляет собой прямоугольный видеоимпульс с некоторой амплитудой $E$ и длительностью $\tau_и$ = 10 мкс. Белый шум на входе фильтра имеет спектральную плотность мощности $W_0 = 25 \cdot 10^{-18}$ В$^2$/Гц. Определить минимальное значение амплитуды $E$, при которой возможно обнаружение сигнала, если приемник регистрирует его присутствие при отношении сигнал/шум $Q_{\text{вых}} = \text{Э}/W_0 = 2$ дБ (в практических схемах надежная регистрация осуществляется при $Q_{\text{вых}} \geq 13$ дБ).

Р е ш е н и е. Требуемое значение отношения сигнал/шум найдем из условия $10 \lg(\text{Э}/W_0) = 2$, откуда $\text{Э}/W_0 = 1{,}59$. Поскольку энергия прямоугольного импульса $\text{Э} = E^2\tau_и$, то

$$E_{\text{мин}} = \sqrt{1{,}58 W_0/\tau_и} \approx 4 \text{ мкВ}.$$

**Импульсная характеристика оптимального фильтра.** Чтобы определить импульсную характеристику оптимального фильтра, вычислим обратное преобразование Фурье от частотного коэффициента передачи (9.17). Используя уже применяемую ранее формулу для определения импульсной характеристики через коэффициент передачи

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

получим

$$h(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathbf{K}(\omega) e^{j\omega t} d\omega = \frac{A}{2\pi}\int_{-\infty}^{\infty} \mathbf{S}^*(\omega) e^{j\omega(t-t_0)} d\omega. \qquad (9.23)$$

Поскольку $\mathbf{S}^*(\omega) = \mathbf{S}(-\omega)$, то, переходя к новой переменной $\omega_2 = -\omega$, после несложных преобразований, запишем

$$h(t) = \frac{A}{2\pi}\int_{\infty}^{-\infty} \mathbf{S}(\omega_2) e^{-j\omega_2(t-t_0)} d\omega_2 = \frac{A}{2\pi}\int_{-\infty}^{\infty} \mathbf{S}(\omega_2) e^{j\omega_2(t_0-t)} d\omega_2 = Au(t_0 - t). \qquad (9.24)$$

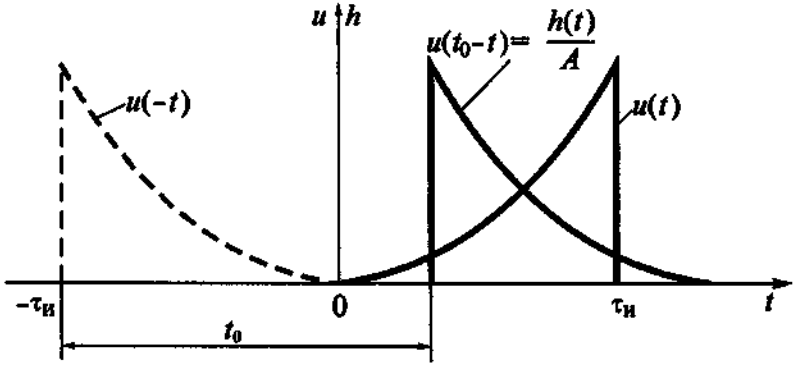



**Рис. 9.23.** К построению импульсной характеристики оптимального фильтра

Следовательно, импульсная характеристика оптимального фильтра совпадает с зеркально отраженной относительно оси ординат копией входного сигнала, сдвинутой на интервал $t_0$ по оси времени. Об этом говорит отрицательный знак при аргументе $t$ в формуле (9.24). На рис. 9.23 показан принцип построения импульсной характеристики оптимального фильтра применительно к некоторому импульсному сигналу $u(t)$ длительностью $\tau_и$. Поскольку при $t < 0$ импульсная характеристика линейной цепи не существует, то временна́я задержка $t_0$ между началом действия сигнала на входе фильтра и моментом образования максимального пика сигнала на его выходе должна быть *не менее длительности сигнала* $\tau_и$. Это одно (но недостаточное) из условий *физической реализуемости* оптимального фильтра, показывающее, что для создания максимального пика сигнала на выходе надо провести обработку фильтром всего входного сигнала $u(t)$.

Фундаментальной особенностью оптимального фильтра является то, что обнаружение с его помощью сигнала в шумах зависит *не от формы, а от его энергии*. В частности, путем увеличения длительности входного импульса можно надежно обнаруживать сигналы небольшой амплитуды. Однако при этом проигрывают в скоростях обработки информации. Как правило, формы полезного сигнала на входе и выходе согласованного фильтра существенно отличаются друг от друга. В частности, задачей согласованного фильтра для двоичной системы является не восстановление формы сигнала, искаженной шумом, а получение одного отсчета, по которому можно судить о присутствии или отсутствии на входе фильтра сигнала известной формы.

За согласованным фильтром в приемнике может находиться *выравнивающий фильтр* (*equalizing filter*), или эквалайзер; он необходим только в цифровых системах связи, в которых сигнал может искажаться вследствие межсимвольной интерференции, внесенной каналом. Принимающий и выравнивающий фильтры являются отдельными устройствами, что подчеркивает различие их функций. Впрочем, в большинстве случаев при использовании эквалайзера для выполнения обеих функций (а следовательно, и для компенсации искажения, внесенного передатчиком и каналом) можно включать единый фильтр. Такой составной фильтр называют просто *выравнивающим* или *принимающим и выравнивающим*.

Согласованным фильтром может быть пассивный фильтр на линиях задержки, или коррелятор, или специальное цифровое устройство, преобразующее входную смесь сигнал/шум в частотную область, умножающее полученный спектр на спектр, комплексно-сопряженный со спектром входного сигнала, на который настроен оптимальный приемник, и возвращающий результат обратно во временну́ю область. Но в любом случае это будет устройство, АЧХ которого повторяет амплитудный спектр сигнала, а ФЧХ — есть зеркальное отражение фазовой характеристики сигнала. Согласованный с неким сигналом фильтр — это линейный четырехполюсник, импульсная характеристика которого является зеркальным отражением этого сигнала.

Отметим, что функцию оптимального фильтра для входного сигнала в приемнике может выполнять коррелятор, что имеет важное практическое значение, поскольку в ряде случаев реализовать коррелятор проще, чем оптимальный фильтр. В возможности выполнять коррелятором функцию оптимального фильтра можно убедиться, сравнив спектры сигналов на выходе оптимального фильтра и коррелятора.

### Реализация согласованных фильтров

Найденные выражения, определяющие амплитудно-частотную и импульсную характеристики согласованного фильтра, дают возможность определить физическую структуру устройства для оптимальной фильтрации входного сигнала известной формы.

Оценим возможности реализации согласованных фильтров. Согласованный фильтр для импульсного сигнала произвольного вида $u(t)$ длительностью $T_c$ можно в принципе построить на основе неискажающей длинной линии с бесконечной плотностью отводов, обеспечивающей задержку сигнала на время $T_c$. Практически можно брать отводы в дискретных точках с разносом $Д = 0{,}5/F_c$, где $F_c$ — эффективная ширина спектра сигнала.

**Согласованный фильтр для прямоугольного видеоимпульса.** Пусть имеется сигнал $u(t)$, представляющий собой видеоимпульс прямоугольной формы с амплитудой $U_0$ и длительностью $\tau_и$. Для нахождения структуры фильтра, согласованного с данным сигналом, применим спектральный метод. Воспользовавшись прямым преобразованием Фурье, вычислим спектральную плотность прямоугольного видеоимпульса

$$\mathbf{S}(\omega) = \int_{-\infty}^{\infty} u(t)\mathrm{e}^{-j\omega t}dt = U_0 \int_0^{\tau_и} \mathrm{e}^{-j\omega t}dt = \frac{U_0}{j\omega}(1-\mathrm{e}^{-j\omega\tau_и}). \qquad (9.25)$$

По формуле (9.17), в которой для конкретности приравняем $t_0 = \tau_и$, т. е. отклик фильтра максимален в момент окончания импульса, находим частотный коэффициент передачи согласованного фильтра

$$\mathbf{K}(\omega) = A\mathbf{S}^*(\omega)\mathrm{e}^{-j\omega t_0} = AU_0\frac{1}{-j\omega}(1-\mathrm{e}^{j\omega\tau_и})\mathrm{e}^{-j\omega\tau_и} = AU_0\frac{1-\mathrm{e}^{-j\omega\tau_и}}{j\omega}. \qquad (9.26)$$

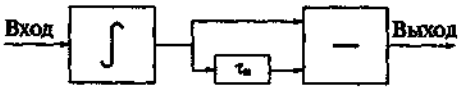



**Рис. 9.24.** Структурная схема согласованного фильтра для прямоугольного видеоимпульса

Показанную на рис. 9.24 структурную схему согласованного фильтра для прямоугольного импульса можно непосредственно построить по коэффициенту передачи (9.26). Входящий в (9.26) множитель $1/(j\omega)$ реализуется интегрирующим звеном, а второй множитель $1-e^{-j\omega\tau_и}$ — вычитающим устройством, на которое после интегрирования подаются прямой и задержанный на время $\tau_и$ сигналы. Известно, что коэффициент передачи идеальной (без потерь) линии задержки $K(\omega) = e^{-j\omega\tau_и}$.

Согласованный фильтр для пачки видеоимпульсов. Рассмотрим пачку из $N$ одинаковых импульсов, следующих с периодом $T$ и имеющих длительность $\tau_и$. Если $S_0(\omega)$ — спектральная плотность одного импульса, то спектральная плотность пачки импульсов согласно теореме линейности преобразования Фурье

$$S_п(\omega) = S_0(\omega)[1 + e^{-j\omega T} + e^{-j2\omega T} + \ldots + e^{-j(N-1)\omega T}]. \qquad (9.27)$$

При синтезе структуры согласованного фильтра для пачки импульсов потребуем, чтобы пик выходного сигнала возникал в момент окончания последнего импульса пачки, т. е. $t_0 = (N-1)T + \tau_и$. Воспользовавшись формулой (9.17), находим частотный коэффициент передачи согласованного фильтра

$$K(\omega) = AS_0^*(\omega)e^{-j\omega\tau_и}\left[1 + e^{j\omega T} + e^{j2\omega T} + \ldots + e^{j(N-1)\omega T}\right]e^{-j(N-1)\omega T} =$$

$$= K_0(\omega)\left[1 + e^{-j\omega T} + e^{-j2\omega T} + \ldots + e^{-j(N-1)\omega T}\right], \qquad (9.28)$$

где $K_0(\omega)$ — коэффициент передачи согласованного фильтра для видеоимпульса.

Формула (9.28) непосредственно определяет структурную схему согласованного фильтра, показанную на рис. 9.25. На входе схемы включен согласованный фильтр для одиночного видеоимпульса с частотным коэффициентом передачи $K_0 = K_0(\omega)$. Основой схемы служит многоотводная линия задержки, обеспечивающая запаздывание импульсов на интервалы $T, 2T, \ldots (N-1)T$.

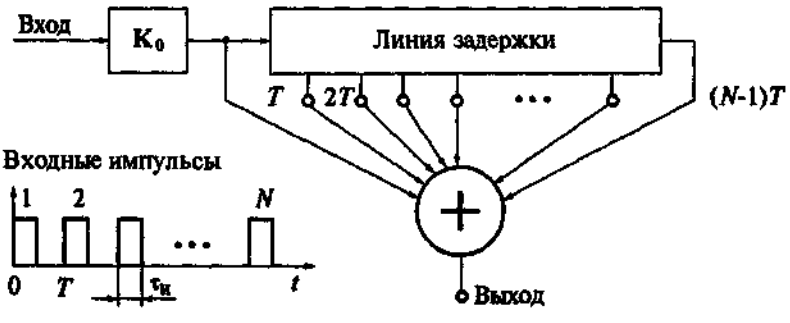



**Рис. 9.25.** Структурная схема согласованного фильтра для пачки видеоимпульсов

Сигналы со всех отводов поступают на сумматор. Нетрудно заметить, что максимальный отклик на выходе сумматора будет наблюдаться тогда, когда полезные сигналы от всех импульсов пачки одновременно окажутся на всех его входах. Эффективность работы схемы тем выше, чем длиннее пачка импульсов.

**Согласованный фильтр для прямоугольного радиоимпульса.** Рассмотрим сигнал, представляющий собой радиоимпульс прямоугольной формы:

$$u_{\text{р}}(t)=\begin{cases}U_{\text{и}}\cos\omega_0 t, & -\tau_{\text{и}}/2\le t\le\tau_{\text{и}}/2;\\ 0, & t\le\tau_{\text{и}}/2,\quad t\ge\tau_{\text{и}}/2.\end{cases} \tag{9.29}$$

Реализуем согласованный фильтр для такого сигнала, используя его импульсную характеристику. Формула (9.24) показывает, что импульсная характеристика согласованного фильтра $h(t) = Au_{\text{р}}(t_0 - t)$. Положим $t_0 = \tau_{\text{и}}$ и для упрощения будем считать, что длительность импульса кратна периоду высокочастотного заполнения, так что $\cos\omega\tau_{\text{и}} = 0$. Тогда

$$h(t)=\begin{cases}AU_{\text{и}}\cos\omega_0 t, & -\tau_{\text{и}}/2\le t\le\tau_{\text{и}}/2;\\ 0, & t\le\tau_{\text{и}}/2,\quad t\ge\tau_{\text{и}}/2.\end{cases} \tag{9.30}$$

т. е. импульсная характеристика согласованного фильтра с точностью до амплитудного множителя $A$ повторяет входной сигнал.

Такую импульсную характеристику можно упрощенно реализовать с помощью устройства, структурная схема которого приведена на рис. 9.26.

На входе линейного фильтра введено идеальное колебательное звено (например, высокодобротный контур без потерь, настроенный на частоту $\omega_0$). Для того чтобы импульсная характеристика согласованного фильтра равнялась нулю при $t > \tau_{\text{и}}$, в схему включены сумматор, на один из входов которого сигнал с выхода колебательного звена подается непосредственно, а на другой — через линию задержки на длительность $\tau_{\text{и}}$, и фазовращатель, изменяющий фазу сигнала на 180°. При таком включении элементов начиная с момента $t = \tau_{\text{и}}$ к входам сумматора приложены два гармонических колебания с равными амплитудами и противоположными фазами, что обращает в нуль сигнал на выходе сумматора.

В радиотехнике для фильтрации сигнала на фоне шума часто вместо согласованных используют фильтры, характеристики которых лишь частично согласованы с характеристиками сигнала. Такие фильтры в радиотехнике называют *квазиоптимальными.*

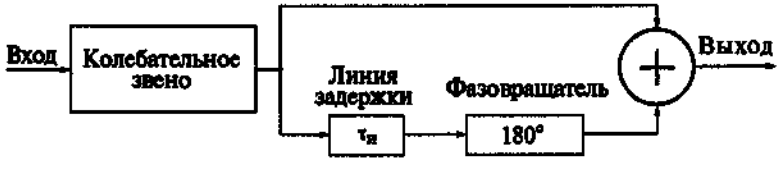



**Рис. 9.26.** Структурная схема согласованного фильтра для прямоугольного радиоимпульса



В.И. Сифоров

В практике радиоприема передаваемых сигналов можно использовать так называемые *квазиоптимальные линейные фильтры*, форма амплитудно-частотных характеристик которых заранее задана и соответствующим образом подобрана под спектр сигнала ширина полосы пропускания фильтра. Квазиоптимальные фильтры исследовались нашим соотечественником Владимиром Ивановичем Сифоровым (1904–1993), который рассматривал прохождение одиночного прямоугольного радиоимпульса через идеальный полосовой фильтр с полосой пропускания $\Delta f$ на фоне квазибелого шума. В.И. Сифоров показал, что при приеме одиночного импульса энергетический выигрыш оптимальной фильтрации по сравнению с квазиоптимальной не велик и не превышает 1 дБ.

Итак, при приеме одиночных радиоимпульсов можно ограничиться квазиоптимальной фильтрацией. Но положение существенно меняется, если надо принимать информационные импульсы (при разных видах импульсной модуляции), следующие друг за другом с таким интервалом, на котором переходные процессы в таком фильтре не успевают затухать. В этих условиях качество приема с квазиоптимальной фильтрацией падает, в то время как при использовании оптимального фильтра качество остается прежним, поскольку сигнал на его выходе концентрируется на ограниченном интервале времени и к моменту отсчета для одного импульса реакция на все предыдущие импульсы равна нулю.

### Скрытая передача сигналов при оптимальной фильтрации

Тот факт, что отношение сигнал/шум на выходе согласованного фильтра определяется лишь энергией полезного сигнала и не зависит от его формы, позволяет «замаскировать» сигнал, «спрятав» его в шумах. Действительно, если сильно уменьшить амплитуду сигнала, увеличив при этом его длительность, чтобы сохранить энергию неизменной, сигнал перестанет как визуально, так и фактически выделяться на фоне шумов. Так как энергия сигнала при этом не изменилась, отношение сигнал/шум на выходе согласованного фильтра останется прежним. Однако если передаваемый сигнал имеет простую форму, т. е. небольшую базу, такая «маскировка» во временно́й области приведет к сильному сужению спектра сигнала (принцип Гюйгенса). Так как энергия сигнала в любом случае остается неизменной, на тех частотах, где сосредоточен спектр сигнала, его спектральная плотность будет превышать (по модулю) спектральную плотность шума. Такой сигнал замаскирован во временно́й области, но легко обнаруживается в частотной области. Если одновременно с увеличением длительности ввести внутриимпульсную модуляцию, можно избежать сужения спектра и «скрыть» сигнал под шумами и во временно́й, и в частотной областях.

Скрытую передачу информации при оптимальной фильтрации в радиоприемниках достаточно легко реализуют в системах передачи информации с шумоподобными (широкополосными) сигналами.

## 9.4. Элементы теории помехоустойчивого приема

В современной теории связи используют как детерминированные модели сигналов, так и вероятностные модели передаваемых сообщений и помех (шумов). Вероятностный подход учитывает случайный характер передачи сообщений и наличие помех в канале связи и позволяет определить оптимальные приемные устройства и предельные показатели систем связи. Для того чтобы количественно оценивать эффективность работы системы связи (или радиоканала) и ее потенциальные возможности, рассмотрим основные понятия теории информации. Математическая формализация понятий об элементах информации относится к числу величайших достижений науки.

Теорию информации используют в самых разнообразных областях науки и технике: связи, радиолокации, телевидении, радиовещании и др. Поэтому понятие «информация» требует более точного определения, основанного на количественных критериях. Поскольку сведениям, поступающим с физического объекта или иного носителя информации, всегда присуща некоторая неопределенность, то данные критерии вытекают из понятий теории вероятностей, т. е. неопределенность неотъемлема от понятия вероятности. Такая взаимная обратимость понятий вероятности и неопределенности послужила основой для использования понятия вероятности при измерении степени неопределенности в теории информации. Фактически теория информации возникла из потребностей технической реализации средств связи. В них автоматически и практически без исключения действует «принцип осциллографа» — процессы, развивающиеся во времени, можно отобразить относительно пространственной координаты. Соответственно усреднение процессов по времени и по пространству эквивалентно. Для случайных процессов это известно как гипотеза эргодичности, использование которой упрощает математические доказательства ряда важных аспектов в теории передачи информации.

### Информационные параметры системы связи

Основными информационными характеристиками систем связи являются *количество информации* в сообщениях, *избыточность* сообщений, *энтропия*, *производительность* источника сообщений, *верность* и *скорость передачи* информации. Среди большого числа всевозможных технических характеристик и параметров системы связи можно выделить две основные — верность и скорость передачи сообщений. Первая характеристика определяет качество передачи, вторая — количество передаваемой информации. В современных, самых разнообразных системах электронного обмена информацией важность обеспечения безошибочной и быстрой связи является одной из приоритетных задач.

В реальной системе связи качество передачи зависит от степени искажений принятого сообщения. Эти искажения определяются свойствами и техническим состоянием системы связи, а также интенсивностью и характером помех. Необ-

ратимые искажения сообщений в грамотно спроектированной и технически исправной системе связи обусловлены лишь воздействием помех. При этом качество передачи полностью определяется помехоустойчивостью системы связи.

При передаче непрерывных сообщений верность (помехоустойчивость) определяется средней квадратической (проще, среднеквадратической) ошибкой. В качестве критерия верности систем передачи дискретных сообщений обычно используется критерий минимума средней вероятности ошибки. Заметим, что в системах передачи информации, в которых используется кодирование источника с целью сокращения избыточности, или помехоустойчивое кодирование в канале, или то и другое вместе, оптимизация системы на основе традиционного критерия минимума средней вероятности ошибки становится затруднительной.

Количественную меру верности выбирают разными способами, в зависимости от характера сообщения и требований получателя. Пусть сообщение передается в виде дискретной последовательности элементов из некоторого конечного множества. Влияние помехи на передачу такого сообщения проявляется в том, что вместо фактически переданного элемента может быть принят какой-либо другой. Подобное событие при передаче сообщения в теории информации называют *ошибкой*. В качестве количественной меры верности можно взять вероятность ошибки $P$ или любую монотонную функцию этой вероятности, например логарифмическую (что часто делают на практике).

В теории связи доказано, что верность передачи информации зависит от отношения средних мощностей сигнала и шума (отношения сигнал/шум). При данном уровне шума вероятность ошибки тем меньше, чем сильнее различаются между собой сигналы, соответствующие разным сообщениям. Задача состоит в том, чтобы для передачи различных сообщений были выбраны существенно отличающиеся сигналы. Наконец, верность передачи зависит и от способа приема. Надо применять такой способ приема, который наилучшим образом реализует различие между сигналами при данном отношении сигнал/шум.

Скорость передачи информации является важнейшим показателем работы системы связи наряду с верностью.

### Оценка количества информации, содержащейся в сообщении

Напомним, что теория информации предназначалась для описания передачи и приема сообщений в процессе деятельности человека. Во всех ее задачах присутствуют понятия передатчика и приемника, сигнала-сообщения, событий и их вероятностей. Ценность сведений, содержащихся в переданном получателю сообщении, характеризует количество заключенной в нем информации (называемое *частным количеством информации*). Для сравнения между собой разных источников сообщений, а также различных каналов связи необходимо ввести какую-то количественную меру, которая дала бы возможность объективно оценить информацию, содержащуюся в сообщении и переносимую сигналом.

Как отмечалось в главе 1, начало пониманию сущности информации было положено Н. Винером в 1948 г. Заявление Н. Винера о том, что «настоящую ин-



Р. Фишер

формацию» измерить нельзя, осталось без внимания, так как он не пояснил, что же такое настоящая информация.

Попытки количественного измерения информации предпринимались неоднократно. Первые отчетливые предложения об общих способах измерения количества информации сделал в 1921 г. математик Рональд А. Фишер (Ronald A. Fisher; 1890–1962) в процессе решения вопросов математической статистики. Проблемами хранения информации, передачи ее по каналам связи и задачами определения количества информации занимались Р. Хартли (1928 г.) и Х. Найквист (1924 г.). Р. Хартли заложил основы теории информации, определив меру количества информации для некоторых задач. Наиболее убедительно эти вопросы были разработаны и обобщены К. Шенноном в 1948 г. С этого времени началось интенсивное развитие теории информации вообще и углубленное исследование вопроса об измерении ее количества в частности.

Оценим понятие частного количества информации. Ее можно определять степенью изменения поведения получателя под воздействием принятого сообщения. В теории связи количественная оценка информации основывается на концепции выбора наиболее важного сообщения из всей совокупности возможных сообщений. При этом, чем менее вероятен выбор данного сообщения, т. е. чем более оно неожиданно для получателя, тем большее количество информации оно содержит (можно сказать, что количество информации пропорционально «удивлению» получателя от того, что оно произошло, причем «чудо», имеющее вероятность появления, близкую к нулю, являет собой количество информации, близкое к бесконечности). Очевидно обратное: достоверное (заранее известное) сообщение нет смысла передавать, поскольку оно не является неожиданным, а значит, не содержит информации. Поэтому любые реальные сообщения следует рассматривать как *случайные события* (*случайные процессы*).

Передаваемое по каналам связи сообщение можно представить в виде набора некоторых смысловых элементов или символов (например, букв алфавита). Если общее число (объем) символов алфавита равно $m$, а одно сообщение может быть составлено из $n$ элементов, то максимальное число возможных сообщений $N = m^n$. Например, с помощью двухразрядного десятичного числа ($n = 2$, $m = 10$) можно представить $N = 10^2 = 100$ различных чисел от 0 до 99. В частности, при средней длине русского слова $n = 5$ букв и алфавите в $m = 32$ буквы можно составить почти $N = 32^5 = 33{,}6$ млн различных слов. Казалось бы, искомая мера количества информации найдена. Ее можно понимать как меру неопределенности исхода опыта, если под опытом подразумевать случайный выбор какого-либо сообщения из некоторого числа возможных. Однако использование числа $N$ в качестве меры информации неудобно, так как не выполняется условие пропорциональности между длиной слова (длительностью сигнала) и количеством содержащейся в нем информации. Между тем удвоение времени передачи сообщений должно приводить к удвоению количества передаваемой информации.

Выход из положения нашел Р. Хартли, который в 1928 г. выдвинул идею, что информация допускает количественную оценку, а теорию информации разработал К. Шеннон в 1948 г. Эти идеи возникли из необходимости построения каналов связи, для определения характеристик устройств, преобразующих сигналы, выбора способов кодирования. Фактически Р. Хартли ввел в теорию передачи информации методологию «измерения количества информации». При этом Р. Хартли четко обозначил, что он имеет в виду под информацией, которую собирался измерять: «... группа физических символов — слов, точек, тире и т. п., имеющих по общему соглашению известный смысл для корреспондирующих сторон». Таким образом, Р. Хартли ставил перед собой задачу ввести какую-то меру для измерения кодированной информации, а точнее последовательности символов, используемых для кодирования информации. Рассматривая передаваемую информацию в виде определенной последовательности символов алфавита, а передачу и прием этой информации в виде последовательных выборов из этого алфавита, Р. Хартли предложил информацию $I$, приходящуюся на одно сообщение, определять логарифмом общего числа возможных сообщений $N$

$$I = \log N = n \log m. \tag{9.31}$$

Если же все множество возможных для передачи сообщений состоит только из одного ($N = m = 1$), то $I = \log 1 = 0$, что соответствует отсутствию информации в этом случае. При наличии независимых источников информации с $N_1$ и $N_2$ числом возможных сообщений

$$I = \log N = \log N_1 N_2 = \log N_1 + \log N_2,$$

т. е. количество информации, приходящееся на одно сообщение, равно сумме количеств информации, которые были бы получены от двух независимых источников, взятых порознь. Формула Хартли удовлетворяет предъявленным требованиям и ее можно использовать для измерения количества информации.

Логарифмическая мера удобна по ряду следующих причин.

1. Ее легко использовать на практике. Параметры, важные в инженерных приложениях, такие как время, пропускная способность, число переключателей и т. д., обычно меняются линейно при логарифмическом изменении числа возможных вариантов. К примеру, добавление одного переключателя удваивает число возможных состояний их группы, увеличивая на единицу его логарифм по основанию 2. Увеличение в два раза времени приводит к квадратичному росту числа сообщений, или удвоению их логарифма, и т. д.

2. Она близка к нашему интуитивному представлению о такой мере. Это тесно связано с предыдущим пунктом, так как мы интуитивно измеряем величины, линейно сравнивая их со стандартами. Так, нам кажется, что на двух одинаковых дисках памяти можно разместить в два раза больше информации, а по двум одинаковым каналам — передать ее в два раза больше.

3. Она удобна математически; многие предельные переходы просты в логарифмах, в то время как в терминах числа вариантов они достаточно сложны.

Выбор основания логарифма безусловно соответствует выбору единицы измерения количества информации. Если взять основание 2, то полученные единицы есть не что иное, как «двоичные цифры», или биты. Устройство с двумя устойчивыми состояниями, такое как переключатель или триггер, способно хранить один бит информации, $N$ таких устройств — $N$ бит, так как полное число состояний $2^N$ и $\log_2 2^N = N$. При использовании же основания 10 единицы называют *десятичными цифрами*. При этом один разряд десятичного кода содержит $I = -\log_2 p = 3{,}32$ бит информации.

Рассмотрим подробнее вопрос о количественной оценке информации, предложенной Р. Хартли. Выберем некоторое случайное событие, относительно которого известно, что оно может появиться с вероятностью $p$. Пусть информация об этом событии передается алфавитом, состоящего из $m$ смысловых символов. Если возможность появления любого символа алфавита также равновероятна, то эта вероятность $p = 1/m$ (при этом $m = 1/p$). Полагая, что $N = m$, получим

$$I = \log N = \log m = \log(1/p) = -\log p, \tag{9.32}$$

т. е. количество информации на каждый равновероятный сигнал равно минус логарифму вероятности отдельного сигнала.

Основание логарифма в (9.32) может быть выбрано произвольным, поскольку это влияет лишь на единицу измерения количества информации. Хартли предложил вычислять количество информации, содержащейся в передаваемом сообщении, по более удобной формуле

$$I = -\log_a p, \tag{9.33}$$

где логарифм может быть взят при любом основании $a$.

Данная формула позволяет для некоторых специфических случаев определить количество информации. Однако для практических целей необходимо задаться единицей его измерения.

Такой подход принципиально изменил понятие информации. Под информацией теперь стали понимать не любые сообщения, передаваемые по системе связи, а лишь те, которые уменьшают неопределенность у получателя информации. Тогда в простейшем случае наличия неопределенности выбор будет производиться между двумя взаимоисключающими друг друга равновероятными сообщениями, например между двумя качественными признаками: положительным и отрицательным импульсами, импульсом и паузой и т. п. Количество информации, переданное в этом простейшем случае, наиболее удобно и следует принять за единицу количества информации. Именно такое количество информации может быть получено, если применить формулу (9.33) и взять логарифм по основанию 2. При этом

$$I = -\log_2 p. \tag{9.34}$$

Тогда

$$I = -\log_2 p = -\log_2(1/2) = \log_2 2 = 1.$$

Полученная единица количества информации является битом (напомним, что термин «бит» имеет два различных значения; одно используется в качестве синонима двоичного символа, а второе обозначает единицу количества информации; скорость передачи данных представляет собой именно количество двоичных символов, передаваемых за одну секунду). Здесь бит является не только единицей количества информации, но и единицей измерения степени неопределенности. При этом имеется в виду неопределенность, которая содержится в одном опыте, имеющем два равновероятных исхода. Данная мера представления информации является универсальной и позволяет сравнить различные сообщения и количественно определить ценность различных источников информации, оценить величину ее потерь при передаче, приеме, обработке, хранении, использовании и т. д. Применение двоичных логарифмов диктуется тем, что сообщение в практических системах связи чаще всего принимает форму отдельных групп (кодовых слов), состоящих только из двух символов, которые можно трактовать как 0 или 1. Каждая такая группа кодирует, например, буквы того или иного естественного языка, из которых составляются отдельные слова.

Если основанием логарифма в формуле (9.34) выбрано $e$ (как в натуральных логарифмах), то информация будет измеряться в натуральных единицах или в *натах* (1 нат ≈ 1,443 бит).

**Пример 9.3.** На одной стандартной странице на русском языке написан текст, содержащий 40 строк по 65 букв в каждой строке. Оценить по формуле Хартли объем информации в данном тексте, полагая, что алфавит состоит из 32 букв.

Р е ш е н и е. Для упрощения будем считать, что появление любой буквы в тексте равновероятно. В этом случае каждая буква содержит количество информации

$$I_б = -\log_2(1/32) = 5 \text{ бит}.$$

Тогда общий объем информации страницы текста составит

$$I = 40 \cdot 65 \cdot 5 = 13\,000 \text{ бит} = 13 \text{ Кбит}.$$

Заметим, что данный расчет является лишь ориентировочным, поскольку не учитывается разница между вероятностями появления различных букв в тексте. На практике же буквы «м» или «е» встречаются гораздо чаще, нежели буквы «щ» или «э». Кроме того, при расчетах не учитывался факт сильной корреляции между отдельными буквами алфавита (в частности, в русском языке сочетание букв «ка» встречаются значительно чаще, чем сочетание «зз»).

**Пример 9.4.** Угадать по формуле Хартли одно число из набора чисел от единицы до ста.

Р е ш е н и е. Используя формулу (9.34), вычислим, какое количество информации для этого требуется: $I = \log_2 100 > 6{,}644$. Таким образом, сообщение о верно угаданном числе содержит количество информации, приблизительно равное 6,644 единицы информации.

**Энтропия источника сообщений.** Подход Р. Хартли основан по существу на теории множеств и комбинаторике, а также нескольких интуитивно ясных и вполне очевидных предположениях. Р. Хартли понимал, что сообщения имеют различную вероятность и, следовательно, неожиданность их появления для получателя неодинакова. Но, определяя количество информации, он пытался полностью исключить фактор «неожиданности». Поэтому формула Хартли (9.33) позволяет определить количество информации в сообщении только для случая, когда появление символов равновероятно и они статистически независимы. На практике эти условия выполняются редко. При определении количества информации необходимо учитывать не только количество разнообразных сообщений, которые можно получить от источника, но и вероятность их получения.

Основной недостаток формулы Хартли — она не отражает случайного характера формирования сообщений. Чтобы устранить этот недостаток, необходимо связать количество информации в сообщениях с вероятностью появления символов. Эта задача была решена К. Шенноном в 1948 г., который применил теоретико-вероятностный подход. Это связано с тем, что исторически теория информации Шеннона выросла из потребностей теории связи, имеющей дело со статистическими характеристиками передаваемых сообщений и каналов связи.

Развивая и обобщая идеи Р. Хартли, К. Шеннон в 1948 г. в фундаментальном труде «Математическая теория связи» ввел два важнейших понятия: *информации, содержащейся в подлежащих передаче по каналу связи сообщениях*, и *энтропии источника сообщений*. Предположение Шеннона состояло в том, что он впервые стал рассматривать статистическую структуру передаваемых сообщений и действующих в канале шумов и, кроме того, он рассматривал не только конечные, но и непрерывные множества сообщений. Созданная им теория информации дала ключ к решению двух основных проблем теории связи: устранение избыточности сообщений и кодирование сообщений, передаваемых по каналу связи с шумами. Одной из задач, которую ставил перед собой Шеннон, заключалась в том, чтобы определить систему кодирования, позволяющую оптимизировать скорость и достоверность передачи информации. К. Шеннон, используя методологию Р. Хартли, обратил внимание на то, что при передаче словесных сообщений частота использования различных букв алфавита не одинакова: некоторые буквы используются очень часто, другие — редко. Существует и определенная корреляция в буквенных последовательностях, когда за появлением одной из букв с большой вероятностью следует конкретная другая.

Пусть объем (полное число символов) какого-то алфавита $X$ равно $m$ и источник передает сообщения этими символами. Положим, что величины всех вероятностей появления любого из символов $p_i$ ($i = 1, 2, ..., m$) — априорные, т. е. известные. Проследим за достаточно длинным отрезком сообщения. Пусть в нем имеется $N_1$ символов первого типа, $N_2$ символов второго типа, ..., $N_m$ сигналов $m$-го типа, причем $N_1 + N_2 + ... + N_i + ... + N_m = N$ — общее достаточно большое число символов в наблюдаемом отрезке, $f_1, f_2, ... f_i, ..., f_m$ — частоты появления соответствующих символов.

При возрастании длины отрезка сообщения каждая из частот появления $i$-го символа стремится к фиксированному пределу, т. е.

$$\lim f_i = p_i, (i = 1, 2, ..., m),$$

где $p_i$ можно считать вероятностью данного символа.

Предположим, получен символ $i$-го типа с вероятностью $p_i$, содержащий, согласно (9.32), $-\log p_i$ единиц информации. Очевидно, что в рассматриваемом отрезке сообщения $i$-й символ встретится примерно $Np_i$ раз, и общая информация, доставленная символами этого типа, будет равна произведению $Np_i \log p_i$. То же относится к символам любого другого типа, поэтому полное количество информации (для нее К. Шеннон ввел метрику $H$), доставленное отрезком из $N$ символов алфавита $X$, будет примерно равно

$$H(X) = -N\sum_{i=1}^{m} p_i \log_2 p_i.$$

Чтобы определить среднее количество информации, приходящееся на один символ передаваемого сообщения, т. е. удельную информативность источника, нужно это число разделить на $N$. При неограниченном росте приблизительное равенство перейдет в точное. В результате получим асимптотическое соотношение — формулу Шеннона

$$H(X) = -\sum_{i=1}^{m} p_i \log_2 p_i. \qquad (9.35)$$

Оказалось, что формула (9.33), предложенная Хартли, представляет собой частный случай более общей формулы Шеннона. Если в формуле Шеннона принять, что $p_1 = p_2 = ... = p_i = ... = p_N = 1/N$, то приходим к формуле Хартли

$$H(X) = -\sum_{i=1}^{N} \frac{1}{N} \log_2 \frac{1}{N} = -\log_2 \frac{1}{N} = \log_2 N. \qquad (9.36)$$

Выражение (9.35), повторяющее по форме выражение для энтропии в статистической механике, К. Шеннон по аналогии назвал *энтропией источника сообщений*. Таким образом, энтропия отражает среднее значение количества информации, приходящееся на один символ алфавита (бит/символ).

Знак минус в формулах Шеннона (9.35) и (9.36) не означает, что энтропия сообщения — отрицательная величина. Объясняется это тем, что вероятность $p$, согласно определению, меньше единицы, но больше нуля. Так как логарифм числа, меньшего единицы, т. е. $\log_2 p_i$ — величина отрицательная, то произведение вероятности на логарифм числа будет положительным.

Как уже отмечалось, именно такой подход принципиально изменил понятие информации. Под информацией теперь понимают не любые сообщения, передаваемые в системе связи, а лишь те, которые снижают энтропию сообщения. При этом увеличивается информативность поступившего сообщения.

Энтропия — это тот минимум информации, который необходимо получить, чтобы ликвидировать неопределенность алфавита, используемого источником информации. Энтропию можно рассматривать как числовую характеристику закона распределения, выражающую неопределенность, которая присуща элементам алфавита, выбранного для передачи сообщений. Как правило, алфавит с небольшой энтропией мало пригоден для практического использования.

Количество информации в сообщении с позиции энтропии определяется уменьшением неопределенности состояния некоторого процесса. В отношении сигнала, несущего информацию, неопределенность выражается неизвестностью его информационных параметров. Пока сигнал не принят и не определены его информационные параметры, о содержании сообщения можно только догадываться с некоторой вероятностью правдоподобия. После приема сигнала неопределенность в содержании сообщения значительно уменьшается. Если есть гарантия, что при передаче сообщения не возникло искажений сигнала, то неопределенность вообще исчезает. Однако имеется всегда, хотя и малая, вероятность ошибки, так как без искажений вообще сигнал не может быть передан. Поэтому некоторая неопределенность все-таки остается.

**Пример 9.5.** Определим по Хартли и Шеннону количество информации, связанное с появлением каждого символа в сообщениях, записанных на русском языке. Будем считать, что русский алфавит состоит из 32 букв.

Р е ш е н и е. По формуле Хартли: $I = \log_2 32 = 5$ бит (здесь считаем, что появление каждой буквы равновероятно). По формуле Шеннона (для неравновероятных исходов) это значение равно $H = 4{,}72$ бит.

Прежде чем пояснить наглядный смысл понятия энтропии, опишем ее основные свойства.

1. Энтропия является положительной непрерывной величиной.

Пусть сообщение передается с помощью двоичного алфавита (двоичного источника), состоящего только из двух символов, например 0 и 1. Если вероятность появления первого символа равна $p$, то вероятность появления второго символа составляет $1 - p$. В этом случае энтропия изменяется от 0 до 1 и

$$H(X) = -[p\log_2 p + (1-p)\log_2 p(1-p)]. \quad (9.37)$$

На рис. 9.27 приведен график зависимости $H(X)$ для двоичного источника, построенный по формуле (9.37). Он показывает, что максимальное значение энтропии, равное 1 бит/символ, достигается при равновероятном появлении обеих букв алфавита. Если же вероятность $p$ приближается к нулю или к единице,

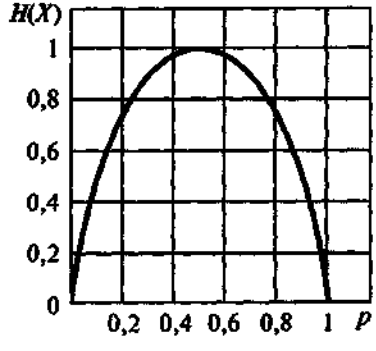



**Рис. 9.27.** Энтропия двоичного источника

энтропия источника становится весьма низкой, что говорит о малой информативности выбранного алфавита. Сообщение превращается, по существу, в детерминированную последовательность символов.

2. Энтропия равна нулю тогда и только тогда, когда вероятность одного из состояний равна единице. Это соответствует случаю, когда одно из состояний источника сообщения достоверно, а другие невозможны т. е. отсутствует какая-либо неопределенность.

3. Энтропия максимальна, когда все символы источника сообщений равновероятны, а значит, $p_1 = p_2 = \ldots = p_m = 1/m$. В этом случае имеем формулу Хартли

$$H(X) = \log_2 m = I. \qquad (9.37)$$

Можно еще добавить уже отмечавшееся свойство: энтропия обладает свойством аддитивности, т. е. энтропии независимых систем можно складывать.

Дадим пояснения этим свойствам. Если ситуация при передаче информации полностью ясна, то никакой неопределенности нет, и энтропия равна нулю. В частности, если ток в цепи равен 20 А, то он не может быть одновременно равным 10 А. На данном примере можно пояснить и другое свойство. Если одно из событий ожидается с очень малой вероятностью, например $p_1 = 0{,}01$, а другое с высокой, например $p_2 = 0{,}99$, то неопределенность невелика, так как почти наверняка получим второе сообщение. Если события равновероятны и $p_1 = p_2 = 0{,}5$, то уже нет уверенности, что будет получено какое-то из сообщений, т. е. неопределенность возрастает. Очевидно, что неопределенность возрастает, если вместо одного из двух сообщений может прийти одно из трех, четырех и более.

Теперь с помощью свойств 1 – 3 можно дать другое общепринятое определение энтропии и наглядно пояснить смысл этого понятия. Энтропия — это средняя информативность источника на один символ, определяющая «непредсказуемость» или «неожиданность» выдаваемых им сообщений. Полностью детерминированный источник, вырабатывающий лишь одну, заранее известную последовательность символов, обладает нулевой информативностью. И наоборот, наиболее «хаотический» источник, выдающий взаимно независимые и равновероятные символы, обладает максимальной информативностью.

Для большей доступности понятия энтропии приведем известный пример с обезьяной, сидящей за пишущей машинкой (в более современном варианте — за клавиатурой компьютера). Если она обучена ударять по клавишам, но, очевидно, не знает грамоты, то «обезьяний» текст окажется примером текста с взаимно независимыми и равновероятными символами. Поэтому он будет обладать наибольшей энтропией, превосходящей энтропию осмысленного текста на каком-либо языке. Несмотря на бесполезность обезьяньего текста передавать его по каналам связи сложнее, чем какой-либо смысловой текст.

**Пример 9.6.** Вычислить энтропию источника, который может выдавать четыре символа с равной вероятностью появления.

**Решение.** Используя формулу (9.30), получаем $H(X) = \log_2 m = \log_2 4 = 2$ бит.

### Оценка пропускной способности канала связи с шумами

Напомним, что предельную скорость передачи информации оценивают *пропускной способностью* (*емкостью*) *канала связи*, численно равную максимальному количеству информации, которое можно передать по каналу за 1 с (в цифровых системах связи, бит/с). Интуитивно нетрудно понять, какими свойствами должна обладать система радиосвязи (или радиоканал), чтобы с высокой эффективностью передавать сообщения от источника к потребителю. Во-первых, такая система радиосвязи должна обеспечивать требуемую скорость передачи информации. Во-вторых, необходимо гарантировать высокую верность приема символов, либо свести вероятность возможных ошибок к приемлемо низкому уровню.

Если теоретически предположить, что в канале связи отсутствуют любые шумы, то задача о пропускной способности теряет смысл — любой объем информации можно было бы передать достаточно быстро и с высокой верностью. Взяв, например, единственный отсчет сигнала и записав его двоичным числом (кодовым словом) достаточной длины, можно с помощью этого числа закодировать сколь угодно длинное сообщение. Реальное наличие в канале связи шумов различного происхождения резко усложняет проблему. Интенсивность информационного потока жестко связана с шириной полосы частот, занимаемых каналом, а также значением отношения сигнал/шум на входе приемника. Совершенно очевиден факт, что канал связи тем совершеннее, чем шире полоса частот и чем больше отношение сигнал/шум.

Количественные оценки, дающие возможность решить данную проблему, были получены впервые К. Шеннoном в его классической работе по теории связи. Следует отметить такой редко упоминаемый ныне факт. В 1939 г. Д.В. Агеев защитил кандидатскую диссертацию на тему «Линейные методы селекции и проблема пропускной способности эфира», в которой были выявлены и исследованы все возможные принципы разделения сигналов линейными методами: частотный, временно́й и по форме. Эти принципы явились основой теории помехоустойчивого приема сигналов, и в частности, по определению предельной пропускной способности частотного канала связи. По существу в данной работе за 10 лет до К. Шеннона была определена предельная пропускная способность канала связи. В основу теории была положена физически реализуемая модель сигналов конечной длительности, впервые применено представление сигналов и помех в многомерном функциональном пространстве.

Упрощенно рассуждения К. Шеннона сводились к следующему. Пусть имеется некоторый канал связи с шириной полосы пропускания $F_к$ [Гц]. Положим, что $P_с$ — средняя мощность полезного сигнала на входе приемника. В канале связи на входе приемника присутствует белый гауссовский шум, мощность которого составляет $P_ш$ (напомним, что возможно ее обозначение как $S$ и $N$). Для надежного различения двух сигналов необходимо, чтобы данные, получаемые при взятии отсчетов, отличались друг от друга на достаточное значение, например не менее чем на среднеквадратическое отклонение, обусловленное шумом. Поскольку дисперсии стационарных случайных процессов прямо пропорцио-

нальны их мощностям, то приходим к выводу, что число полностью «различимых» значений при взятии каждого отсчета составит

$$\sqrt{\frac{P_c + P_ш}{P_ш}} = \sqrt{1 + \frac{P_c}{P_ш}}.$$

Пусть длительность передаваемой информации $T_c$ [с]. По теореме отсчетов при этом берется не менее $2T_cF_к$ отсчетных значений. Тогда общее число различных сигналов, которые могут быть построены описанным способом,

$$M = \left(\sqrt{1 + \frac{P_c}{P_ш}}\right)^{2T_cF_к}. \qquad (9.38)$$

Поскольку все сигналы равновероятны, то вероятность выбора одного конкретного сигнала равна $p_1 = 1/M$. Очевидно, что число битов информации, которое можно безошибочно передать за время $T_c$, согласно (9.34), составит

$$I = -\log_2 p_1 = -\log_2(1/M) = \log_2 M. \qquad (9.39)$$

Максимальная скорость передачи информации [бит/с], достигаемая при этом, и является *пропускной способностью канала*

$$C = \frac{I}{T_c} = \frac{\log_2 M}{T_c} = F_к \log_2\left(1 + \frac{P_c}{P_ш}\right). \qquad (9.40)$$

Это соотношение известно как *формула Шеннона* для пропускной способности канала аналоговой системы связи с ограниченной полосой частот и ограниченной средней мощностью сигнала при наличии белого гауссовского шума. Для стандартного телефонного канала $F_к$ = 3 кГц, $P_c/P_ш$ = 30 дБ, значит, теоретический предел для телефонной сети равен около 30 Кбит/с. Ослабление в телефонных скрученных парах составляет 15 дБ/км, дополнительные ограничения возникают из-за перекрестных наводок. Стандартные проводные линии связи имеют ослабление 6 дБ/км на частоте 800 Гц, или 10 дБ/км на частоте 1 600 Гц.

Формула Шеннона является важным инструментом при проектировании разных каналов связи. Анализ формулы (9.40) показывает, что существуют два пути повышения пропускной способности канала связи: за счет *расширения полосы частот канала* $F_к$ и *повышения отношения сигнал/шум* $P_c/P_ш$. Однако эти пути неравноценны. Если $F_к$ = const, а $P_c/P_ш$ возрастает, то пропускная способность $C$ также увеличивается, но ее рост оказывается весьма медленным, поскольку он подчиняется логарифмическому закону. При фиксированном значении отношения сигнал/шум и увеличении полосы пропускания канала $F_к$ пропускная способность возрастает линейно, однако можно показать, что в полосе пропускания не заключены неограниченные возможности увеличения $C$. Пропускная способность $C = 0$ только при $P_c = 0$, т. е. непрерывный канал обеспечивает передачу информации даже в том случае, если уровень шумов превышают уровень сигнала ($P_c/P_ш < 0$), это используют для скрытой передачи.

Понятие пропускной способности для канала аналоговой системы связи позволяет судить о его потенциальных возможностях по передаче информации, однако для того, чтобы определить их в более конкретных терминах верности и скорости передачи, необходимо обратиться к теореме кодирования Шеннона для непрерывного канала связи.

Приведем характеристики некоторых реальных каналов связи (табл. 1).

**Таблица 1. Характеристики некоторых каналов связи**

| Вид связи | $F_m$, Гц | $P_c/P_ш$ | $C$, бит/с |
|---|---|---|---|
| Телеграф | 120 | $2^6$ | 640 |
| Телефон | $3\cdot10^3$ | $2^{17}$ | $5\cdot10^4$ |
| Телевидение | $7\cdot10^6$ | $2^{17}$ | $130\cdot10^6$ |
| Компьютерная сеть | — | — | До $10^9$ |
| Слух человека | $20\cdot10^3$ | | $5\cdot10^4$ |
| Глаза человека | | | $5\cdot10^6$ |

Из сопоставления данных табл. 1 видно, что пропускная способность телефонного канала связи почти совпадает с пропускной способностью органов слуха человека. Однако она значительно выше скорости обработки информации человеком, которая составляет не более 50 бит/с. Другими словами, человеческие каналы связи допускают значительную избыточность информации, поступающей в мозг.

## 9.5. Кодирование сообщений в системах связи

Далее будут описаны основные подходы к способам кодирования источников сообщений, так называемого *канального кодирования* (*channel coding*). Причем в отличие от математической теории кодирования здесь рассмотрены не столько детали доказательств различных свойств кодов, сколько демонстрация скоростных, энергетических или спектральных выигрышей, которые обеспечивают данные методы по сравнению с «некодированной» передачей сообщений.

Сообщение может иметь и аналоговую и дискретную форму. Простейшим примером дискретного сообщения является текст. Любой текст состоит из *алфавита источника сообщения*, который представляет собой определенную совокупность элементов: букв, цифр, знаков препинания. Поскольку число элементов в алфавите конечно, то их можно пронумеровать, т. е. закодировать, и тогда передача сообщения сведется к передаче последовательности чисел.

Напомним, что *кодирование сигнала* — его представление в определенной форме, удобной или пригодной для передачи сообщений. Говоря строже, это правило, описывающее отображение одного набора знаков (символов) в другой набор знаков набора знаков (символов). В процессе кодирования элементы передаваемого сообщения преобразуются в соответствующие кодовые символы. В

частности, для передачи 32 основных букв русского алфавита необходимо передать числа от 0 до 31. Чтобы передать любое число, записанное в десятичной форме, требуется передача десяти цифр от 0 до 9 (для данного случая нужны десять элементарных сигналов). Реализация кодирования на передающей стороне всегда предполагает применение обратной процедуры — *декодирования* — для восстановления сообщения по принимаемым кодовым символам.

**Код** — полная совокупность условных символов, которую применяют для кодирования сообщений. Взаимосвязь символов (или комбинаций символов, если кодируются не отдельные символы) исходного алфавита с их кодовыми комбинациями составляет таблицу соответствия (или таблицу кодов). Множество возможных кодовых символов (элементарных сигналов) называют *кодовым алфавитом*, а их число $m$ — *основанием кода*. Код с основанием $m = 2$ — *бинарный*, с другими основаниями — *многопозиционный*. При основании кода $m$ правила кодирования $K$ элементов сообщения сводятся к правилам записи $K$ различных чисел в $m$-ичной системе счисления.

Простейшими электрическими сигналами, образующими символы при кодировании сообщений, являются импульсы тока или напряжения. Код принимаемого языка определяется алфавитом символов и правилами их комбинирования, т. е. правилами построения знаков различного ранга.

При кодировании каждый элемент сообщения записывают определенной совокупностью кодовых символов, которую называют *кодовой комбинацией*. Число разрядов $n$, образующих кодовую комбинацию, называют *разрядностью кода* или *длиной кодовой комбинации*.

Чтобы закодировать сообщение источника, необходимо установить определенные правила. Совокупность правил кодового обозначения объектов называют *системой кодирования*. Такую систему можно выразить различным образом. Часто очень удобной системой кодирования является кодовая таблица, в которой приводятся алфавит кодируемых сообщений и соответствующие им кодовые комбинации.

Наиболее широко при кодировании дискретных сообщений применяют двоичный код, имеющий в технике особое значение. Это объясняется тем, что наличие и отсутствие посылки наиболее уверенно различимы. Кроме того, очень просто практически реализовать устройство, имеющее два возможных состояния (открыто – закрыто). В частности, устройство с двумя устойчивыми состояниями, например электронный переключатель или триггер, способно хранить один бит информации, $N$ таких устройств — $N$ бит. Поэтому двоичный код находит преимущественное применение не только в системах связи, но и в компьютерах, автоматике и т. д.

Необходимое количество разрядов для кодирования сообщения при заданном максимальном числе уровней шкалы квантования $U_{max}$ определятся из соотношения $n = \log_2 N_{max}$. Если кодовая группа содержит $n$ символов 0 и 1, то с помощью подобного $n$-разрядного двоичного кода можно закодировать числа до $N_{max} = 2^n$. Так, при $n = 7$ $N_{max} = 128$, при $n = 8$ $N_{max} = 256$ и т. д.

Если все кодовые слова имеют одинаковое число символов, то код называют *равномерным*, в противном случае — *неравномерным*. Длину кодовой комбинации определяют числом входящих в нее единичных элементов. Равномерный код при $n = 3$ называют трехбитным, при $n = 5$ — пятибитным или пятиэлементным кодом, при $n = 7$ — семибитным кодом и т. д. Для равномерного кода общее число различных кодовых комбинаций $N = m^n$.

Код Бодо ($m = 2$, $n = 5$) — это код, элементами которого являются посылка и отсутствие посылки одинаковой длительности и абсолютного значения, но разной полярности. Код Бодо равномерный, каждой букве соответствует пять элементарных знаков. Всего можно составить $N = 2^5 = 32$ комбинаций. Для увеличения числа знаков в аппарате Бодо применяют второй регистр с 32 знаками.

Базовый и известный пример неравномерных кодов — код Морзе, у которого кодовые слова имеют разную длину: наиболее часто встречающимся знакам присваиваются наименее короткие и наоборот. В коде Морзе символы 1 и 0 используются только в двух сочетаниях — как одиночные (1 и 0) или как тройные (111 и 000). Сигнал, отражаемый одной единицей (короткой посылкой тока), соответствует точке (.), трем единицам (втрое более длинной посылкой тока) — тире (нуль тока или пауза). Символ 0 используется как знак, отделяющий точку от тире, точку от точки и тире от тире. Совокупность знаков 000 используется как разделительный знак между кодовыми комбинациями. Помимо этого совокупность знаков 00000 используется для разделения слов. Код Морзе основан на статистике английского языка. Заметим, что код Морзе — это кодирование не с двоичным входным алфавитом (точка и тире), как кажется на первый взгляд, а с троичным алфавитом: точка, тире и пробел. При радиосвязи кодом Морзе передатчик излучает радиоимпульсы разной длины, соответствующие точкам и тире кодовых комбинаций. В приемнике с помощью специального генератора и смесителя радиоимпульсы преобразуют в отрезки колебаний звуковой частоты. Меняя частоту колебаний генератора, оператор имеет возможность выбрать наиболее благоприятную для себя частоту звуковых колебаний. Обычно она устанавливается равной 800...1000 Гц. Скорость передачи знаков при слуховой телеграфной связи зависит от квалификации оператора и обычно составляет около 100 знаков в минуту. К достоинствам кода Морзе следует отнести высокую помехоустойчивость приема на слух, благодаря чему он находит широкое применение в радиосвязи. При приеме на слух преобразование сигнала в сообщение осуществляет человек, который, обладая определенным опытом, может различать сигналы, значительно пораженные помехами.

**История кодирования.** Зародилось кодирование в давние времена и использовалось как для представления информации в символическом виде, так и для шифрования сообщений и тайнописи.

В теории информации доказывается, что как для неортогональных, так и для ортогональных сигналов значение вероятности ошибки $p$ при оптимальном приеме однозначно определяется числом символов $m$ и параметром отношения сигнал/шум $Q_{вых0} = Э_0/W_0$, где $Э_0$ — энергия элемента сигнала; $W_0$ — спектраль-

ная плотность мощности белого шума. Отсюда, казалось бы, можно однозначно сделать вывод, что неограниченное повышение верности передачи сообщений (т. е. убывание вероятности ошибки $p$ к нулю) может быть получено лишь за счет неограниченного возрастания энергетического параметра $Q_{вых0}$ (путем увеличения мощности сигнала). В частности, если заданы средняя мощность сигнала $P_c$ и спектральная плотность мощности белого шума $W_0$, то это может происходить только при стремлении к нулю скорости передачи символов $v$, поскольку $Э_0 = P_c T_c$, а длительность сигнала $T_c = 1/v$.

Первым человеком, «ошеломившим» специалистов по теории связи возможностью абсолютно надежно передавать информацию по каналам связи с помехами не за счет увеличения мощности сигнала или уменьшения скорости передачи, а за счет усложнения методов модуляции-демодуляции и введения кодирования сообщений, был К. Шеннон. Фактически выход статьи «Математическая теория связи» в 1948 г. считают началом помехоустойчивого кодирования. К. Шеннон доказал принципиальную возможность безошибочной передачи сигналов, *если скорость передачи меньше пропускной способности канала связи*, которая тем больше, чем выше отношение сигнал/шум на входе приемного устройства. Это указывало на то, что энергетика линий связи определяет только их пропускную способность, а сколь угодно высокой помехоустойчивости приема сообщений можно достичь применением специальным образом построенных кодов. Революционные для того времени идеи Шеннона осуществили переворот в сознании связистов, ведь до создания теории информации считалось само собой разумеющимся, что единственные возможности повышения помехоустойчивости приема сигналов состоят в увеличении мощности передатчика или в многократной передаче по каналу связи одного и того же сообщения. Эти способы приводят к весьма низкой эффективности использования пропускной способности канала связи. Фактически Шеннон утверждал, что *мощность сигнала, шум в канале и полоса частот ограничивают лишь скорость передачи*, а не ее точность. Шеннон нашел необходимые и достаточные условия убывания вероятности ошибки до нуля. Кроме того, он сделал это для различных моделей каналов связи, в частности с ограниченной полосой частот, для фиксированного числа форм сигналов, для источников с неравновероятными или с зависимыми символами (называемых *избыточными источниками*), для источников непрерывных сообщений, для критериев верности приема дискретных сообщений, отличных от вероятности ошибки символа $p$, и т. д. Это и составило сущность новой теории информации и теории кодирования. Однако К. Шеннон не указал, как найти подходящие коды, а лишь доказал их существование.

### Классификация основных методов кодирования и кодов

Применяемые в технике связи методы кодирования и коды можно классифицировать по ряду специфических признаков. Приведенная классификация не исчерпывает всего изобилия методов кодирования и кодов, но отражает все те коды, которые являются предметом дальнейшего изложения (рис. 9.28).




**Рис. 9.28.** Классификация основных методов кодирования и кодов

По назначению кодирование (как и соответствующие коды) принято разделять на примитивное, экономное и помехоустойчивое.

С помощью кодирования решают следующие задачи:

- согласование алфавита источника сообщений с алфавитом канала — примитивное (первичное, простое, безызбыточное) кодирование;
- «сжатие» информации (уменьшение или полное устранение избыточности, содержащейся в сообщении) — экономное (эффективное) кодирование;
- обнаружение (*detection*) или исправление (*correction*) ошибок, возникающих в канале связи из-за помех и искажений сигнала, — помехоустойчивое (корректирующее, избыточное) кодирование. Помехоустойчивое кодирование часто называют *канальным кодированием*.

### Примитивное кодирование

*Примитивное* (*primitive*), или *безызбыточное*, *кодирование* согласует алфавит источника с алфавитом дискретного канала, а в ряде случаев используется для шифрования передаваемой информации, для ее защиты от несанкционированного доступа, а также для повышения устойчивости работы устройств синхронизации систем связи. Термин «примитивное» — теоретико-числовое понятие, требующее алгебраического рассмотрения.

Отличительное свойство примитивного кодирования состоит в том, что избыточность дискретного источника, образованного выходом примитивного кодера, равна избыточности источника на входе кодера. Упрощенно можно сказать, что при примитивном кодировании символ одного алфавита заменяют на символ другого алфавита.

Шифрование (при примитивном кодировании), которое используют для обеспечения секретности связи, предотвращает понимание сообщения несанкционированным пользователем и введение в систему ложных сообщений. В этом случае правило примитивного кодирования выбирается так, чтобы вероятность появления на выходе кодера длинной последовательности, состоящей только из единиц или только из нулей, была минимальной. Подобный кодер называют *скремблером* (от англ. *scramble* — перемешивать). Скремблирование имеет цель придать потоку символов свойства, похожие на случайную последовательность равновероятных и независимых двоичных символов. Скремблирование осуществляют пропуская поток символов через регистр сдвига с обратными связями, в котором эти символы «перемешиваются» и преобразуются. Разумеется, характеристика скремблера должна быть обратимой. Тогда принятые после демодулятора символы подвергаются обратной операции — дескремблированию для восстановления исходной последовательности. Скремблирование часто применяют для повышения надежности системы синхронизации. Но оно также улучшает работу декодера, особенно в системах с обратной связью. Коды, у которых все возможные кодовые комбинации используют для передачи информации, называют *примитивными* или *кодами без избыточности*. В простых равномерных кодах превращение одного символа комбинации в другой, например 0 в 1 или 1 в 0, приводит к появлению новой *разрешенной комбинации*, т. е. может привести к ошибке в принятом сообщении.

### Экономное кодирование

*Экономное* (иначе эффективное, статистическое) *кодирование*, или *сжатие данных*, используют для повышения скорости передачи информации и приближения ее к пропускной способности канала связи. Отличительное свойство экономного кодирования состоит в том, что избыточность источника, образованного кодером, меньше, чем избыточность источника на входе кодера. Наибольшее применение экономное кодирование находит в компьютерах при архивации информации. Так, последние версии операционных систем содержат в своем составе программы сжатия данных (динамические компрессоры и архиваторы), а новые стандарты на модемы для связи между компьютерами по телефонным сетям общего пользования включают сжатие в процедуры обработки данных.

Теоретической основой построения экономных кодов является первая теорема кодирования Шеннона о передаче информации, доказывающая (приведена в упрощенном виде), что *для канала связи без помех всегда можно создать систему экономного кодирования дискретных сообщений, у которой среднее количество двоичных кодовых сигналов на один символ сообщения будет приближаться как угодно близко к энтропии источника сообщений*. Используя понятие избыточности кода, можно дать более короткую формулировку теоремы: *при отсутствии помех передачи всегда возможен такой вариант кодирования сообщения, при котором избыточность кода будет сколь угодно близкой к нулю*. Экономное кодирование основывается на использовании неравномерных кодов.

При этом более короткие кодовые комбинации используются для передачи более вероятных символов, более длинные — для передачи менее вероятных символов. Экономный код должен обладать следующими свойствами:

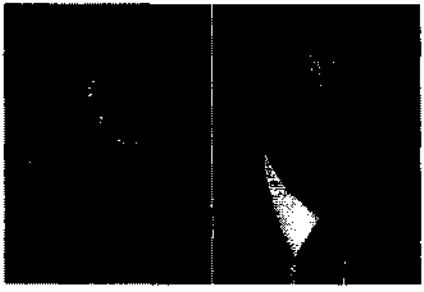
- на передачу одного символа должно затрачиваться минимальное число нулей и единиц двоичной кодовой последовательности;
- код должен обеспечивать однозначное декодирование, т. е. по принятой кодовой последовательности восстанавливать переданный символ.

Для обеспечения однозначности декодирования эффективный код должен удовлетворять требованию: никакое более короткое слово (последовательность) эффективного кода не должно являться началом другого более длинного слова.

Рассмотрим экономное кодирование источников с известной статистикой сообщений. Очевидно, что средняя длина неравномерного кода будет минимизироваться тогда, когда с более вероятными сообщениями источника будут сопоставляться более короткие комбинации канальных символов. Проблема, однако, состоит в том, что у неравномерного кода на приемной стороне оказываются неизвестными границы этих комбинаций. Если же мы попытаемся их выделить, используя известный способ кодирования, то декодирование может оказаться неоднозначным (действительно, если, например, с буквой А сопоставлена комбинация 0, с буквой Б — 1, а с буквой В — 10, то невозможно определить по принятой комбинации 10, передавались ли буква В или пара букв А и Б). Для того чтобы используемый код обладал свойством однозначной декодируемости, он, очевидно, должен удовлетворять некоторым условиям. Однозначное декодирование будет обеспечено, если ни одно кодовое слово не является началом другого кодового слова. Такие коды называются *префиксными*. Неравномерное префиксное кодирование устраняет избыточность источника, вызванную неодинаковой вероятностью сообщений.

Экономное кодирование сообщений источника без памяти обеспечивается неравномерным префиксным кодом, который строится оптимальным образом методом Хаффмана, а приближенно — методом Шеннона–Фано. Входными данными для обоих методов является заданное множество исходных символов для кодирования с их частотами; результат — экономные коды.

**Экономное кодирование методом Шеннона–Фано** (предложен независимо Шенноном и Фано в 1948–49 гг.; Роберт Фано — *Robert Fano*; американский ученый; 1917 г.). Упрощенно принцип алгоритма Шеннона–Фано заключается в следующем. Символы алфавита источника сообщения записывают в порядке убывающих вероятностей. Затем их разделяют на две части так, чтобы суммы вероятностей символов, входящих в каждую из частей, были примерно одинаковыми. Всем символам первой части в качестве первого символа комбинации кода приписывают нуль, а символам второй части в качестве первого символа комбинации — единицу. Потом каждую из этих частей (если она содержит более одного сообщения) делят в свою очередь на две, по возможности, равновероятные части и к ним применяют то же правило кодирования. Процедуру повторяют до тех пор, пока в каждой из частей не останется по одному сообщению.



Р. Фано Д. Хаффман

**Экономное кодирование методом Хаффмана** (Девид Хаффман — *David Huffman*; американский ученый; р. 1944 г.). Этот метод кодирования передаваемой информации появился очень давно и получил широкое распространение из-за простоты реализации и высокой скорости передачи. Метод не требует огромных вычислительных мощностей в отличие от арифметического кодирования, производящего целочисленные умножения и деления.

Для работы алгоритма необходимо иметь таблицу значений вероятностей различных символов, которые могут встретиться в данный момент кодирования.

На основе таблицы строят *бинарное дерево Хаффмана* следующим образом:

- отсортировать символы по возрастанию вероятности их появления;
- первые два символа в получившемся ряде объединить в один, сопоставив первому символу нуль, второму символу — единицу; вероятности этих двух символов сложить. Если в ряде остался один символ, то закончить, иначе перейти к первому пункту.

Основным и существенным недостатком этого метода является то, что он кодирует символы с помощью целого числа бит, что снижает степень сжатия и сводит на нет точное предсказание вероятностей, которое дают некоторые отличные современные алгоритмы моделирования. Отметим, что алгоритм создания кода Шеннона–Фано называют *сверху вниз*, а Хаффмана — *снизу вверх*. Преимуществами данных методов являются их очевидная простота реализации и, как следствие этого, высокая скорость кодирования/декодирования. Основным недостатком является их неоптимальность в общем случае.

### Помехоустойчивое кодирование

Проблема повышения верности приема обусловлена несоответствием между требованиями, предъявляемыми при передаче данных, и качеством реальных каналов связи. В сетях передачи данных требуется обеспечить верность не хуже $10^{-6}...10^{-9}$, а при использовании реальных каналов связи и простого (первичного) кода верность не превышает $10^{-2}...10^{-5}$. Одним из методов повышения помехоустойчивости систем связи является оптимизация устройств приема и обработки сигналов с помехоустойчивыми видами модуляции. Другим методом повышения помехоустойчивости приема дискретной информации является введение избыточности в передаваемые сообщения, что позволяет обнаруживать и исправлять ошибки в принятой информации. Методы внесения избыточности в передаваемый сигнал связаны с увеличением объема сигнала, т. е. с увеличением либо энергии сигнала, либо ширины спектра, либо времени передачи. Наиболее часто применяют введение избыточности увеличением времени передачи сигнала за счет использования помехоустойчивых кодов.

Первичные коды, полученные при примитивном кодировании, характеризуются тем, что отдельные его кодовые комбинации могут отличаться друг от друга лишь одним элементом. Поэтому даже один ошибочно принятый элемент в кодовой комбинации приводит к замене одной кодовой комбинации другой и, следовательно, к неправильному приему кодовой комбинации в целом.

*Помехоустойчивые* (избыточные, корректирующие) коды строят таким образом, что для передачи информации используют лишь часть кодовых комбинаций (разрешенные комбинации), отличающихся друг от друга более чем в одном разряде. Все остальные комбинации для передачи не используются и являются запрещенными. При использовании таких кодов ошибка в одном разряде приводит к замене разрешенной кодовой комбинации неразрешенной, что позволяет обнаружить ошибку. При достаточно большом отличии разрешенных комбинаций друг от друга возможно обнаружение ошибок большой кратности, поскольку они приведут к образованию запрещенных комбинаций, а переход одной разрешенной комбинации в другую будет происходить под действием ошибок более высокой кратности, являющихся результатом помех.

Предположим, что передаваемая по каналам связи информация может быть представлена в двоичной форме. Эта двоичная информация подлежит передаче по каналу, подверженному случайным ошибкам. Задача помехоустойчивого кодирования состоит в таком добавлении к информационным символам дополнительных символов, чтобы в приемнике эти искажения с помощью дополнительных символов могли быть найдены и исправлены. Иначе говоря, последовательность информационных символов представляется в виде некоторой более длинной последовательности символов, избыточность которой достаточна для защиты передаваемого сообщения.

Например, пусть для передачи информации по каналу связи используются четырехэлементные кодовые комбинации, отличающиеся друг от друга не менее чем в двух разрядах: 0011, 0110, 1001, 1010, 1100, 1111, 0101, 0000. Если при передаче любой из этих комбинаций произошла одиночная ошибка, то принятая комбинация будет запрещенной, что и свидетельствует о наличии в ней ошибки. Комбинация 1011, которая является запрещенной, получается за счет одиночной ошибки в первой, третьей, четвертой и шестой кодовых комбинациях соответственно в первом, втором, третьем и четвертом разряде.

Возьмем четырехэлементные комбинации, отличающиеся всеми четырьмя разрядами: 0011 и 1100. Легко убедиться, что при использовании этих комбинаций обнаруживаются одно-, двух- и трехкратные ошибки. Этот код, состоящий из двух кодовых комбинаций, можно использовать для исправления одиночных ошибок. Если, как и ранее, будет принята комбинация 1011, она будет отличаться от разрешенной комбинации 0011 одним разрядом, а от другой разрешенной 1100 — тремя, т. е. принятая комбинация «ближе» к комбинации 0011, чем к комбинации 1100, что и дает основание считать, что была комбинация 0011.

Повышенная помехоустойчивость двух рассмотренных кодов связана с имеющейся в них избыточностью. Так, первый код использует восемь кодовых

комбинаций из возможных шестнадцати ($2^4 = 16$), и можно было бы для их образования применить трехэлементный код ($2^3 = 8$). Второй код, обладающий большей помехоустойчивостью, потребовал соответственно еще большей избыточности — трех дополнительных разрядов.

Если экономное кодирование сокращает избыточность источника сообщений, то *помехоустойчивое* (*канальное*) кодирование, наоборот, заключается в целенаправленном введении избыточности (дополнительных символов) для того, чтобы появилась возможность обнаруживать и (или) исправлять ошибки, возникающие при передаче по каналу связи. Отличительное свойство помехоустойчивого кодирования состоит в том, что избыточность источника, образованного выходом кодера, больше, чем избыточность источника на входе кодера.

Базой для помехоустойчивого кодирования служит вторая теорема кодирования Шеннона, в которой отмечается, *что для канала связи с помехами всегда можно создать систему кодирования, при которой сообщения будут переданы со сколь угодно высокой степенью верности, если только производительность источника не превышает пропускной способности канала связи.* Упрощенный вариант теоремы гласит: *если скорость передачи не превышает пропускной способности канала связи с шумом, то всегда найдется способ кодирования, при котором сообщение будет передаваться с требуемой достоверностью.*

### Параметры помехоустойчивы кодов

Любой код принято характеризовать следующими параметрами:

- основание кода $m$;
- длина комбинации — длина слова (общее число символов в слове) $n$;
- длина информационной последовательности $k$;
- число проверочных символов $r$;
- число кодовых комбинаций $N_k$;
- кодовое расстояние (иначе, метрика) $d_{ij}$;
- скорость кода $R$;
- избыточность кода $v$;
- вес кодовой комбинации $\omega$;
- кратность ошибки;
- вероятность обнаружения ошибки (искажения) $p_{оо}$;
- вероятность необнаружения ошибки (искажения) $p_{но}$.

Проверку кода на безошибочность приема проводят по контрольному сигналу, называемому *синдромом*. Если синдром равен нулю, то ошибок в коде нет. Каждому коду соответствует свой синдром.

Цель внесения избыточных символов в помехоустойчивые коды — сформировать различные закодированные сообщения, как можно больше отличающиеся друг от друга. Для измерения степени отличия Хемминг ввел понятие *кодового расстояния Хемминга* (*Hamming distance*) $d_{ij}$ между двумя кодовыми словами — число несовпадающих в них разрядов, т. е. число единиц в сумме двух комбинаций по модулю 2.

Наименьшее значение $d_{ij}$ для всех пар кодовых последовательностей, т. е. минимальное число символов, которыми отличаются два кодированных сообщения, называют *минимальным кодовым расстоянием* и обозначают $d_0$. Очевидно, что для некодированных сообщений кодовое расстояние $d_0 = 1$. От кодового расстояния зависит *корректирующая способность* кода — возможность обнаруживать и/или исправлять с его помощью некоторые из ошибок. При декодировании сообщения с кодовым расстоянием $d_0$ можно обнаружить ошибки кратности $t_0 = (d_0 - 1)$ и исправить ошибки кратности

$$t_0 = [(d_0 - 1)/2)]. \tag{9.41}$$

Говорят, что в канале произошла ошибка кратности $t_0$, если в кодовой комбинации $t$ символов приняты ошибочно. Кратность ошибки есть не что иное, как расстояние Хэмминга между переданной и принятой кодовыми комбинациями, иначе, *вес вектора ошибки*. Параметр $t$ указывает, что все комбинации из $t_0$ или менее ошибок в любой принятой последовательности могут быть исправлены.

В теории кодирования показано, что систематический код обладает способностью обнаружить ошибки лишь тогда, когда минимальное кодовое расстояние для него больше или равно $2t_0$ (в случае обнаружения одиночных ошибок $t_0 = 1$). Это означает, что между соседними кодовыми комбинациями должна существовать по крайней мере одна кодовая комбинация. Запишем основные зависимости между кратностью обнаруживаемых ошибок $t_0$, исправляемых ошибок $t_н$, исправлением стираний $t_с$ и минимальным кодовым расстоянием $d_0$ кода:

$$d_0 \geq t_0 + 1; \qquad d_0 \geq t_0 + t_н + 1, \qquad \text{при } t_0 > t_н;$$

$$d_0 \geq 2t_н + 1; \qquad d_0 \geq 2t_н + t_с + 1, \qquad d_0 \geq t_с + 1.$$

*Избыточностью кода* называют разность между средней длинной слова и энтропией. Избыточность кода определяется как

$$R = \frac{n - k}{n}. \tag{9.42}$$

При этом скорость передачи

$$v = k/n. \tag{9.43}$$

Например, для блокового кода с длиной слова $n = 15$, исправляющего две ошибки, $k = 7$ и $R = (15 - 7)/15 \approx 0{,}533$, а при длине слова $n = 15$ и исправлении трех ошибок $k = 5$ и $R = (15 - 5)/15 \approx 0{,}666$. Соответственно $v = 0{,}467$ и $0{,}333$. Значит, при данной длине блоков $n$ повышение корректирующей способности связано с увеличением избыточности $R$ и уменьшением скорости передачи $v$. Так как число различных сообщений $M$, которые могут кодироваться блоком длиной $n$, то снижение скорости передачи $v$ эквивалентно уменьшению потока передаваемых сообщений. Отметим, что число проверочных символов $r = n - k$, а число ненулевых элементов в кодовой комбинации определяет ее вес $\omega$.

Коды, у которых требуемая корректирующая способность достигается при минимальной избыточности, называются *оптимальными*.

К. Шенноном доказано, что при кодировании длинных последовательностей (блоков) символов ($n \to \infty$) существуют коды с как угодно малой избыточностью ($R \to 0$) при как угодно малой вероятности ошибочного опознания сообщений. В используемых сейчас кодах обычно с увеличением $n$ корректирующая способность улучшается, однако при $n \to \infty$ у большинства кодов $R \to 1$.

Различают коды, обнаруживающие ошибки и исправляющие (корректирующие) коды, которые дополнительно к обнаружению еще и исправляют ошибки. Устранение ошибок с помощью *корректирующих кодов* (это называют *Forward Error Control*) реализуют в симплексных каналах связи. В дуплексных каналах достаточно применения кодов, *обнаруживающих* ошибки (*Feedback or Backward Error Control*), так как сигнализация об ошибке вызывает повторную передачу от источника. Это основные методы, используемые в сетях связи.

*Код с обнаружением ошибок* уменьшает число неверно опознанных сообщений, позволяет «стирать» или особо отмечать сообщения, в которых установлено присутствие ошибки, а в некоторых случаях (системы с обратными информационными связями) принять меры к повторной передаче и приему неопознанных сообщений.

Простейшим из помехоустойчивых кодов является код с проверкой на четность. Этот код позволяет обнаруживать одиночную ошибку. Его суть заключается в следующем. На передающем конце канала связи устройство кодирования проводит подсчет числа логических 1 в передаваемом двоичном кодовом слове. Если сумма 1 оказывается нечетной, в конец передаваемой кодовой комбинации добавляется 1, а если нет, то 0. На приемном конце канала связи проводится аналогичный подсчет, и если контрольная сумма (число единиц в принятой кодовой комбинации) будет нечетной, то принимается решение о том, что при передаче произошло искажение информации, в противном случае принятая информация признается достоверной. В описанном способе помехоустойчивого кодирования используется один добавочный контрольный разряд в кодовом слове. Это позволяет обнаруживать ошибку передачи только одного разряда в одном кодовом слове сообщения. При контроле на четность единственный способ получить достоверную информацию — повторная передача кодового слова. Для этого приемник формирует специальную команду, которую передают по каналу связи в обратном направлении к передатчику сообщений. Однако такие коды недостаточно надежны, особенно при появлении пачек ошибок. Более надежными обнаруживающими ошибки кодами являются *циклические коды*.

*Код с исправлением* (*коррекцией*) *ошибок* позволяет получать верные сообщения, несмотря на наличие некоторого числа ошибок при опознании символов. Коррекция (обнаружение и исправление) ошибок достигается лишь при использовании в кодовом слове определенного числа избыточных символов. Общий принцип построения такого кода аналогичен принципу построения кода, обнаруживающего одиночную ошибку. Однако при этом из всех кодовых слов длиной $n$ символов для использования в системе передачи необходимо отобрать лишь определенную часть, считая остальные слова запрещенными. При этом,

чтобы исправить одну ошибку, все отобранные кодовые слова должны отличаться друг от друга *как минимум на три символа*. В этом случае одиночная ошибка переведет переданное слово в одно из запрещенных, что и позволит обнаружить ошибку, но полученное запрещенное слово отличается от переданного лишь одним символом, а от остальных разрешенных слов — не менее чем на два символа. Значит, принятое слово «ближе» к действительно переданному и менее похоже на остальные разрешенные слова.

Если при передаче используют все возможные кодовые комбинации, то ошибка любой кратности остается незамеченной. Действительно, любое число замен превращает одну кодовую комбинацию в другую, а так как любая комбинация может встретиться в передаче, то нет сомнений в правильности принятой комбинации (если не использовать статистику, контекст и пр.).

Рассмотрим в качестве примера передачу четырех символов двухзначным двоичным кодом ($n = 2$; $k = 2$):

| сообщение | А | Б | В | Г, |
|---|---|---|---|---|
| код | 00 | 01 | 10 | 11. |

При передаче сообщения Б одиночная ошибка в первом знаке приведет к приему сообщения Г, а ошибка во втором знаке — к А, двойная ошибка — к В. Эти ошибки не различимы из-за того, что все сообщения отличаются лишь в одном или в двух знаках.

Для того, чтобы можно было обнаружить одиночную ошибку, достаточно взять такие кодовые комбинации, которые отличались бы между собой не менее чем в двух знаках. Таким образом, принцип построения кода, обнаруживающего одиночную ошибку, состоит в использовании не всех возможных комбинаций, а только половины. Вторая половина должна образовывать запрещенные комбинации. Одиночная ошибка превращает разрешенную комбинацию в запрещенную. Для того же примера код с обнаружением ошибки имеет вид:

| сообщение | А | Б | В | Г, |
|---|---|---|---|---|
| код | 000 | 011 | 101 | 110. |

Здесь каждая комбинация отличается от другой уже двумя знаками. Для этого пришлось взять трехзначный код ($N = 2^n = 2^3 = 8$).

Код, позволяющий исправить ошибку, строится из кодовых комбинаций, различающихся не менее чем в трех знаках. Ошибочно принятая комбинация всегда будет ближе к истинной, чем к любой другой разрешенной комбинации:

| сообщение | А | Б | В | Г, |
|---|---|---|---|---|
| код | 00000 | 01101 | 10110 | 11011. |

Здесь применен пятизначный код. Пусть принята комбинация 01001, которой в кодовой таблице нет. Следовательно, появилась ошибка. Принятая комбинация отличается от А и Г в двух знаках, от Б — в одном, от В — в пяти, значит, передавали сообщение В. Это эквивалентно *исправлению одиночной ошибки*.

Если в кодовой комбинации два знака заменяются ошибочными, то такая ошибка называется — двойной, если три — тройной и т. д. Чем больше исправляющая способность кода, тем выше кратность исправляемых ошибок, тем больше требуется знаков. Само исправление происходит путем сравнения принятого кода с разрешенными, и если обнаружена ошибка, то истинной считается та комбинация, от которой принятая наименее отличается. Это достаточно сложная задача, приводящая к громоздким техническим решениям. Поэтому разрабатываются специальные коды, строение которых позволяет осуществить декодирование более простыми способами. Одним из таких кодов являются *систематические коды*. Число информационных знаков определяется числом $N$ различных сообщений, которые нужно передать, т. е. $N = 2^k$. Из $r$ контрольных знаков образуется $2^r$ двоичных комбинаций.

Число $r$ определяется в зависимости от ситуации:

- указать, есть ли ошибка или нет;
- если ошибка есть, то указать, в какой из $n$ позиций она находится.

Исправление сводится к замене в указанном месте нуля единицей или наоборот. Таким образом, нужна одна кодовая комбинация для ответа «да» или «нет» на вопрос о наличии ошибки и $n$ кодовых комбинаций для указания номера ошибочной позиции. Отсюда должно выполняться неравенство

$$2^r \geq n + 1,$$

или

$$N \leq \frac{2n}{n+1}.$$

Итак, повышение помехоустойчивости, достигаемое с помощью корректирующих кодов, связано с увеличением значности кода, что предполагает либо увеличение длительности передаваемого сигнала, либо расширение полосы его частот, занимаемой сигналом.

### Линейный и нелинейный коды

Если избыточные символы кодовой комбинации (кодовой последовательности) образуются путем применения к информационным символам некоторых линейных операций, код называют *линейным*. В противном случае получается *нелинейный* код. Как линейные, так и нелинейные коды образуют обширные классы, содержащие много различных конкретных видов помехоустойчивых кодов. Среди линейных блочных кодов наибольшее значение в практике кодирования имеют коды с одной проверкой на четность, $M$-коды (симплексные), ортогональные, биортогональные, Хэмминга, БЧХ (названные по первым буквам имени открывших их независимо ученых Боуза, Чоудхури и Хоквингема (1959 – 1960 гг.; *Bose–Chaudhuri–Hocquenguem code — BCH*), Голея, Рида–Соломона (*Reed–Solomon*, иначе РС-коды). Коды Рида–Соломона были предложены в 1960 г. специалистами Ирвином Ридом (*Irving Reed*; р. 1923) и Густавом Соломоном (*Gustave Solomon*; 1930–1996).



И. Рид Г. Соломон М. Голей Р. Хэмминг

*Линейные коды* отличаются от нелинейных замкнутостью кодового множества относительно некоторого линейного оператора, например сложения или умножения слов кода, рассматриваемых как векторы пространства, состоящего из кодовых слов — векторов. Линейность кода означает, что каждый символ этого кода может быть линейно выражен через другие символы, что упрощает его построение и реализацию. При большой длине слов практически могут быть использованы только линейные коды. Вместе с тем часто нелинейные коды обладают лучшими параметрами по сравнению с линейными. Для относительно коротких кодов сложность построения и реализации линейных и нелинейных кодов примерно одинакова. Почти все схемы кодирования, применяемые на практике, основаны на линейных кодах. Двойные линейные блоковые коды часто называют *групповыми кодами*, поскольку кодовые слова образуют математическую структуру, называемую группой.

Помехоустойчивые коды можно разделить на три больших класса — блоковые (иногда, блочные, или полиномиальные) коды, сверточные коды и турбокоды (*turbo code*).

### Блоковые коды

При блоковом (сосредоточенном) кодировании каждому $i$-му дискретному сообщению соответствует кодовое слово с определенным числом $n_i$ и значением символов. При использовании *блоковых кодов* исходное сообщение делится на блоки и каждый блок кодируется независимо, т. е. добавляемые к нему избыточные символы не зависят от предыдущих блоков. Блоковые коды реализуются как систематические коды, т. е. коды «без памяти», поскольку у них $k$ входных информационных символов порождают блок из $n = k + r$ выходных символов, причем каждый блок из $n$ символов зависит от своего входного блока и $k$ символов и не зависит от других блоков, т. е. отсутствует память между блоками.

Блоковые коды бывают *разделимыми* и *неразделимыми*.

К разделимым относятся коды, в которых передаваемые символы по их назначению могут быть разделены на информационные и проверочные. Такие коды принято обозначать как ($n$, $k$)-коды, где $n$ — длина кода; $k$ — число информационных символов. Число комбинаций в коде не превышает $2^k$.

К неразделимым относятся коды, символы которых нельзя разделить по их назначению на информационные и проверочные.

Первые блоковые коды, позволявшие корректировать одиночные ошибки, были разработаны в 1949 и 1950 гг. известными американскими учеными Марселем Голеем (*Marcel Golay*; 1902 – 1989) и Ричардом Хэммингом (*Richard Hamming*; 1915 – 1998). При блочном кодировании последовательность информационных символов разбивается на отдельные блоки определенной длины. Для каждого такого блока формируются дополнительные проверочные символы, которые образуются путем сложения по модулю 2 определенных информационных символов. Блоки информационных и проверочных символов передаются в канал связи. Коды Хэмминга имели простой алгоритм декодирования и позволяли корректировать одну ошибку в кодовой комбинации. Для декодирования двоичных кодов большой длины могут быть использованы простейшие, мажоритарные (от франц. *Majoritaire* — право) декодеры. Кодирующие устройства используют регистры сдвига с обратными связями по модулю 2. Регистр сдвига имеет $k$ ячеек, что равно числу информационных символов. В эти ячейки записывается слово, состоящее из информационных символов, и после сигнала пуск с выхода регистра сдвига за $n$ тактов получаем слово циклического кода. При мажоритарном декодировании для каждого информационного символа формируется нечетное число оценок путем сложения по модулю 2 определенных комбинаций символов принятого кода. Решение об истинном значении принятого символа принимается по мажоритарному принципу — если большее число оценок равно 1, то принимается именно такое решение.

Простейшим случаем блокового кодирования является организация проверки на четность. Двоичный код с одним проверочным разрядом четности, доводящим количество единиц в кодовом слове (включая проверочный разряд) до четного числа, позволяет обнаруживать одну ошибку. Для этого надо на приемной стороне сосчитать число единиц в принятом кодовом слове. Если полученное число не является четным, то в канале произошла ошибка. Код с одной проверкой на четность имеет очень скромные возможности. Он находит применение только в каналах с малой вероятностью ошибок.

Увеличивая избыточность и вводя дополнительные проверочные разряды, можно не только обнаруживать сам факт принятой ошибки, но и определять позицию ошибки и исправлять ее. Это можно сделать, например, представляя данные одного информационного слова в виде двумерной матрицы и вводя дополнительные разряды четности в строки и столбцы матрицы. Для исправления ошибки надо найти строку и столбец матрицы, в которых произошли нарушения четности числа единиц, и заменить символ, принадлежащий найденным строке и столбцу, на противоположный. Этот способ кодирования хорошо иллюстрирует принцип исправления ошибок, но для использования на практике он мало подходит, так как избыточность его очень велика, хотя она и сокращается с ростом длины информационной части слова.

Увеличение исправляющей способности блоковых кодов достигается увеличением длины формируемого блока, что ведет к повышению сложности реализации кодера и декодера (кодека).

### Сверточные коды

Сверточные коды являются частным случаем (линейной реализацией) так называемых *решетчатых кодов*. Можно также полагать, что решетка является просто другим (иногда более удобным) способом представления и обычных сверточных кодов, которые могут быть систематическими и несистематическими. В последовательности кодовых символов систематического кода без изменения содержится последовательность информационных символов. В несистематическом коде положение информационных символов в кодовой последовательности указать нельзя. При сверточном кодировании символы кода распределяются во времени на значительном интервале, что повышает его устойчивость к одиночным пакетам ошибок. Кодер для сверточного кода представляет собой устройство с памятью, и добавляемые избыточные символы зависят не только от текущего фрагмента сообщения, но и от внутреннего состояния кодера, т. е. в конечном счете от предыдущих фрагментов сообщения.

Структурная схема сверточного кодера показана на рис. 9.29. Внутренняя структура кодера образуется сочетанием многоразрядных регистров сдвига и сумматоров по модулю 2. В зависимости от удобства схемотехнической реализации, формирования и сопряжения различных тактовых частот в состав кодера может входить один или несколько регистров. Входная информационная последовательность с тактовой частотой $F_T$ в последовательно-параллельном преобразователе ($S/P$) разделяется на блоки на $k_0$ потоков, которые непрерывно поступают в соответствующие входы кодера, задавая набор его входных данных. Регистры кодера тактируются с частотой $F_T/k_0$. На выходе кодера в каждом такте формируются $n_0$ выходных символов, которые также преобразуются в последовательный поток с повышенной тактовой частотой $(F_T/k_0)n_0 = F_T/R$ и поступают в канал связи. Выходная последовательность кодера может быть представлена как цифровая свертка входной информационной последовательности и импульсного отклика кодера (отсюда название кодов — *сверточные*).

Для представления сверточных кодов можно использовать порождающие многочлены и матрицы, кодовые деревья, диаграммы состояний или решетчатые структуры. С математической и схемотехнической точек зрения наиболее удобно описание с помощью порождающих многочленов, использующих операторы задержки $D$. Порождающие многочлены определяют для каждого из мно-

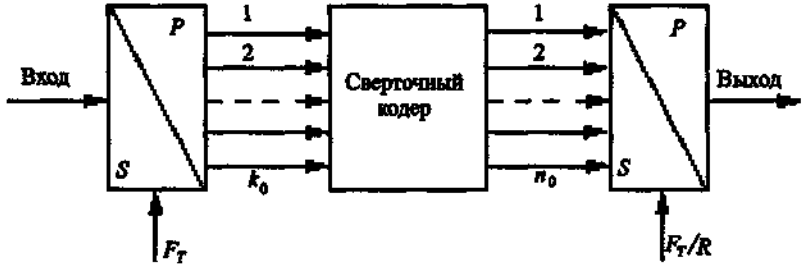



**Рис. 9.29.** Структурная схема сверточного кодера

говходовых сумматоров по модулю 2 номера разрядов регистра сдвига, к которым подключаются входы сумматоров.

Сверточные коды относят к рекуррентным. При рекуррентном кодировании кодируемую информационную последовательность не разбивают на блоки конечной длины, а кодовые символы вычисляют непрерывно по мере поступления символов информационной последовательности по определенным рекуррентным соотношениям, выбранным для данного типа кода. В случае использования линейных рекуррентных соотношений получается сверточный код. При этом кодовые символы образуются как свертка символов информационной последовательности и некоторой порождающей последовательности, которая задает линейные рекуррентные правила кодирования. Для сверточных кодов понятие кодового слова (кодовой комбинации) не имеет смысла, так как кодовые символы вычисляются по определенному отрезку последних информационных символов.

**Методы декодирования сверточных кодов.** В настоящее время известны три важнейших метода декодирования сверточных кодов: пороговое, последовательное и декодирование по максимуму правдоподобия (алгоритм Витерби).

При *пороговом декодировании* сверточных кодов вычисляют синдромы, затем синдромы или же некоторые последовательности, полученные посредством линейного преобразования синдромов, подаются на входы порогового элемента, где путем «голосования» и сравнения результатов с порогом выносится решение о значении декодируемого символа. Характерным для метода является простота реализации. Однако потенциальные корректирующие способности сверточного кода реализуются при этом не полностью. Более того, не все коды могут быть декодированы этим методом. Они должны обладать дополнительным свойством ортогональности. Алгоритм порогового декодирования основан на алгебраической структуре кода и применении мажоритарного принципа вынесения решения о каждом информационном символе. Алгоритм заведомо не оптимален.

Алгоритм *последовательного декодирования* основан на поиске наиболее вероятного пути на кодовом дереве путем последовательных проб с возможностью возвращения назад. С ростом числа ошибок в канале резко возрастает число проб и для декодирования требуется буферная память на входе декодера. Переполнение буфера ведет к отказу от декодирования и стиранию части принятых сообщений.

Алгоритм сверточного декодирования Витерби. Весьма значительным достижением в области теории сверточного кодирования явилась разработка в 1967 г. Эндрю Витерби (*Andrew Viterbi*; американский ученый; р. 1935 г.) эффективного с вычислительной точки зрения алгоритма декодирования сверточных кодов *по принципу максимального правдоподобия* (*maximum likelihood detector*). Этот алгоритм, в отличие от алгоритма последовательного декодирования, исследует все возможные пути по кодовой решетке на длине кодового ограничения $k$.



Э. Витерби

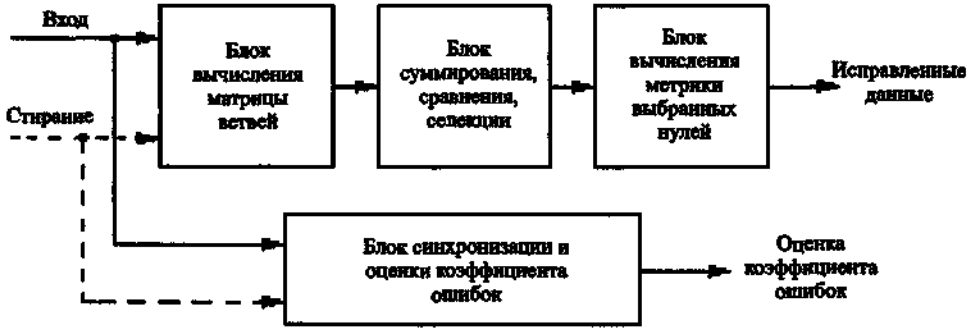



**Рис. 9.30.** Структурная схема декодера Витерби

Алгоритм Витерби основан на использовании вероятностных характеристик принимаемых сигналов. Алгоритм максимального правдоподобия является оптимальным и позволяет полностью реализовать корректирующую способность кода. Отыскание максимально правдоподобной последовательности можно организовать перебором всех возможных вариантов. Однако с ростом длины кодовой последовательности объем вычислений резко возрастает. При этом осуществляется динамический перебор всех возможных путей по решеточной диаграмме с одновременным отбрасыванием маловероятных отрезков путей. Алгоритм Витерби обеспечивает высокую помехоустойчивость и является простым при его технической реализации на специализированных микросхемах. Пример одной такой реализации показан на рис. 9.30.

Преимущество — сложность реализации декодера с мягким решением мало отличается от сложности декодера с жестким решением. Недостаток — экспоненциальный рост сложности схемы декодера в зависимости от длины кодового ограничения сверточного кода, которая по этой причине должна быть ограничена значением, примерно равным 10.

Отметим, что кодирование по Витерби используется для *прямой коррекции ошибок* (*Forward Error Correction — FEC*) за счет передачи с избыточностью.

### Циклические коды

Обширным подклассом линейных блоковых кодов являются циклические коды (*Cyclic Redundance Code — CRC*) — подкласс линейных кодов, обладающие тем свойством, что циклическая перестановка символов в кодированном блоке дает другой возможный кодированный блок того же кода. Циклические коды основаны на применении идей алгебраической теории полей Галуа.

Фактически все наборы, образованные циклической перестановкой любой кодовой комбинации, также являются кодовыми комбинациями. Так, например, циклические перестановки комбинации 1000101 будут также кодовыми комбинациями 0001011, 0010110, 0101100 и т. д. Это свойство позволяет в значительной степени упростить кодирующее и декодирующее устройства, особенно при обнаружении ошибок и исправлении одиночной ошибки. Внимание к цикличе-

ским кодам обусловлено тем, что присущие им высокие корректирующие свойства реализуют на основе сравнительно простых алгебраических методов. В то же время для декодирования произвольного линейного блокового кода чаще применяют табличные методы, требующие большой объем памяти декодера.

Итак, *циклическим кодом* называют линейный блоковый $(n,k)$-код, который характеризуется свойством цикличности, т. е. сдвиг влево на один шаг любого разрешенного кодового слова дает также разрешенное кодовое слово, принадлежащее этому же коду, и у которого множество кодовых слов представляется совокупностью многочленов степени $(n\text{-}1)$ и менее, делящихся на некоторый многочлен $g(x)$ степени $r = n - k$, являющийся сомножителем двучлена $x^{n+1}$. Многочлен $g(x)$ называется *порождающим*.

Как следует из определения, в циклическом коде кодовые слова представляются в виде многочленов (полиномов)

$$A(x) = A_{n-1}x^{n-1} + A_{n-2}x^{n-2} + A_i x^i + \ldots + A_1 x^1 + A_0 x^0,$$

где $n$ — длина кода; $A_i$ — коэффициенты из так называемого *поля Галуа* (значений кодовой комбинации).

Например, для кодовой комбинации 101101 полиномиальная запись имеет вид

$$A(x) = 1\cdot x^5 + 0\cdot x^4 + 1\cdot x^3 + 1\cdot x^2 + 0\cdot x^1 + 1 = x^5 + x^3 + x^2 + 1.$$

Примерами циклических кодов являются коды с четной проверкой, коды с повторениями, коды Хэмминга, РС-коды и турбокоды.

**Код Хемминга.** Возможности исправления ошибок в коде Хемминга связаны с минимальным кодовым расстоянием $d_0$. Исправляются все ошибки кратности $q = \text{ent}(d_0-1)/2$ (здесь ent означает целая часть) и обнаруживаются ошибки кратности $d_0 - 1$. Так, при контроле на нечетность $d_0 = 2$ и обнаруживаются одиночные ошибки. В коде Хемминга $d_0 = 3$. Дополнительно к информационным разрядам вводится $L = \log_2 Q$ избыточных контролирующих разрядов, где $Q$ — число информационных разрядов, $L$ округляется до ближайшего большего целого значения. $L$-разрядный контролирующий код есть инвертированный результат поразрядного сложения (т. е. сложения по модулю 2) номеров тех информационных разрядов, значения которых равны 1.

**Пример 9.7.** Пусть имеем основной код 100110, т. е. $Q = 6$.

Р е ш е н и е. Находим, что $L = 3$ и дополнительный код равен

$$010 \,\square\, 011 \,\square\, 110 = 111,$$

где $\square$ — символ операции поразрядного сложения, и после инвертирования имеем 000. Теперь с основным кодом будет передан и дополнительный. В приемнике вновь рассчитывают дополнительный код и сравнивают с переданным. Фиксируется код сравнения (поразрядная операция отрицания равнозначности), и если он отличен от нуля, то его значение есть номер ошибочно принятого разряда основного кода. Так, если принят код 100010, то рассчитанный дополнительный код равен инверсии от $010 \,\square\, 110 = 100$, т. е. 011, что означает ошибку в 3-м разряде.

Обобщением кодов Хэмминга являются циклические коды БЧХ, которые позволяют корректировать многократные ошибки в принятой кодовой комбинации. Декодирование кодов БЧХ производят на основе решения определенных алгебраических уравнений.

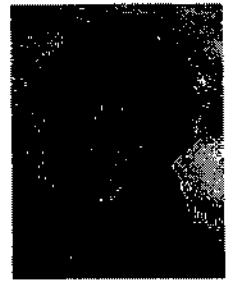
**Коды Рида–Соломона** базируются на специальном разделе математики — полях Галуа или конечных полях. Арифметические действия сложение, вычитание, умножение, деление и т. д. над элементами конечного поля дают результат, который также является элементом этого поля. Кодировщик или декодер Рида–Соломона должны обязательно выполнять эти операции. Все операции для своей реализации требуют специального оборудования или специализированного программного обеспечения.

### Турбокоды

Тот или иной избыточный код может применяться как самостоятельно, так и в виде некоторого объединения нескольких кодов, когда наборы символов одного избыточного кода рассматриваются как элементарные информационные символы другого избыточного кода. Такое объединение стали называть *каскадным кодом*. Огромным достоинством каскадных кодов является то, что их применение позволяет существенно упростить кодер и в особенности декодер по сравнению с аналогичными устройствами обычных (некаскадных) кодов той же длины и избыточности.

Каскадное кодирование привело к созданию так называемых *турбокодов*. Турбокоды были предложены около 10 лет назад. Турбокодом называют параллельную структуру сигнала, состоящую из двух или большего числа систематических кодов. Основной принцип их построения — использование нескольких (как правило, не более трех) параллельно работающих компонентных кодеров. В качестве компонентных можно использовать как блочные, так и сверточные коды. Использование перфорации (выкалывания) позволяет увеличить относительную скорость турбокода, адаптировав его исправляющую способность к статистическим характеристикам канала связи. Принцип формирования турбокода состоит в следующем. Входной сигнал $x$, состоящий из $K$ бит, подается параллельно на $N$ перемежителей. Каждый из последних представляет собой устройство, осуществляющее перестановку элементов в блоке из $K$ бит в псевдослучайном порядке. Выходной сигнал с перемежителей — символы с измененным порядком следования — поступает на соответствующие элементарные кодеры. Двоичные последовательности $x_i$ ($i = 1,2..., N$) на выходе кодера представляют собой проверочные символы, которые вместе с информационными битами составляют единое кодовое слово. Применение перемежителя позволяет предотвратить появление последовательностей коррелированных ошибок при декодировании турбокодов, что немаловажно при использовании традиционного в обработке сигналов итеративного (рекурентного) способа декодирования. Чаще всего в турбокодах применяют в качестве внутреннего двоичный код БЧХ, а внешнего — недвоичный РС-код.

### Арифметика конечного поля Галуа


Э. Галуа

Многие важнейшие помехоустойчивые коды систем передачи информации, в частности циклические, основаны на структурах конечных полей Галуа (Эварист Галуа — *Evariste Galois*; 1811–1832; французский математик; убит на дуэли).

Поля Галуа оказались необходимым рабочим инструментом для конструирования различных кодеров и декодеров систем передачи информации.

*Полем* называется множество элементов (математических объектов), которые можно «складывать», «вычитать», «умножать», «делить». Названия операций взяты в кавычки потому, что они не всегда являются общепринятыми арифметическими операциями. В поле всегда имеется нулевой элемент (обозначается 0), или нуль, и единичный элемент (обозначается 1), или единица.

Если число $q$ элементов поля ограничено, такое поле называется *конечным полем*, или *конечным полем Галуа*, и обозначается GF($q$), где $q$ — порядок поля. Наименьшим полем Галуа является двухэлементное поле GF(2), состоящее всего из двух элементов 1 и 0. Для того чтобы выполнение операций над элементами GF(2) не приводило к выходу за пределы этого поля, они осуществляются по модулю 2 (вообще это определяется порядком поля для случая так называемых *простых полей Галуа*).

Математическое понятие «поле» не является чем-то новым, с чем специалисты раньше не имели дела. Мы знакомы с полями в виде, например, множеств вещественных, комплексных, рациональных и т. д. чисел. Но это все поля с бесконечным количеством элементов.

Поле обладает рядом специфических математических свойств. Для элементов поля определены операции *сложения* и *умножения*, причем результаты этих операций должны принадлежать этому же множеству.

1. Для операций сложения и умножения выполняются обычные математические правила ассоциативности ($a + (b + c) = (a + b) + c$), коммутативности ($a + b = b + a$ и $a \cdot b = b \cdot a$) и дистрибутивности ($a \cdot (b + c) = a \cdot b + a \cdot c$).

2. Для каждого элемента поля $a$ обязательно должны существовать *обратный элемент по сложению* ($-a$) и, если $a$ не равно нулю, *обратный элемент по умножению* ($a^{-1}$).

3. Поле должно содержать *аддитивную единицу* — элемент 0, такой, что $a + 0 = a$ для любого элемента поля $a$.

4. Поле должно содержать *мультипликативную единицу* — элемент 1, такой, что $a \cdot 1 = a$ для любого элемента поля $a$.

Например, существуют поля вещественных чисел, рациональных чисел, комплексных чисел. Эти поля содержат бесконечное число элементов. Очевидно, не существует поля целых чисел, так как обратный элемент по умножению не принадлежит множеству целых чисел.

**Сигнально-кодовые конструкции в модемах**

Применение многопозиционной КАМ в чистом виде сопряжено с проблемой недостаточной помехоустойчивости. Поэтому в современных высокоскоростных протоколах КАМ используют совместно с решетчатым кодированием — специальным видом сверточного кодирования. В результате появилась треллис-модуляция. Выбранная определенным образом комбинация КАМ и помехоустойчивого кода относится к сигнально-кодовой конструкции. СКК позволяют повысить помехозащищенность передачи информации наряду со снижением требований к отношению сигнал/шум в канале на 3...6 дБ. При этом число сигнальных точек увеличивается вдвое за счет добавления к информационным битам одного избыточного, образованного путем сверточного кодирования. Расширенный таким образом блок битов подвергается все той же КАМ. В процессе демодуляции производится декодирование принятого сигнала по алгоритму Витерби. Именно этот алгоритм за счет использования введенной избыточности и знания предыстории процесса приема позволяет по критерию максимального правдоподобия выбрать из сигнального пространства наиболее достоверную эталонную точку.

## 9.6. Модемные устройства систем связи

**Общие сведения.** Современные системы связи — это модемные технологии, реализованные под конкретные скорости передачи информации. Модем такой системы связи, выполняющий сложнейшие математические операции, можно приравнять к компьютеру третьего поколения. Сведения о внутреннем устройстве и архитектуре последних моделей модемов не настолько доступны, как, например, информация об устройстве персональных компьютеров. Одной из причин этого является отсутствие каких бы то ни было промышленных стандартов на конструкцию модемов. Другая причина состоит в том, что современные модемы, как правило, строятся на наборах специализированных микросхем, которые реализуют основные модемные функции. Это приводит к тому, что в отличных по конструкции модемах одни и те же методы и протоколы реализованы различными способами. Вместе с тем практически все современные модемы имеют похожие структурные схемы, содержащие (рис. 9.31):

- порты канального и *DTE – DCE* интерфейсов;
- основной или универсальный процессор (*PU*);
- цифровой сигнальный процессор (*Digital Signal Processor — DSP*)
- модемный процессор (собственно модулятор/демодулятор);
- постоянное запоминающее устройство (ПЗУ, *ROM*);
- перепрограммируемое запоминающее устройство (ППЗУ, *ERPROM*, часто *Non-Volatile RAM, NVRAM* — неразрушающаяся память с прямым доступом);
- оперативное запоминающее устройство (ОЗУ, *RAM*);
- схемы индикаторов состояния модема.

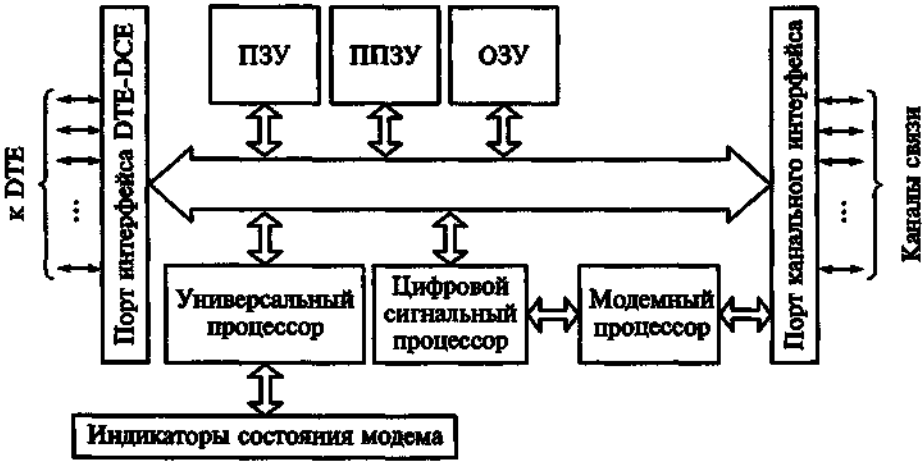



**Рис. 9.31.** Структурная схема современного модема

Аббревиатурой *DTE* (*Data Terminal Equipment* — оконечное оборудование передачи данных) обозначают оконечные цифровые устройства, генерирующие или получающие данные. Аббревиатурой *DCE* (*Data Communication Equipment* — оборудование передачи данных) обозначают модемы. Линия связи между *DCE* — аналоговая, между *DCE* и *DTE* — цифровая.

*Порт интерфейса DTE – DCE* обеспечивает взаимодействие с *DTE*. Если для связи *DTE* и *DCE* используется унифицированный цифровой интерфейс, то это зачастую дает возможность связать два расположенных рядом *DTE* прямой цифровой линией — так называемым *нуль-модемным кабелем*. В случае разнесения *DTE* на большое расстояние в разрыв вместо нуль-модемного кабеля включаются пара модемов и аналоговая линия связи, обеспечивая прозрачное соединение и передачу данных.

*Универсальный процессор* выполняет функции управления взаимодействием с *DTE* и схемами индикации состояния модема. Именно он выполняет посылаемые *DTE* AT-команды (служат для получения сведений о состоянии модема, изменении режимов его работы, наборе номера, установке/завершении связи и тестировании модема и линии) и управляет режимами работы остальных составных частей модема. Универсальный процессор фактически является встроенным микрокомпьютером, отвечающим за прием и выполнение команд, буферизацию, обработку некоторых данных и т. п., а также за управление сигнальным процессором. В большинстве модемов используются специализированные процессоры на основе типовых наборов микросхем. Интеллектуальные возможности модема определяются в основном типом используемого *PU* и микропрограммой управления модемом, хранящейся в *ROM*.

На *цифровой сигнальный процессор*, как правило, возлагаются задачи по реализации основных функций протоколов модуляции (кодирование сверточным кодом, относительное кодирование, скремблирование, декодирование, компрессии/декомпрессии, т. е. сжатия/распаковки).

*Модемный процессор* (собственно модулятор/демодулятор) занимается непосредственно операциями с сигналом — модуляцией/демодуляцией, разделением частотных полос, эхо-подавлением и т. п. В зависимости от типа и сложности модема основная интеллектуальная нагрузка смещается в сторону модемного процессора. В низкоскоростных (300...2 400 бит/с) модемах основную работу выполняет модемный процессор, в скоростных (4 800 бит/с и выше) — сигнальный процессор.

В *постоянном запоминающем устройстве* хранятся программы для основного и сигнального процессоров, а также микропрограмма управления модемом — прошивка, включающая в себя наборы команд и данных для управления модемом. ПЗУ может быть однократно программируемым (*PROM*), перепрограммируемым со стиранием ультрафиолетом (*EPROM*) или перепрограммируемым электрически (*EEPROM*). Последний тип ПЗУ позволяет оперативно менять прошивки по мере исправления ошибок или появления новых возможностей. Путем замены или перепрограммирования ROM иногда можно достичь существенного улучшения свойств модема, т. е. произвести его модернизацию, или апгрейд (*upgrade*). Такого рода модернизация некоторых моделей модемов может обеспечить поддержку новых протоколов или сервисных функций, таких как автоматическое определение номера (АОН) вызывающего абонента. Для облегчения такой модернизации в последнее время вместо микросхем *ROM* стали широко применяться микросхемы флэш-памяти (*FLASHROM*). Флэш-память позволяет легко обновлять микропрограмму модема, исправляя ошибки разработчиков и расширяя возможности устройства.

*Перепрограммируемое запоминающее устройство* позволяет сохранять установки модема в так называемых *профайлах* или *профилях* на время выключения. Большинство команд изменения состояния действует только на текущий набор параметров, теряющий свои значения при отключении или сбросе модема. Содержимое текущего набора может быть записано в один из сохраненных ранее наборов в ППЗУ; кроме этого, ряд команд может непосредственно изменять содержимое ППЗУ. Обычно имеется две сохраняемые установки — основная (*profile* 0) и дополнительная (*profile* 1). По умолчанию для инициализации используется основной набор, но есть возможность переключиться на дополнительный. Ряд модемов имеет более двух сохраняемых наборов.

*Оперативное запоминающее устройство* интенсивно используют для временного хранения данных, работы алгоритмов сжатия и выполнения промежуточных вычислений как универсальным, так и цифровым сигнальным процессорами. В ОЗУ хранится также текущий набор параметров модема (*active profile*).

Современный телефонный модем располагает еще и аналоговой частью, ответственной за сопряжение модема с телефонной сетью, — устройство набора номера, усилитель, АЦП и ЦАП, ответственные за преобразование сигнала из аналоговой формы в цифровую, и наоборот, соответственно. Практически все телефонные (и другие) модемы производят обработку информации в цифровой форме, без сколь либо сложной аналоговой предобработки, так как это позволя-

ет добиться высокой стабильности и в значительной степени упростить разработку и анализ алгоритмов. При этом обычно частота дискретизации находится в пределах 7...12 кГц и более. Теоретически частота дискретизации должна быть как минимум в два раза выше максимальной частоты сигнала для того, чтобы сигнал представлялся отдельными отсчетами без потерь. Количество уровней квантования для ЦАП и АЦП современных модемов достигает десятков тысяч. Обычно, поскольку с «цифровой стороны» ЦАП и АЦП пишутся или читаются в виде числа, говорят о количестве разрядов у ЦАП/АЦП, т. е. количестве разрядов двоичного числа, требуемого для представления всех возможных уровней, например 16-разрядный АЦП может распознавать 65 536 уровней, обозначаемых числами от –32 768 до +32 767.

### Некоторые виды модуляции в модемах

В модемах применяют несколько уже упоминавшихся в главе 2 видов модуляции:

- частотная манипуляция (*Frequency Shift Keying — FSK*);
- бинарная фазовая манипуляция (*Binary Phase-Shift Keying — BPSK*);
- относительная фазовая манипуляция (*Differential Phase Shift Keying — DPSK*);
- квадратурная амплитудная модуляция (КАМ; *quadrature amplitude modulation*, QAM), позволяет осуществить кодирование 8 бит на бод;
- квадратурная относительная фазовая модуляция (КОФМ; *Quadrature phase shift keying* — QPSK);
- треллис-модуляция (*Trellis Coded Modulation*; *TCM*).

### Классификация модемов

В настоящее время строгой классификации модемов не существует и, вероятно, не может существовать по причине большого разнообразия как самих модемов, так и сфер применения и режимов их работы. Тем не менее можно выделить ряд классификационных признаков модемов: область применения; функциональное назначение; тип используемого канала; конструктивное исполнение; поддержка протоколов модуляции, исправления ошибок и сжатия данных. Можно выделить и более детальные технические признаки, такие как применяемый способ модуляции, интерфейс сопряжения с *DTE* и т. д. Можно разделить модемы и по скоростям передачи информации (14 400 бит/с, 28 800 бит/с, 33 600 бит/с, 56 Кбит/с).

**По области применения** современные модемы можно разделить на несколько групп:

- для сотовых систем связи;
- радиомодемы;
- для пакетных радиосетей;
- для локальных радиосетей;
- для физических линий;

- спутниковые;
- для цифровых систем передачи данных (*ISDN*);
- кабельные;
- для выделенных телефонных каналов (*DSL*-модемы);
- *PLC*-модемы;
- для оптоволоконных линий;
- для коммутируемых телефонных каналов;
- голосовые.

Сотовые модемы используют для мобильной радиотелефонии, к которой относится и сотовая связь. Интересно, что эти модемы не содержат радиостанции (в отличие от радиомодемов), а лишь передают в нее свой сигнал. Модемы для сотовых систем связи отличаются компактностью исполнения и поддержкой специальных протоколов модуляции и исправления ошибок, позволяющих эффективно передавать данные в условиях сотовых каналов с высоким уровнем помех и постоянно изменяющимися параметрами. При пересечении границы сот (в случае сотовой связи) происходит переключение на другую радиостанцию и связанное с этим временное пропадание сигнала (*hand-off*). Большинство обычных модемов в этих условиях пытается или возобновить соединение или рвет его, что неправильно. За счет отражения сигнала от зданий приходят несколько сигналов (*multipath*) и накладываются, сигнал искажается или вообще периодически пропадает. Ясно, что для такой работы нужны специальные протоколы. Ведущие производители поставляют модемы такого типа.

Сотовые модемы могут работать и как обычные модемы, т. е. выполнены как модемы с дополнительными возможностями, но часто имеют портативное исполнение. Главное для сотовых модемов — значительные изменения параметров линии и сигнала в течение сеанса. Поэтому модемы используют специальные протоколы передачи, главное свойство которых — динамическая адаптация к линии. Например, при наступлении неблагоприятных условий модемы уменьшают скорость или размер блока, а при улучшении — снова увеличивают.

Радиомодемы используют свободное пространство как среду передачи сигналов. Поэтому вместо телефонного разъема радиомодем имеет антенный разъем, куда вставляется антенна или антенный кабель. Кроме того, радиомодем содержат передатчик/приемник. Внешне радиомодем выглядит как настольный и подключается к компьютеру через стандартный интерфейс *RS*-232C, только имеет антенный вывод. В него подключается или входящая в комплект штыревая антенна небольшого размера (около 30 см) или, если нужна большая дальность, антенный кабель, усилитель и направленная антенна.

В современных радиомодемах часто стала использоваться шумоподобная технология передачи сигналов, которая достаточно устойчива к обычным помехам и ставит практически непреодолимые препятствия для перехвата данных. Однако используемая высокая частота (порядка 900 МГц и выше) требует прямой видимости, хотя есть возможность обойти это ограничение, построив ретрансляцию по ломаной линии.

Если система связи занимает несколько точек, разбросанных по городу, то для их объединения в общую сеть можно применять специальные радиомодемы, работающие в СВЧ-диапазоне. Они обеспечивают устойчивую связь на расстоянии от сотен метров до десятков километров. Конечно, разнесенные на несколько километров локальные сети можно объединить и через оптоволоконный кабель, однако это весьма дорогостоящее решение.

В первую очередь можно классифицировать такое оборудование по диапазону частот, в котором оно работает. Причем от того, в каком диапазоне работает оборудование, зависят такие показатели, как дальность связи, скорость передачи информации и требования к обеспечению «прямой видимости», зависимость качества связи от погодных условий. Наиболее распространенные диапазоны частот, предназначенные для систем передачи данных, 136...174 МГц, 400...512 МГц, 820...960 МГц, 2...4 ГГц, 10...12 ГГц, 30...35 ГГц и выше. При этом зависимость параметров следующая: чем выше частота, тем выше скорость передачи данных, меньше дальность, выше требования к обеспечению прямой видимости и больше чувствительность к перемене погоды. Эта зависимость иллюстрируется следующими примерами, показывающими, каких параметров связи можно добиться при использовании различных диапазонов частот:

- 136...174 МГц — скорость передачи данных до 19,2 Кбит/с, дальность связи до 50...70 км, связь может осуществлять «из-за угла» и за горизонтом за счет искривления пути прохождения радиолуча у земли;
- 400...512 МГц — скорость передачи данных до 128 Кбит/с, дальность связи до 40...50 км.

Возможна радиосвязь при помощи приема сигналов, отраженных от различных зданий и сооружений, гор и т. д., хотя наличие прямой видимости желательно в диапазонах 800...960 МГц и выше. Возможна организация каналов передачи данных со скоростью свыше 2 Мбит/с, при этом обязательным является условие прямой видимости между антеннами. С ростом используемой частоты увеличивается влияние погодных условий и уменьшается дальность связи, так как условия распространения радиоволн в этом диапазоне приближаются к условиям распространения света. Кроме того, дальность связи зависит от мощности передатчиков, чувствительности приемников и характеристик применяемых антенн и радиочастотного кабеля.

Передачу данных в системах проводной связи абонентского доступа можно существенно улучшить с помощью модемной технологии (аппаратура «последней мили»), решив проблему увеличения скорости передачи информации на участке «абонент–станция» без замены традиционных телефонных кабелей на оптоволоконные.

Пакетные радиомодемы предназначены для передачи данных по радиоканалу между мобильными пользователями. При этом несколько радиомодемов используют один и тот же радиоканал в режиме многостанционного доступа, например, многостанционного доступа с контролем несущей. Радиоканал по своим характеристикам близок к телефонному и организуется с исполь-

зованием типовых радиостанций, настроенных на одну и ту же частоту в метровом или дециметровом диапазоне. Пакетный радиомодем реализует методы модуляции и многостанционного доступа.

Локальные радиосети являются развивающейся перспективной сетевой технологией, дополняющей обычные локальные сети. Ключевой их элемент — специализированные *радиомодемы локальных радиосетей* (адаптеры локальных радиосетей). В отличие от пакетных радиомодемов такие модемы обеспечивают передачу данных на небольшие расстояния (до 300 м) с высокой скоростью (2...10 Мбит/с), сопоставимой со скоростью передачи в проводных локальных сетях. Кроме того, радиомодемы локальных радиосетей работают в определенном диапазоне частот с применением сигналов сложной формы, таких как сигналы с псевдослучайной перестройкой частоты. Радиосети передачи данных применяют в тех случаях, когда организация проводных или оптоволоконных каналов связи невозможна физически, либо если существующие проводные каналы связи не удовлетворяют потребителей с точки зрения скорости передачи информации, или их использование является экономически нецелесообразным.

Модемы для физических линий отличаются от других типов модемов тем, что полоса пропускания физических линий не ограничена значением 3,4 кГц, характерным для телефонных каналов. Однако полоса пропускания физической линии также является ограниченной и зависит в основном от типа физической среды (экранированная и неэкранированная витая пара, коаксиальный кабель и др.) и ее длины. С точки зрения используемых для передачи сигналов модемы для физических линий могут быть разделены на *модемы низкого уровня* (линейные драйверы или модемы на короткие расстояния — *short range modems*), использующие цифровые сигналы, и *модемы основной полосы* (*baseband modems*), в которых применяются методы модуляции, аналогичные применяемым в модемах для телефонных каналов. В модемах первой группы обычно используются цифровые методы биимпульсной передачи, позволяющие формировать импульсные сигналы без постоянной составляющей и часто занимающие более узкую полосу частот, чем исходная цифровая последовательность. В модемах второй группы часто используются различные виды квадратурной амплитудной модуляции, позволяющие радикально сократить требуемую для передачи полосу частот. В результате на одинаковых физических линиях такими модемами может достигаться скорость передачи до 100 Кбит/с, в то время как модемы низкого уровня обеспечивают только 19,2 Кбит/с.

Модемы для передачи на короткие расстояния используют для связи между компьютерами, маршрутизаторами и другой аппаратурой цифровой связи, например, внутри зданий, в границах города. Эти модемы проектируют с целью преодолеть ограничение в дальности действия цифровых интерфейсов. Модемы для передачи на короткие расстояния позволяют преодолеть ограничение в дальности и проблемы помех благодаря применению специальной модуляции сигнала и методов коррекции линии, которые обеспечивают свободную от ошибок передачу информации на более дальние расстояния.

Спутниковые модемы предназначены для передачи информации по спутниковым каналам связи. Выпускаемые в настоящее время спутниковые модемы работают в различных диапазонах частот, имеют возможность перестройки и установки основных параметров, включая рабочую частоту, коэффициент усиления, выходную мощность, тип модуляции, скорость кодирования, тип скремблирования, размеры буферов для данных и т. д. Эти параметры могут изменяться с малым шагом в широком диапазоне значений. Также следует отметить, что любой современный модем имеет систему встроенных процессоров с развитым программным обеспечением, позволяющую изменять конфигурацию модема либо с помощью его собственных органов управления и дисплея, либо через его порт контроля и управления.

По максимальной скорости передачи информации следует выделить три основные группы спутниковых модемов:

- низкоскоростные (максимальная скорость до 2 048 Кбит/с), позволяющие организовать стандартные каналы на скоростях 16, 24, 32, 64, 128, 256, 512, 1 024, 2 048 Кбит/с;
- среднескоростные — максимальная скорость от 2 048 до 8 448 Кбит/с;
- высокоскоростные (максимальная скорость свыше 8 448 Кбит/с), работающие на скоростях 51,840, 69,632, 139,264 и 155,52 Мбит/с.

Модемы для цифровых систем передачи данных напоминают модемы низкого уровня. Однако в отличие от них обеспечивают подключение к стандартным цифровым каналам, таким как ISDN, и поддерживают функции соответствующих канальных интерфейсов.

Кабельные модемы используют для обмена данными по специализированным кабелям — к примеру, через кабель коллективного телевидения. Вместо телефонных линий кабельные модемы используют коаксиальные кабели с широкой полосой пропускания, применяемые для передачи видеоинформации. До сотни телевизионных каналов — лишь небольшая часть той информации, что фактически может быть передана в квартиру. Если бы весь кабель был использован для передачи информации, ее можно было бы получать со скоростью, превышающей 750 Мбит/с, что в тысячи раз быстрее телефонного соединения. Для полноты обзора нужно упомянуть о подключении с их помощью к *Internet*. Через коаксиальный телевизионный кабель с помощью кабельного модема можно передавать данные со скоростью порядка 1 Мбит/с, что не так уж и плохо, несмотря на то, что в обратном направлении (в сторону *Internet*) данные уходят медленнее. За рубежом в виде эксперимента некоторые жилые дома и кварталы подключаются к Internet с помощью кабельных модемов, однако у нас этот способ подключения пока еще не получил широкого распространения.

*DSL*-модемы (*DSL — Digital Subscriber Line* — цифровая абонентская линия) используют для организации связи выделенные (некоммутируемые), как правило, обычные телефонные линии. Отличаются от коммутируемых модемов кодированием сигналов. Любой абонент, пользующийся в настоящий момент обычной телефонной связью, имеет возможность с помощью технологии *DSL*

значительно увеличить скорость соединения, например, с сетью модемов для физических линий. В результате он получает круглосуточный доступ в сеть *Internet* с сохранением нормальной работы обычной телефонной связи. Благодаря многообразию технологий *DSL* пользователь может выбрать подходящую именно ему скорость передачи данных — от 32 Кбит/с до более чем 50 Мбит/с.

PLC-модемы (*PLC — Powerline Communication* — связь по силовым электросетям) используют технологию передачи данных по проводам бытовой электрической сети. Широкополосный доступ *PLC* на территории России известен как *Internet через розетку* — еще одно его название — *Internet*-электросеть (выход в *Internet* через обычную электросеть).

Модемы для оптоволоконных линий могут работать как на одномодовом, с длиной волны 860 нм, так и многомодовом оптоволокне с длиной волны 1 300 или 1 550 нм:

- 860 нм — наиболее популярны, но имеют существенное ограничение на длину кабеля — до 5 км по многомодовому кабелю. Источник излучения — светодиод;
- 1 300 нм — более универсальны — до 20 км по одномодовому волокну со светодиодом, до 50 км с применением полупроводникового лазера;
- 1 550 нм — по одномодовому с применением полупроводникового лазера до 100 км. Максимальная дистанция зависит и от диаметра кабеля.

Подавляющее большинство выпускаемых модемов предназначено для использования на коммутируемых телефонных каналах. Модемы для коммутируемых телефонных каналов должны работать с автоматическими телефонными станциями, различать их сигналы и передавать свои сигналы набора номера.

Голосовой модем (термин «голосовой» точнее, чем звуковой, так как телефонная сеть передает лишь голосовую часть звукового диапазона) имеет функцию оцифровки сигнала с телефонной линии и воспроизведение произвольного звука в линию. Часть голосовых модемов имеет встроенный микрофон. Это позволяет осуществить:

- передачу голосовых сообщений в режиме реального времени на другой удаленный голосовой модем, прием сообщений от него и воспроизведение их через внутренний динамик;
- использование такого модема в режиме автоответчика и для организации голосовой почты.

Этот тип модема позволяет: принимать из телефонной сети голосовые сообщения, записывая их в файл (*File*); воспроизводить в телефонную сеть ранее сформированные голосовые (в общем случае звуковые) файлы. Сформированный голосовым модемом файл можно затем проиграть несколькими способами, например: на динамик компьютера; более качественно звуковой платой; на трубку телефонного аппарата.

Никаких специальных подсоединений, как в предыдущем случае, делать не надо. Наговорить файлы можно и в динамик телефонной трубки, если подсоединить ее подходящим образом к модему.

**По методу передачи** модемы делятся на асинхронные и синхронные. Говоря о синхронном либо асинхронном методе передачи обычно подразумевают передачу по каналу связи между модемами. Как правило, синхронизация реализуется одним из двух способов, связанных с тем, как работают тактовые генераторы отправителя и получателя: независимо друг от друга (асинхронно) или согласованно (синхронно).

Асинхронный режим передачи используется, главным образом, тогда, когда передаваемые данные генерируются в случайные моменты времени, например пользователем. При такой передаче получающее устройство должно восстанавливать синхронизацию в начале каждого получаемого символа. Для этого каждый передаваемый символ обрамляется дополнительным стартовым и одним или более стоповыми битами. Такой асинхронный режим часто применяют при передаче данных по интерфейсу *DTE – DCE*. При передаче цифровых данных по каналу связи возможности применения асинхронного режима во многом ограничены низкой эффективностью и необходимостью использования простых методов модуляции, таких как амплитудная и частотная. Более совершенные методы модуляции, такие как ОФМ, КАМ, требуют поддержания постоянного синхронизма опорных тактовых генераторов отправителя и получателя.

**Контроль ошибок.** Вероятность ошибок никогда не исключена, поэтому в асинхронной передаче используется специальный бит — *бит четности*. Схема проверки и коррекции ошибок, которая его применяет, называется *контролем четности*. Число посланных и принятых единичных битов при контроле четности должно совпадать.

Синхронный режим передачи основан на синхронизации, согласованной между двумя устройствами. Ее цель — выделить биты из группы при передаче их блоками (кадрами). Для установления синхронизации и проверки правильности ее работы используются специальные символы. Поскольку информационные биты передаются по линии в синхронном режиме, стартовые и стоповые биты не нужны. Передача данных завершается в конце одного кадра и начинаются в начале другого. Этот метод более эффективен, чем асинхронная передача. В случае ошибки синхронная схема распознавания и коррекции ошибок повторяет передачу кадра. Чтобы принимающее устройство могло обеспечить различные уровни синхронизации, должны выполняться следующие требования:

- передаваемая последовательность битов не должна содержать длинных последовательностей нулей или единиц для того, чтобы принимающее устройство могло устойчиво выделять тактовую частоту синхронизации;
- каждый кадр должен иметь зарезервированные последовательности битов или символов, отмечающие его начало и конец.

Кроме высокоскоростной передачи данных собственно по физическим каналам систем связи синхронный режим часто применяется и для передачи по интерфейсу *DTE – DCE*. В этом случае для синхронизации устройств используются дополнительные интерфейсные цепи, по которым передается сигнал тактовой частоты от отправителя к получателю.

**По конструкции** различают модемы:

- внешние;
- внутренние (*internal*);
- портативные;
- групповые.

Внешние модемы представляют собой автономные устройства, подключаемые к компьютеру или другому *DTE* посредством одного из стандартных интерфейсов *DTE – DCE*.

Внутренний модем — это плата расширения, вставляемая в соответствующий слот компьютера. Каждый из вариантов конструктивного исполнения имеет свои преимущества и недостатки, которые будут рассмотрены далее.

Портативные модемы предназначены для использования мобильными пользователями совместно с компьютерами класса *Notebook*. Они отличаются малыми габаритными размерами и высокой ценой. Их функциональные возможности, как правило, не уступают возможностям полнофункциональных модемов. Часто портативные модемы оснащены интерфейсом *PCMCIA*.

Групповыми модемами называют совокупность отдельных модемов, объединенных в общий блок и имеющих общие блок питания, устройства управления и отображения. Отдельный модем группового модема представляет собой плату с разъемом, устанавливаемую в блок, и рассчитан на один или небольшое число каналов.

Передача данных организуется на основе набора протоколов, каждый из которых устанавливает правила взаимодействия связывающихся устройств.

**Протоколы**, используемые в модемах, делятся на четыре основные группы:

- модуляции и передачи данных;
- коррекции ошибок;
- сжатия передаваемых данных;
- связи *DTE* и *DCE*.

Первые три группы относятся только к связи *DCE – DCE*, последняя — только к связи *DCE – DTE*.

Первая группа протоколов устанавливает правила вхождения модемов в связь, ее поддержки и разрыва, параметры аналоговых сигналов, правила модуляции и кодирования. Протоколы непосредственно относятся к сигналам, передаваемым по межмодемной аналоговой линии связи. Соединение двух модемов возможно лишь в случае поддержки ими каких-либо общих или совместимых протоколов этой группы. В семиуровневой иерархии протоколов связи OSI эта группа протоколов имеет уровень 1 (физический) и формирует канал цифровой связи в реальном времени, однако не защищенный от ошибок передачи.

Протоколы физической связи могут быть *симплексными* — реализующими в каждый момент времени передачу только в одну сторону, и *дуплексными* — с одновременной двунаправленной передачей. Чаще всего применяются дуплексные протоколы, которые могут быть симметричными, когда скорости передачи в обоих направлениях равны, и несимметричными, когда скорости различаются.

Несимметричный дуплекс применяется для повышения скорости передачи в одну сторону за счет ее снижения в обратную сторону, когда поток передаваемых данных имеет выраженную асимметрию. Для определения направления передачи в физическом канале используются понятия *вызывающего* (инициирующего соединение) и *отвечающего модемов*; направление передачи определяется со стороны вызывающего модема.

Вторая группа устанавливает правила обнаружения и коррекции ошибок, возникающих на этапе передачи с помощью протоколов первой группы. Эти протоколы имеют дело лишь с цифровой информацией; для проверки целостности информации она разделяется на блоки (пакеты), снабжаемые контрольными избыточными кодами. При несовпадении контрольного кода на приемном конце переданный пакет считается ошибочным и запрашивается его повторная передача. Эта группа протоколов формирует из ненадежного физического канала надежный (защищенный от ошибок) канал более высокого уровня, но это приводит к потере связи в реальном времени и дается ценой определенных накладных расходов. В модели *OSI* эта группа соответствует уровню 2 (канальному).

Третья группа устанавливает правила сжатия передаваемых данных путем уменьшения их избыточности. При этом на передающем конце происходит их анализ и упаковка, а на приемном — распаковка в исходный вид. Сжатие позволяет повысить скорость передачи сверх физической пропускной способности канала за счет уменьшения объема реально передаваемых данных. Реализация сжатия также требует некоторых накладных расходов на анализ информации и формирование пакетов; в случае неэффективного сжатия скорость передачи может оказаться ниже скорости физического канала.

Последняя группа протоколов задает правила взаимодействия *DCE* и *DTE*. Они подразделяются на физические, касающиеся кабелей, разъемов и сигналов взаимодействия, и информационные, относящиеся к формату и смыслу передаваемых сообщений. Посредством этих протоколов реализуется общение *DTE* с *DCE* во время подготовки к вхождению в связь, организации вызова и ответа, а также в процессе самого обмена данными.

**Интеллектуальные возможности модемов.** Современное понятие модема значительно шире, чем просто совокупность модулятора и демодулятора. Сейчас модемы являются интеллектуальными устройствами, позволяющими помимо своей главной задачи — преобразования передаваемых сигналов, реализовать множество других функций, предоставляя дополнительные удобства пользователю. Эти модемы называют *интеллектуальными* или *Smart-модемами*. Интеллектуальные возможности модемов реализуются благодаря наличию схемы управления, выполненной на основе того или иного микропроцессора. Каждый конкретный модем может воспринимать определенное множество команд, в общем случае не совпадающее командами, поддерживаемыми другими модемами. Однако для удобства применения модемов и совместимости коммуникационных программ необходимо иметь стандартный набор таких команд.

## Контрольные вопросы

1. Какими основными показателями характеризуется передающее устройство?
2. Приведите структурную схему передатчика с амплитудной модуляцией.
3. Приведите структурную схему передатчика с частотной модуляцией.
4. От чего зависит конструкция и элементная база передатчика?
5. Какой главный эффект достигается применением синтезатора частоты?
6. На каком принципе построены схемы синтезаторов частоты?
7. Почему необходимо суммировать мощности в передатчике?
8. Назовите основные перспективные направления развития передающих систем.
9. Какими показателями характеризуется радиоприемная система?
10. Назовите признаки, по которым классифицируются приемники.
11. Сравните достоинства и недостатки схем приемника прямого усиления и супергетеродинного приемника.
12. Для каких целей в схему приемника вводится АРУ?
13. Приведите схему цифрой АРУ.
14. С какой целью в приемниках применяется АПЧ?
15. На каком принципе основано действие ФАПЧ?
16. Приведите схему ЦФАПЧ.
17. Для каких целей вводится двойное преобразование частоты?
18. Какие функции должен выполнять согласованный фильтр?
19. Как связаны коэффициент передачи согласованного фильтра и спектральная плотность обрабатываемого сигнала?
20. Каким двум условиям должен отвечать согласованный фильтр?
21. Какому условию должно удовлетворять время задержки при обработке оптимальным фильтром известного сигнала?
22. Как строится импульсная характеристика оптимального фильтра?
23. Что представляет собой верность передачи сообщений?
24. Какими показателями характеризуется потенциальная помехоустойчивость непрерывного и дискретного канала связи?
25. От чего зависит пропускная способность канала связи?
26. Какими основными свойствами характеризуются количество информации и энтропия?
27. Чем хорош или плох источник сообщений с большой энтропией?
28. Что такое кодирование и декодирование в непрерывных и дискретных каналах связи? Что общего и чем они отличаются друг от друга?
29. Какое значение имеет теорема кодирования в канале с помехами?
30. Каково назначение кодирования в канале без помех?
31. Как устроен и работает современный модем?
32. Какие способы модуляции используются в модемной связи?
33. Как устроен интерфейс модема и DTE?
34. Как организуют передачу данных посредством модемов?
35. Чем различаются синхронные и асинхронные режимы?
36. Как скорость передачи зависит от параметров канала?

# Глава 10. ЭЛЕМЕНТЫ ВЕЙВЛЕТНОГО И ФРАКТАЛЬНОГО АНАЛИЗА СИГНАЛОВ

В последние годы в радиотехнике широко применяют вейвлетный и фрактальный методы анализа сигналов, а также различные способы обработки информации на их основе.

Основная область применения вейвлетных преобразований — анализ и обработка сигналов, нестационарных во времени или неоднородных в пространстве, когда результаты должны содержать не только общую частотную характеристику сигнала, но и сведения об определенных локальных координатах, на которых происходят быстрые изменения частотных составляющих сигнала. По сравнению с разложением сигналов на ряды Фурье, вейвлеты способны с гораздо более высокой точностью представлять локальные особенности сигналов, вплоть до разрывов 1-го рода. Вейвлет-преобразование одномерных сигналов обеспечивает двумерную развертку, при этом частота и координата рассматриваются как независимые переменные, что дает возможность анализа сигналов сразу в двух пространствах.

В основе концепции фрактальной обработки сигналов лежит некоторая идеализация реального мира — *гипотеза самоподобия* (скейлинга): вид фрактальной структуры объектов существенно не меняется при ее масштабных преобразованиях в определенном диапазоне значений масштаба. В самом простом случае небольшая часть фрактала содержит информацию обо всем фрактале, а это позволяет во много раз сжимать информацию, отражающую изображение. Фрактальные методы сжатия основаны на том, что с помощью коэффициентов системы итерируемых функций изображение представляется в более компактной форме.

## 10.1. Основы вейвлет-анализа сигналов

Важное достоинство преобразования Фурье состоит в возможности единообразного представления разных по форме функций $u(t)$ в виде ряда (или интеграла), составленного из гармонических сигналов с разными частотами. Однако с практической точки зрения и с позиций точного представления произвольных сигналов преобразование Фурье имеет ряд ограничений и недостатков.

Преобразование Фурье дает достаточно простые для расчетов формулы и прозрачную интерпретацию результатов, однако не обладает временны́м разрешением; даже для одной заданной частоты оно требует знание сигнала не только в прошлом, но и в будущем, а это является теоретической абстракцией. Еще

один недостаток — невозможность отразить некоторые локальные особенности функции, связанные с кратковременными резкими изменениями амплитуды или частоты и фазы сигнала. Далее, преобразование Фурье иногда не дает прямой возможности отличить амплитудные спектры некоторых сигналов, если не известны их фазовые спектры. В частности, это относится к спектрам сигналов, один из которых является суммой двух синусоид разной частоты, а другой представляет собой сумму последовательного включения этих синусоид.

**Пример 10.1.** Исследуем и сравним спектры двух разных сигналов. Первый состоит из суммы двух синусоид с разными частотами (рис. 10.1, *а*):

$$u(t) = U_1 \cos(2\pi f_1 t) + U_2 (2\pi f_2 t); \tag{10.1}$$

второй — сумму тех же колебаний, сдвинутых во времени (рис. 10.1, *б*):

$$s(t) = \begin{cases} U_1 \cos(2\pi f_1 t) & \text{при } t \le t_0, \\ U_2 \cos(2\pi f_2 t) & \text{при } t \ge t_0. \end{cases} \tag{10.2}$$

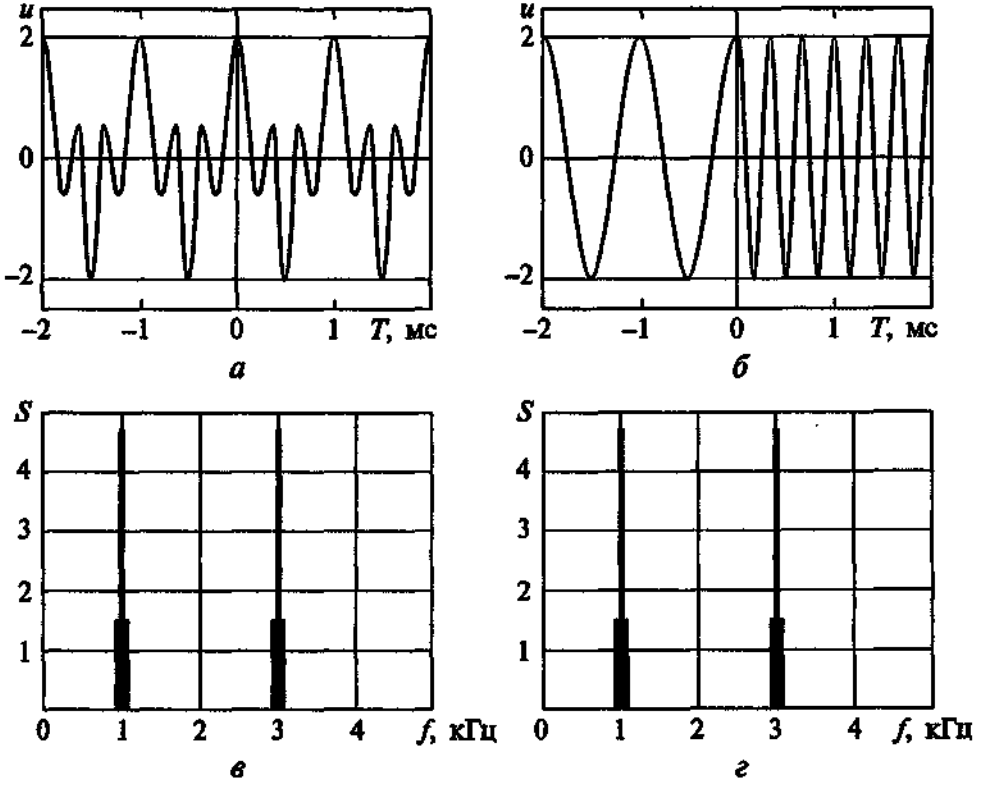



**Рис. 10.1.** К преобразованию Фурье:
*а* — сигнал состоит из суммы двух синусоид с разными частотами;
*б* — сигнал состоит из суммы тех же колебаний, сдвинутых во времени;
*в* — спектр сигнала *а*; *г* — спектр сигнала *б*

Р е ш е н и е. Расчеты показывают, что амплитудные спектры сигналов $u(t)$ и $s(t)$ практически идентичны (рис. 10.1, *в*, *г*). Значит, в рассмотренном примере, несмотря на различие сигналов, идентифицировать их по частотным спектрам, построенным согласно преобразованию Фурье, практически невозможно.

Итак, если сигнал не имеет четкого периодического характера, то, хотя алгоритм преобразования Фурье и остается полностью в силе, его эффективность существенно падает; в частности, не удается сэкономить объем данных за счет перехода от математической модели сигнала во временно́й области к соответствующей модели в частотной области. Обусловлено это тем, что базисной функцией при разложении в ряд Фурье является гармоническое колебание или экспонента, определенные на временно́м интервале от $-\infty$ до $+\infty$. Локальные особенности сигнала (разрывы, ступеньки, пики и т. п.) содержат едва заметные оставляющие амплитудного спектра, по которым обнаружить эти особенности практически невозможно. В этом случае очень сложно и точное восстановление сигнала из-за проявления эффекта Гиббса. Кроме того, на практике не все сигналы стационарны, а для нестационарных сигналов трудности преобразования Фурье возрастают многократно, делая его практически невозможным.

Очевидно, что наиболее эффективным было бы использование не тригонометрических и экспоненциальных, а некоторых локализованных во времени компактных базисов функций, коэффициенты разложения по которым сохраняют информацию о резком изменении параметров анализируемого сигнала. В случае если требуется временна́я локализация спектральных компонент, необходимо обратиться к частотно-временно́му представлению сигнала. Для решения этой задачи в последние годы был предложен новый подход, на базе которого возникло целое направление в теории и технике сигналов, получившее название *вейвлет-анализа*, который является обобщением спектрального анализа на основе преобразования Фурье.

Основы вейвлет-анализа разработали в середине 80-х гг. XX в. Жан Морле (*Jean Morlet*; 1931–2007; французский геофизик) и Александр Гроссман (*Alexander Grossman*; хорватский физик; р. 1930) как альтернативу преобразованию Фурье для исследования сигналов с ярко выраженной неоднородностью.

Занимаясь сейсмологией Ж. Морле обратил внимание на то, что радиоимпульсы, посланные в глубь Земли, имеют при высокой несущей частоте слишком большую длительность отраженных импульсов. Это мешало различить принятые радиоимпульсы, отраженные от тонких, прилегающих друг к другу слоев породы. Поэтому он предложил посылать на высоких частотах короткие специфические волновые образования различной длительности, полученные масштабированием исходной функции. Ж. Морле и А. Гроссман создали базис функций, названных *вейвлетами*. Термин «вейвлет» (от англ. *wavelet* — *вейвлет* — перевод франц. *onde-i lette*) они ввели для анализа свойств сейсмических и акустических сигналов. Вейвлетные преобразования получались в результате растяжения-сжатия и смещения по времени одной *порождающей* (самоподобной, называемой *скейлинг-функцией* — *scaling function*, *scalet*) функции — *гаус-*

*сиана* (гауссов импульс). В отличие от преобразования Фурье, локализующего частоты сигнала, но не дающего временно́го разрешения процесса, вейвлет-преобразование, обладающее самонастраивающимся подвижным частотно-временны́м окном, одинаково хорошо выявляет как низкочастотные, так и высокочастотные характеристики сигнала на разных временны́х масштабах.

С аналитической точки зрения вейвлеты — обобщенное название семейств математических функций определенной формы, которые локальны во времени $t$ и по частоте $f$ и в которых все функции получаются из одной базовой (порождающей) посредством ее сдвигов и растяжений по оси времени. Поэтому сигнал можно анализировать одновременно в физическом (время, координата) и в частотном пространствах. Если сигнал на достаточно большом интервале времени не имеет резких скачков, то вейвлет растягивают по оси $t$ и на нем проводят анализ; при этом существенно экономятся ресурсы обработки. При наличии резкого кратковременного выброса сигнала вейвлет сжимается до интервала этого выброса, и анализ проводится на данном коротком интервале времени. По этой причине вейвлет-анализ часто сравнивают с «математическим микроскопом», вскрывающим внутреннюю структуру существенно неоднородных объектов.

Условно способ многомасштабного, или кратно-масштабного, анализа (*multiresolution analysis*), состоит в том, чтобы взглянуть на исследуемый объект (в данном случае сигнал) сначала с помощью микроскопа, потом — через лупу, затем посмотреть, отойдя на пару шагов, и, наконец, оценить издалека. Глядя в бинокль, микроскоп или кинокамеру и пытаясь уловить максимально четкое и крупное изображение интересующих объектов, во многих других случаях наблюдатель, не сознавая этого, проводит подобный анализ, и даже удивительно, почему он так долго не был математически исследован. Оказалось, что в возможности использования разномасштабного анализа и кроется большой выигрыш в представлении сигнала с помощью вейвлетов.

В вейвлет-терминологии функции базиса вейвлета принято называть *масштабами*, и обычно они обозначаются через $\psi(t)$ или $\psi(x)$. Коэффициенты такого разложения несут важнейшую информацию об эволюции сигнала в частотной и временно́й областях. Для каждой конкретной практической задачи можно подобрать наиболее эффективный вейвлет.

В настоящее время теорию вейвлет-анализа делят на два класса: непрерывный и дискретный. По аналогии с преобразованием Фурье, непрерывный вейвлет-анализ часто называют *непрерывным вейвлет-преобразованием* (НВП; *continuous wavelet transform — CWT*), а дискретный вейвлет-анализ (ДВА; *discrete wavelet transform — DWT*) — *ортогональным*.

### Непрерывный вейвлет-анализ

Основные приложения *непрерывного вейвлет-анализа*: локализация и классификация особых точек сигнала, частотно-временно́й анализ нестационарных сигналов. Например, у таких сигналов, как музыка и речь, спектр радикально меняется во времени, а характер этих изменений — очень важная информация, которая используется при цифровой передаче подобных сигналов по каналам

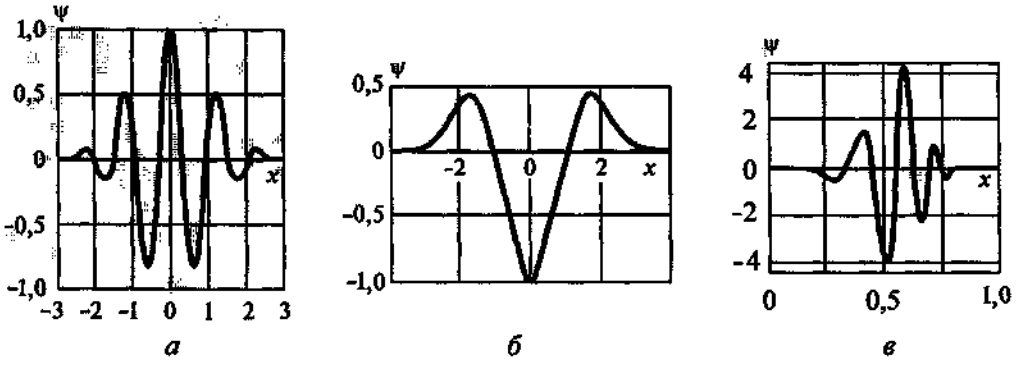



**Рис. 10.2.** Вейвлеты:
*а* — Морле; *б* — мексиканская шляпа; *в* — Добеши

связи. При этом встает вопрос о сокращении избыточности (сжатии) информации. Глобальная задача сжатия информации — сократить ее объем при сохранении приемлемого качества. В этом существенную роль и играет вейвлет-анализ.

Фактически, для того чтобы функция $\psi(t)$ могла принадлежать к классу вейвлетов, необходимо выполнение двух условий:

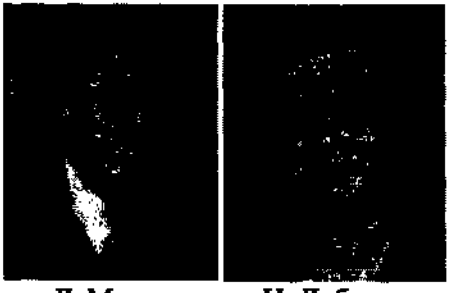
• график этой функции должен быть локален и осциллировать вокруг нуля в окрестности некоторой точки на оси времени, и резко убывать при $t \to \pm\infty$; при этом ее среднее значение (т. е. интеграл по времени) равно нулю

$$\int_{-\infty}^{\infty}\psi(t)dt = 0; \tag{10.3}$$

• норма функции (ее энергия) должна быть конечной

$$\|\psi(t)\| = \sqrt{\int_{-\infty}^{\infty}\psi^2(t)dt} < +\infty. \tag{10.4}$$

В радиотехнике непрерывный вейвлет-анализ часто осуществляют с помощью трех основных функций, показанных на рис. 10.2. Это вейвлеты Морле (*wavelet Morlet*), мексиканская шляпа (*Mexican hat* — часто, *MHAT*-вейвлет; перевернутое «сомбреро») и Добеши (Ингрид Добеши — *Ingrid Daubechies* — одна из основателей теории вейвлет-преобразования; бельгийский математик; р. 1954). График практически любого вейвлета выглядит примерно так же. Необходимо отметить, что, хотя речь идет об аппроксимации непрерывных интегрируемых функций, базисные функции вейвлетов имеют весьма сложный вид. Тем не менее существует много аналитических выражений, корректт-



Д. Морле И. Добеши

но удовлетворяющих условиям вейвлета. В частности, распространены вейвлеты, описываемые формулами:

$$\psi(t) = \sqrt{2}\left[1 - |t|^2\right] e^{-\frac{|t|^2}{2}}, \tag{10.5}$$

$$\psi(t) = \frac{1}{\sqrt{2}}\left[1 - |t|^2\right] e^{-\frac{|t|^2}{2}}. \tag{10.6}$$

С ряда точек зрения для вейвлетов хороши функции, локализованные и во времени и в частотном спектре.

При перемасштабировании результата вейвлет-преобразования произведение временно́го и частотного диапазонов остается постоянным. Оно интерпретируется как площадь ячейки на «частотно-временно́й» (фазовой) плоскости, информацию о которой можно получить с помощью данной функции. Идеальным инструментом анализа сигналов в этом смысле является гауссиан

$$\psi(x) = e^{-x^2/2}. \tag{10.7}$$

Ее преобразование Фурье имеет точно такой же вид «колокола», как и она сама. Вторая производная гауссова импульса и является вейвлетом, называемым вейвлетом *сомбреро* (рис. 10.2, *б*):

$$\psi(x) = (x^2 - 1)e^{-x^2/2}. \tag{10.8}$$

Непрерывное вейвлет-преобразование является сверткой сигнала $u(t)$ с растянутым (сжатым) в $a$ раз вейвлетом $W(x, a)$. При этом вейвлет-спектр зависит от двух временны́х переменных — дрейфовой $x$ и масштабной $a$:

$$W(x,a) = \frac{1}{a}\int_{-\infty}^{\infty} \psi\left(\frac{t-x}{a}\right) u(t)dt. \tag{10.9}$$

Значения вейвлет-коэффициентов $W(x, a)$ показывают, насколько характерный период колебаний $a$ (или частота $a^{-1} = 1/f$) представлен в сигнале в окрестности момента времени $x$. Фактически параметр $x$ является аналогом координаты времени $t$ (т. е. характеризует смещение сигнала во времени), а параметр $a$ — аналогом периода осцилляции (т. е. обратной частоте). При этом вследствие принципа неопределенности частота осцилляций $a^{-1}$ находится с точностью, обратно пропорциональной характерному времени наблюдений $x$ (равному ширине вейвлета). Значит, вейвлет-преобразование $W(x, a)$ одновременно содержит информацию об амплитудах вейвлет-спектра и частотных и временны́х (или пространственных) свойствах сигнала. Это и позволяет изучить сложный сигнал более детально, чем с помощью преобразования Фурье.

Итак, по своему смыслу вейвлет-преобразование функций полностью соответствует преобразованию Фурье. Однако здесь ядром интегрального преобра-

зования вместо экспоненциальной функции $e^{-j\omega t}$ в формуле Фурье (2.54) служит вейвлет $\psi = (t - x)/a$. В отличие от преобразования Фурье, в котором временна́я координата $t$ трансформируется в одну частотную переменную $f$ (или $\omega$), вейвлет-преобразование $W(x, a)$ является двумерным и определяет трехмерную поверхность в пространстве переменных $a$ и $x$.

Заметим, что в зарубежной учебной и технической литературе уже сравнительно давно принято спектр сигнала, определенный по Фурье, называть *single spectrum*, а спектр, полученный на основе вейвлет-преобразования, — *time-scale spectrum*, или *wavelet spectrum*.

### Свойства непрерывного вейвлет-преобразования

Результаты вейвлет-преобразования, как скалярного произведения вейвлета и анализируемого сигнала, содержат комбинированную информацию о сигнале и вейвлете. Получение определенной объективной информации об анализируемом сигнале базируется на свойствах вейвлет-преобразования, общих для вейвлетов всех типов. Рассмотрим основные из этих свойств.

*Линейность*:

$$W[\alpha u_1(t) + \beta u_2(t)] = \alpha W[u_1(t)] + \beta W[u_2(t)] = \alpha W_1(x, a) + \beta W_2(x, a). \qquad (10.10)$$

*Инвариантность относительно сдвига во времени*: сдвиг сигнала во времени на $t_0$ приводит к сдвигу вейвлет-спектра также на $t_0$

$$W[u(t - t_0)] = W(x, a - t_0). \qquad (10.11)$$

*Инвариантность относительно растяжения (сжатия) сигнала*: растяжение (сжатие) сигнала приводит к растяжению (сжатию) его вейвлет-спектра

$$W\left[u\left(\frac{t}{a_0}\right)\right] = \frac{1}{a_0} W\left(\frac{x}{a_0}, \frac{a}{a_0}\right). \qquad (10.12)$$

Это свойство позволяет, в частности, определять наличие и характер особенностей анализируемой функции.

*Дифференцирование*: операция дифференцирования вейвлет-преобразования дает такую формулу:

$$W\left[\frac{\partial^n u(t)}{dt^n}\right] = (-1)^n \int_{-\infty}^{\infty} u(t) \frac{\partial^n [\psi_{xa}(t)]}{dt^n} dt. \qquad (10.13)$$

Отсюда следует, что не важно, что дифференцировать: либо сигнал, либо анализирующий вейвлет.

Если вейвлет задан формулой, то это может быть очень полезным для анализа сигналов. Проанализировать особенности высокого порядка или мелкомасштабные вариации сигнала $u(t)$ с игнорированием крупномасштабных полиномиальных составляющих можно дифференцированием нужного числа раз либо вейвлета, либо самого сигнала. Это свойство особенно полезно, когда сигнал $u(t)$ представляет собой дискретный ряд чисел.

**Способы графического представления результатов вейвлет-анализа**

Вейвлет-преобразование одномерного сигнала преобразует функцию одной переменной в набор вейвлет-коэффициентов, который представляет собой функцию двух переменных — масштаба $x$ и смещения $a$. Если при вычислении преобразования эти параметры изменяются в достаточных пределах, коэффициенты $W(x, a)$ заключают в себе полную информацию об исходном сигнале. Вейвлет-спектр $W(x, a)$ представляет собой некоторую поверхность в трехмерном пространстве, поэтому способы ее визуализации могут быть различны. Вместо изображения трехмерной поверхности спектра вейвлет-преобразования часто удобнее изобразить эту поверхность в виде проекции на плоскость параметров $x$ и $a$, получая при этом:

- карту проекций на плоскость с изображением изоуровней или изолиний функции $W(x, a)$ на плоскость $(x, a)$, позволяющих проанализировать изменение интенсивности амплитуд вейвлет-преобразования на различных частотных и временны́х масштабах;
- карты линий локальных экстремумов поверхности $W(x, a)$, названных *скелетонами* (от англ. *sceleton* — скелет, каркас), линиями, которые показывают распределение точек локальных экстремумов вейвлет-спектра, четко выявляющих структуру анализируемого сигнала;
- зависимости коэффициентов вейвлет-преобразования для выбранного масштаба преобразования от времени.

Наиболее часто результаты вейвлет-преобразования анализируемого сигнала графически изображают проектируя линии постоянного уровня поверхности $W(x, a)$ на плоскость с переменными параметрами: сдвига (по оси абсцисс) и масштабом (по оси ординат) с градиентной заливкой оттенками серого цвета между линиями. На рис. 10.3 при графическом вейвлет-представлении применен вариант заливки, при котором темные области соответствуют отрицательным, а светлые — положительным значениям вейвлет-преобразования $W(x, a)$; при этом оттенками серого цвета в каждой из областей выделены диапазоны изменения значений $W(x, a)$.

На приведенных на рис. 10.3 графиках $W(x, a)$ показаны результаты вейвлет-преобразования ряда тестовых сигналов, представленных на конечном интервале времени. В качестве анализирующего вейвлета использован *MHAT*-вейвлет. Верхняя часть рисунка — исследуемый сигнал, нижняя — изолинии поля $W(x, a)$ спектра. Изображены как линии локального максимума, так и минимума. В седловых точках поверхности происходит слияние линий.

Сигнал, показанный на рис. 10.3, *а*, является гармоническим колебанием. Картина изолиний уровня (рис. 10.3, *б*) указывает на периодический характер сигнала и в мелкомасштабной области представляет регулярную систему ячеек с поочередно повторяющимися положительными (светлые области) и отрицательными (темные области) значениями поля $W(x, a)$. Граница раздела ячеек совпадает с положением нулей $u(t)$. Вертикальный размер ячеек определяет масштаб, отвечающий периоду колебаний.

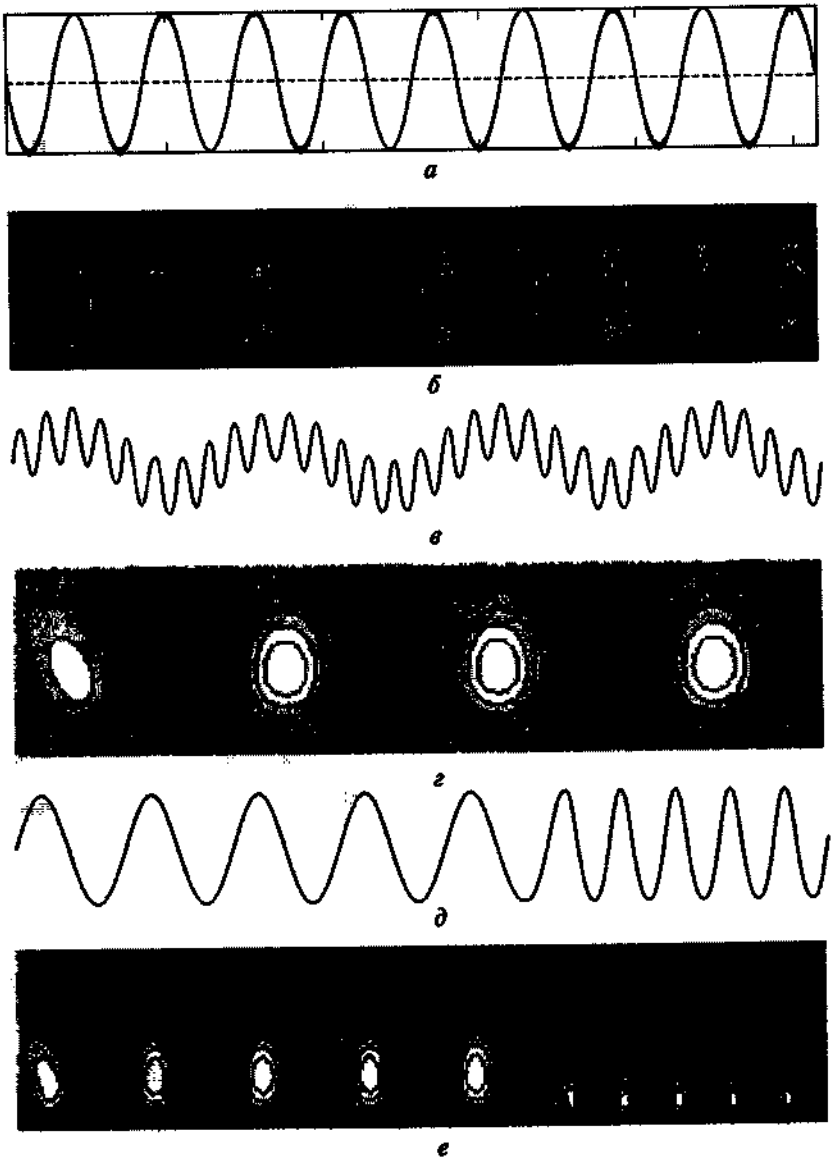


**Рис. 10.3.** Принцип вейвлет-анализа сложного сигнала:
*а, б* — гармонический сигнал и его вейвлет-спектр; *в, г* — сигнал в виде суммы двух гармонических колебаний с разными частотами и его вейвлет-спектр; *д, е* — сигнал в виде суммы тех же колебаний, включенных последовательно, и его вейвлет-спектр

На рис. 10.3, *в, г* приведены результаты расчета вейвлет-спектра сигнала, являющегося суммой двух синусоид с разными частотами, а на рис. 10.3, *д, е* — для варианта их последовательного включения. Как было показано в примере 10.1, с точки зрения амплитудного Фурье-спектра, эти два сигнала практически не различимы. Вейвлет-анализ выявляет два характерных масштаба, периодичность следования ячеек смены знака указывает на периодический характер сиг-

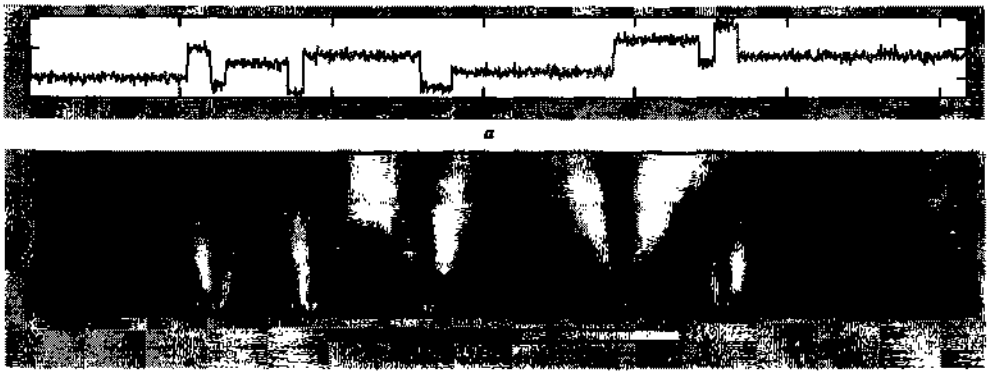


**Рис. 10.4.** Вейвлет-преобразование сигнала сложной формы:
*а* — сигнал; *б* — вейвлет-спектр

налов, их пространственное распределение показывает специфику каждого из сигналов. Постоянство расстояний между линиями экстремумов и их параллельность свидетельствуют, что составляющие сигнала — два гармонических колебания разных частот.

Рассмотренные графические результаты вейвлет-преобразования относятся к сигналам простой формы и не дают полного представления об этом виде анализа. На рис. 10.4 приведен пример реального сигнала сложной формы и графического отображения поверхности его вейвлет-преобразования. Вид поверхности определяет изменения во времени спектральных компонент различного масштаба и называется *частотно-временны́м спектром*. Для расширения диапазона масштабов можно применять логарифмическую шкалу. В общем случае вейвлет-спектр (вейвлет-коэффициенты) сложного сигнала, принимающего вещественные значения, на плоскости с координатами $(a, x)$ можно достаточно условно представить как *структуру горного хребта* разной ширины по параметру $a$ и бесконечной длины по относительной временно́й координате $x$ (рис. 10.4, *б*). Амплитудой спектра является текущая величина $W(x, a)$. Процедура перехода структуры осцилляции одной формы сигнала к другой выглядит как слияние или расщепление этих *горных кряжей* вдоль координаты $a$.

Наиболее наглядно результаты (трехмерная поверхность) непрерывного вейвлет-преобразования (вейвлет-спектр) реального сигнала представлены на рис. 10.5.

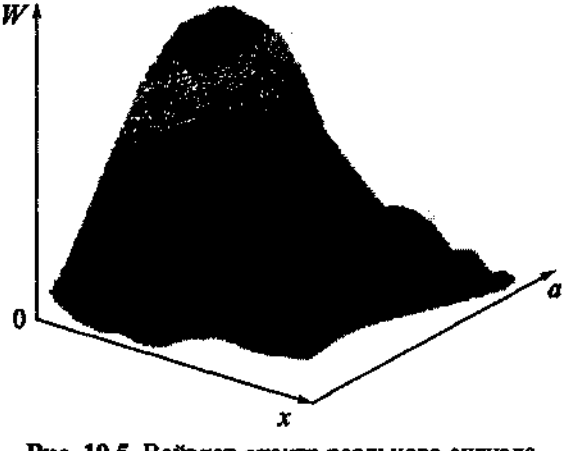



**Рис. 10.5.** Вейвлет-спектр реального сигнала

По аналогии с обратным преобразованием Фурье можно представить алгоритм восстановления (синтеза) исследуемого сигнала с помощью обратного вейвлет-преобразования по вейвлет-коэффициентам и базису вейвлетов:

$$u(t)=\frac{1}{\|\psi(t)\|}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{a^2}W(a,x)\psi(t)dadx, \qquad (10.14)$$

где $\psi(t)=(t-x)/a$ — вейвлет.

Данная формула окончательно устанавливает сходство непрерывного вейвлет-анализа и метода интегрального преобразования Фурье.

### Дискретный (ортогональный) вейвлет-анализ

Непрерывное *W*-преобразование может быть проведено аналитически лишь для простейших функций, а вычисление его на компьютерах при помощи квадратур является весьма трудоемким. Отметим также, что базисы на основе непрерывных вейвлетов, как правило, не являются строго ортонормированными, поскольку элементы базиса бесконечно дифференцируемы и экспоненциально спадают на бесконечности. Поэтому в приложениях обычно используется дискретный вариант, который при специальном выборе базисных функций может быть выполнен крайне эффективно и без дополнительных затрат памяти (вектор коэффициентов рекурсивно замещает вектор исходных значений). Дискретное *W*-преобразование может выполняться аппаратно на специализированных процессорах, как и быстрое преобразование Фурье.

Дискретным (ортогональным) вейвлет-анализом называют представление *дискретных сигналов* в виде обобщенного ряда Фурье по системе ортогональных базисных функций, формирующихся из некоторого исходного (порождающего) вейвлета $\psi(t)$. Применяется ортогональный вейвлет-анализ в основном для сжатия данных (информации) и подавления шумов.

В науке о вейвлет-анализе естественным образом возникают музыкальные термины — набор значений вейвлет-преобразования при фиксированном масштабе $a$ называют *голосом*, диапазон масштабов (или, что то же самое, частот) от $a$ до $2a$ — *октавой*. Согласно музыкальной классике, на каждую октаву следует брать, по крайней мере, 12 голосов. Так, если сигнал имеет скромную длину 2 048 цифровых слов, то 8 октав займут 196 608 слов. Избыточность такого представления информации огромна, и она занимает большой объем памяти цифровой техники, времени обработки и каналов связи при передаче сигнала на расстояние. Поэтому ставится задача сжатия (компрессии) данных, причем допускаются несущественные потери информации при «разжатии». Идея состоит в том, чтобы запомнить вместо полного сигнала его «детали» на каждом из масштабных уровней, а также самую грубую версию сигнала. Очень быстро стало ясно, что для вполне точного восстановления исходного сигнала достаточно знать вейвлет-преобразование на некоторой довольно редкой решетке в плоскости координат $(x, a)$. Следовательно, и вся информация о сигнале содержится в этом сравнительно небольшом наборе значений.

Дело в том, что в реальных сигналах, несущих информацию об изображениях, резкие скачки на малых масштабах встречаются редко. Большую часть этих величин в цифровой форме сигнала можно заменить нулями (на восстановленном изображении это почти не скажется), а остальные записать более короткими цифровыми словами, чем основной сигнал. Такой массив можно довольно сильно сжать обычными, обратимыми методами.

Упрощенно идею многомасштабного дискретного вейвлет-анализа можно трактовать так: «большое» пространство «всех анализируемых сигналов» надо заполнить растянутыми или сжатыми копиями некоего «эталонного» пространства, порожденного элементарным сигналом $\psi(x)$ и его сдвигами по оси времени. По определению это эталонное пространство принимается за набор сигналов, которые можно представить с разрешением 1.

Рассмотрим пример многомасштабного дискретного вейвлет-анализа, показанный на рис. 10.6. Пусть функция $\varphi(x)$ — исходный единичный заштрихованный «столбик» на интервале [0, 1]; его сдвиги — аналогичные столбики на интервалах ..., [−2, −1], [−1, 0], [1, 2],... (рис. 10.6, *а*) Пространство сигналов (анализируемый сигнал $u(t)$) состоит из «лесенок», полученных умножением столбиков на определенные числа. Остальными пространствами сигналов служат перемасштабированные копии наборов столбиков. Одно из них получено растяжением вдвое (в примере — лесенки из столбиков удвоенной ширины — рис. 10.6, *б*), второе — в четыре раза, третье — сжатием в два раза исходного фиксированного сигнала $\varphi(x)$ (столбики половинной ширины), и так для всех положительных и отрицательных степеней двойки. Очевидно, что, уменьшая ширину столбиков, можно аппроксимировать ими практически любой реальный сигнал.

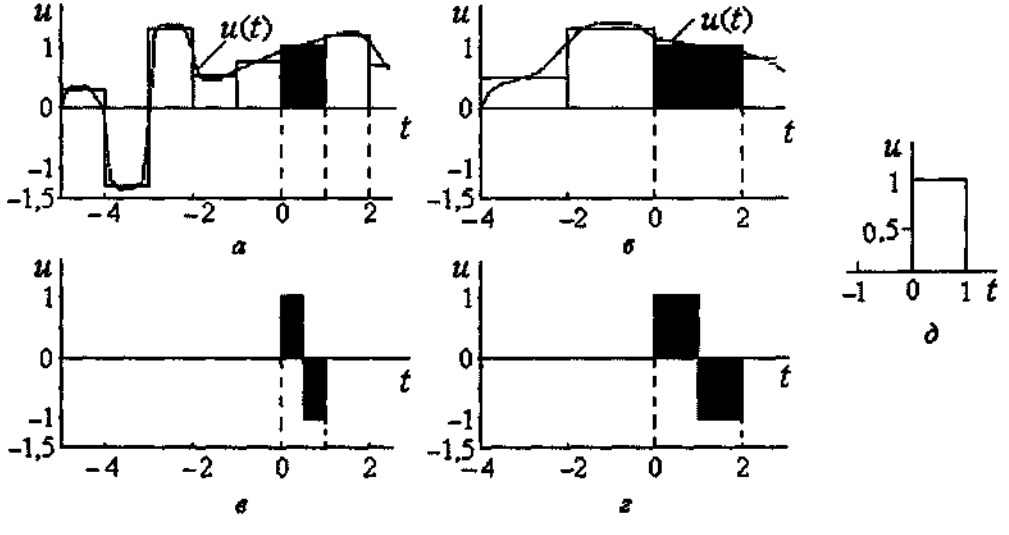



**Рис. 10.6.** К выводу уравнения самоподобия:
*а* — сигнал с разрешением 1; *б* — сигнал с разрешением 2; *в* — порождающий вейвлет Хаара; *г* — растянутый в два раза вейвлет Хаара; *д* — функция Хаара

Для сигнала $\varphi(x)$ необходимо, чтобы сдвиги $\varphi(x-1)$, $\varphi(x-2)$, ..., $\varphi(x-k)$, ... были друг другу ортогональны (т. е. некоррелированны, что в данном примере, безусловно, выполняется). Сигналы, полученные с разрешением 2, можно представить и с разрешением 1 (достаточно столбик удвоенной ширины разделить на два одинаковых столбика единичной ширины); в общем случае, растянутая вдвое скейлинг-функция должна выражаться через сдвиги исходной скейлинг-функции. Этот факт выражается фундаментальным уравнением:

$$\varphi(x) = \sum_{k=-\infty}^{\infty} h_k \varphi(2x-k), \tag{10.15}$$

где $h_k$ — некоторые коэффициенты.

Это *функциональное уравнение самоподобия*, или *скейлинговое уравнение*, и есть «уравнение колебаний» дискретного (ортогонального) вейвлет-преобразования. Для рассматриваемого примера $h_k = 1$ и уравнение имеет вид

$$\varphi(x) = \varphi(2x) + \varphi(2x-1). \tag{10.16}$$

Графически столбик удвоенной ширины есть сумма двух соседних столбиков единичной, исходной ширины (рис. 10.5, *б*).

Оказывается, что если для функции $\varphi(x)$ выполняется уравнение (10.15) и ее сдвиги *ортогональны* друг другу, то появляется очень эффективный способ представления сигналов. А именно, существует еще одна функция — обозначим ее как $\psi(x)$, обладающая следующими свойствами:

- она выражается через сдвиги исходной функции $\varphi(x)$ по формуле, аналогичной (10.15), но с другими значениями коэффициентов $h_k$;
- ортогональны друг другу и сдвиги функции и все ее перемасштабированные версии;
- функция является вейвлетом.

В рассматриваемом примере у функции $\varphi(x)$ коэффициенты $h_k$ равны 1 и 1, а у $\psi(x)$ соответственно 1 и –1.

Ортогональные вейвлеты замечательны в первую очередь тем, что существует очень быстрый алгоритм разложения по ним любого сигнала. Сам вейвлет показан на рис. 10.6, *в* и *г*. Система его сдвигов и двоичных растяжений и сжатий — широко известный базис на основании функции Альфреда Хаара (*Alfred Haar*; венгерский математик; 1885–1933), разработанной им еще в 1909 г. Хаар предложил использовать в математике специфическую функцию — импульс единичной амплитуды, сдвинутый по оси времени (рис. 10.6, *д*):

$$\varphi(t) = \begin{cases} 1 \text{ при } & -t \le 0, \quad t-1 < 0; \\ 0, & \text{иначе.} \end{cases} \tag{10.17}$$

Из функции Хаара и был создан вейвлет, напоминающий зигзагообразную прямоугольную волну и в общем виде описываемый формулой

$$\psi(t) = -\varphi(2t) + \varphi(2t-1). \tag{10.18}$$

Аналитически данную функцию вейвлета Хаара (*HAAR*-вейвлет) нагляднее представить в относительном времени $x$:

$$\psi(x)=\begin{cases}1, & 0\le x<0{,}5;\\ -1, & 0{,}5\le x<1;\\ 0, & x<0,\ x\ge 1.\end{cases} \tag{10.19}$$

Этот вейвлет имеет резкие границы во временно́й области и, как следствие, бесконечно осциллирующие, но убывающие «хвосты» в частотной области. Кроме того, *HAAR*-вейвлет является несимметричным.

Полный ортонормированный базис вейвлетов Хаара имеет вид

$$\psi_{jk}(x)=2^{j/2}\psi(2^j x-k), \tag{10.20}$$

где $2^{j/2}$ — некоторая постоянная; $j$, $k$ — положительные и отрицательные целые числа, включая нуль.

Нетрудно убедиться в том, что функции $\psi_{jk}(x)$ взаимно ортогональны:

$$\int_{-\infty}^{\infty}\psi_{jk}(x)\,\psi_{mn}(x)dx=0, \tag{10.21}$$

если одновременно не выполняются равенства $j=m$ и $k=n$.

Действительно, если $j\ne m$, но $k=n$, то один из вейвлетов целиком размещается на том отрезке оси $t$, где другой вейвлет постоянен. Это ведет к обращению в нуль интеграла (10.21). Если $j=m$, а $k\ne n$, то хотя бы один из сомножителей в подынтегральном выражении (10.21) равен нулю. Кажется, что число функций в этом базисе в четыре раза больше, чем в известной записи обобщенного ряда Фурье, так как в последнем фигурируют только положительные целочисленные индексы. На самом деле их ровно столько же; просто использована другая нумерация бесконечной последовательности чисел.

Построив ортонормированную базисную систему вейвлетов Хаара, можно представить произвольный сигнал $u(t)$ в виде обобщенного ряда Фурье:

$$u(t)=\sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty}2^{j/2}c_{jk}\psi(2^j x-k). \tag{10.22}$$

Согласно (10.22), коэффициенты ряда Фурье $c_{jk}$ являются скалярными произведениями анализируемого сигнала и соответствующей базисной функции:

$$c_{jk}=\int_{-\infty}^{\infty}u(t)2^{j/2}\psi(2^j x-k)dt. \tag{10.23}$$

Формально обобщенный ряд Фурье вида (10.22) отличается от изучавшихся ранее тем, что суммирование проводится не по одному, а по двум индексам. Это обстоятельство несущественно, поскольку и одноиндексная и двухиндексная системы в равной мере принадлежат одному и тому же классу бесконечных счетных, т. е. перечислимых множеств.

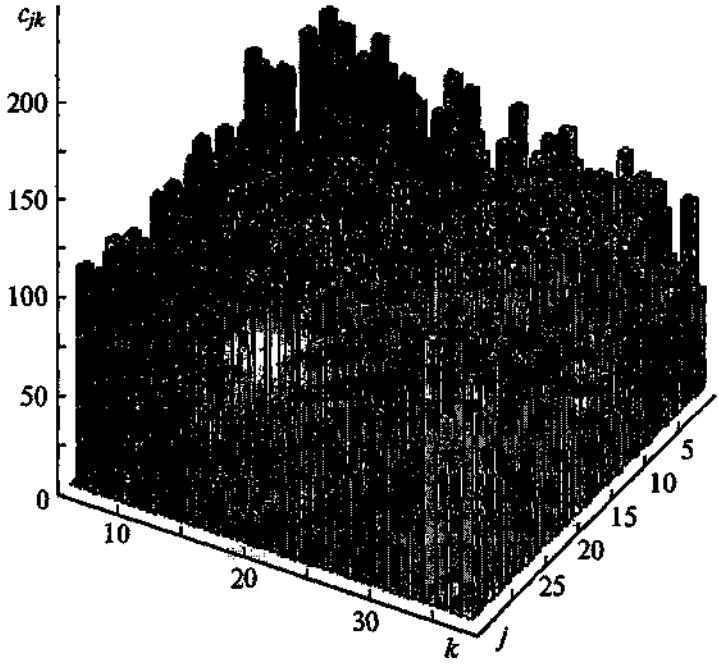



**Рис. 10.7.** Коэффициенты дискретного вейвлет-анализа

В достаточно образной форме коэффициенты дискретного вейвлет-анализа (вейвлет-спектр) сигнала, принимающего вещественные значения, можно представить как «лес» из вертикальных прямоугольных столбиков, размещенных на *jk*-плоскости в точках с целочисленными координатами. При этом координата *j* указывает на скорость изменения сигнала (отражает частоту спектра), а координата *k* — на положение вдоль оси времени (рис. 10.7).

Отметим интересные аналогии. Ортогональные вейвлеты реализуют специфическое покрытие плоскости «время — частота». Другими словами, сигнал раскладывается по октавам, без промежуточных голосов, т. е. ортогональный вейвлет-анализ является предельно экономным разложением, в противоположность непрерывному вейвлет-преобразованию, учитывающему вклад от каждой точки на плоскости время – частота.

### Сжатие видеоинформации на основе дискретного вейвлет-преобразования

Простейший алгоритм сжатия видеоинформации (телевизионных и прочих изображений) при помощи вейвлет-преобразования можно представить так:

- выполнить вейвлет-преобразование;
- упорядочить вейвлет-коэффициенты;
- отбросить «хвост» упорядоченного массива, энергия которого равна допустимому (по условиям задачи) значению;
- запомнить оставшиеся коэффициенты и их положение в исходном массиве;
- при восстановлении заменить отброшенные коэффициенты нулями.

Подобное сжатие информации используют не только для ее хранения или передачи по каналам связи, но и для решения задач анализа и распознавания. При этом значительно сокращается объем вычислений распознающей системы, что ускоряет обработку.

**Вейвлетный (рекурсивный) алгоритм сжатия информации**. В радиотехнике и теории и технике связи вейвлеты позволили разработать высокоэффективные методы сжатия различной видеоинформации, например телевизионных изображений. Такие изображения — своеобразный тип данных, характеризуемый следующей особенностью: они требуют намного больше цифровой памяти, чем текст. Так, скромная, не очень качественная иллюстрация на обложке книги размером $500 \cdot 800$ точек занимает 1,2 Мб — столько же, сколько художественная книга из 400 страниц (60 знаков в строке, 42 строки на странице).

Избыточность изображения. Поток данных об изображении имеет излишнюю информацию, которая может быть устранена практически без визуальной заметности его изменений. Существует два типа избыточности:

- статистическая избыточность, связанная с корреляцией и предсказуемостью данных; как правило, соседние точки как по горизонтали, так и по вертикали в изображении близки по цвету; кроме того, можно использовать подобие между цветовыми плоскостями $R$, $G$ и $B$; эту избыточность можно устранить без потери информации, исходные данные при этом можно полностью восстановить;
- визуальная (субъективная) избыточность, которую можно устранить с частичной потерей данных, мало влияющих на качество воспроизводимых изображений; человеческое зрение при анализе изображения оперирует контурами, общим переходом цветов и сравнительно нечувствительно к малым изменениям в изображении; таким образом можно создать эффективные алгоритмы сжатия, в которых восстановленное изображение не будет совпадать с оригиналом, однако человек этого не заметит.

Наиболее простая идея алгоритма сжатия заключается в том, что в файл записывается разница между средними значениями соседних пикселов в изображении, которая обычно принимает значения, близкие к 0. Так, два числа $a_i$ и $a_{i+1}$ всегда можно представить в виде:

$$p_i = (a_i + a_{i+1})/2; \qquad q_i = (a_i - a_{i+1})/2. \tag{10.24}$$

Полезный эффект такого представления заключается в следующем — если в сигнале присутствуют значительные корреляции между соседними отсчетами, то полуразность $q$ мала (в пределе $a = p$, $q = 0$) и может быть представлена в компьютере меньшим числом битов. При этом не происходит потери информации, поскольку имеется обратное преобразование:

$$a = p - q, \qquad b = p + q. \tag{10.25}$$

Отметим, что в рассматриваемой паре отсчетов $a$ и $b$ значение $a$ может рассматриваться, как прогноз для следующего отсчета $b$, для уточнения которого используется коэффициент разложения $q$.

Если выполнить описанное преобразование для всех соседних пар отсчетов сигнала длины $2^n$, то он делится на две дискретные последовательности отсчетов половинной длины каждая

$$p_n = (a_n + a_{n+1})/2; \qquad q_n = (a_n - a_{n+1})/10. \tag{10.26}$$

Вектор средних значений отсчетов $p_n$ можно рассматривать как огрубленное (но *сжатое*!) представление исходного вектора $a_n$ сигнала, а вектор разностей $q_n$ — как детализирующую информацию о сигнале, необходимую для обратного перехода из сжатого представления к исходному. Затем такое же преобразование можно применить к уже сжатому сигналу $p_n$, переходя к еще более компактному и огрубленному представлению исходного сигнала. Если рекурсивно выполнить преобразование $n$ раз, то получим из исходного сигнала $n$ его версий с огрублением на разных масштабах. Самое грубое представление $p_0$ — это просто среднее значение всего исходного сигнала.

Описанная процедура обработки информации и отражает дискретный вейвлет-анализ по базису вейвлетов Хаара (*Haar transform*) — разложение сигналов по ортонормальному базису вейвлетов Хаара (10.20). Полное число коэффициентов разложения равно исходному числу отсчетов $2^n$, однако их информационная нагрузка (и, соответственно, требуемая точность представления) различна. Обратное преобразование выполняется также рекуррентно, по обратным формулам. Полное число выполняемых операций пропорционально числу преобразований $n$ (что эффективнее, чем быстрое преобразование Фурье, требующее, как известно, около $n\log n$ операций).

**Пример 10.2.** Необходимо сжать строку изображения из восьми значений яркости пикселов ($a_i$): (120; 111; 112; 118; 117; 114; 110; 102).

Р е ш е н и е. Согласно формуле (10.26), получим следующие последовательности коэффициентов $p_i$ и $q_i$ алгоритма сжатия изображения: (115,5; 115; 115,5; 106) и (4,5; –3; 0,5; 4). Заметим, что значения $q_i$ достаточно близки к 0. Повторим операцию, рассматривая уже $p_i$ как $a_i$. Данное действие выполняется как бы рекурсивно, откуда и название алгоритма. Теперь получим из (115,5; 115; 115,5; 106): (115,25; 110,75), (0,25; 4,75). Эти коэффициенты, округлив до целых, можно поместить в файл, как сжатую информацию. Заметим, что здесь применено преобразование к цепочке пикселов только два раза. Реально же можно применять вейвлет-преобразования 4…6 раз. Эти приемы позволяют достичь заметных коэффициентов сжатия информации.

**Пример 10.3.** Пусть из архивированного файла информации (см. пример 10.2) восстановлена цепочка значений (115; 111); (0; 5); (6; –3; 1; 4). Постройте строку изображения из восьми значений яркости пикселов, которую воссоздаст алгоритм вейвлетного сжатия.

Р е ш е н и е. В качестве примера приведем представление вектора сигнала из четырех элементов по базисным векторам (векторным вейвлетам) Хаара

$$u=\begin{bmatrix}10\\1\\4\\0\end{bmatrix}=3\begin{bmatrix}1\\1\\1\\1\end{bmatrix}+2\begin{bmatrix}1\\1\\1\\-1\end{bmatrix}+4\begin{bmatrix}1\\-1\\0\\0\end{bmatrix}+1\begin{bmatrix}1\\0\\-1\\-1\end{bmatrix}.$$

Согласно этому разложению, значение вектора каждого исходного элемента равно сумме векторных составляющих по соответствующей строке базиса.

Дискретное вейвлет-преобразование Хаара имеет по меньшей мере два ценных практических свойства, определяющие перспективность применения их в радиотехнике, для систем связи и сжатия видеоинформации:

- при передаче и хранении больших объемов информации степень ее сжатия и детали представления можно варьировать в широких пределах, обнуляя (не передавая и не храня) разностные коэффициенты $q_n$, начиная с требуемого уровня подробности;
- передавать сигнал можно последовательно — фрагментами (кадрами, окнами), начиная с самых грубых его уровней, несущих в предельно компактной форме основной объем информации, с дальнейшим последовательным уточнением, вплоть до полного восстановления. На достаточном уровне подробности передачу данного фрагмента можно прекратить и перейти к следующему фрагменту. При поиске информации в базе данных скорость доступа можно увеличивать за счет снижения подробности представления.

### Практические приложения вейвлет-анализа

Вейвлеты позволяют чрезвычайно эффективно применять фундаментальную математическую теорию к решению важных задач в повседневной жизни. Коротко остановимся на некоторых приложениях вейвлет-анализа в радиотехнике и теории передачи информации.

**Распознавание речевых сигналов.** Физический механизм производства звука голосовыми связками определяет базовые характеристики этого класса сигналов. В рамках известной модуляционной модели реальный речевой сигнал представляется в виде суммы:

$$u(t)=\sum_{k=1}^{N}E_k\cos\varphi_k(t)+x(t), \qquad (10.27)$$

где $E_k$ — амплитуда $k$-й первичной гармонической компоненты речи; $\varphi_k(t)$ — ее мгновенная фаза (для упрощения пусть начальные фазы равны нулю); $N$ — число гармонических составляющих; $x(t)$ — вклад шума и ошибок моделирования.

Установлено, что длительное произношение односложного звука близко к чистым музыкальным тонам (гармоникам), т. е. прямое преобразование Фурье оптимально описывает этот процесс (амплитуда $E_k$, мгновенная фаза $\varphi_k(t)=\omega_k t$ и круговая частота $\omega_k$ — константы). Реально же в живой речи человека происхо-

дит быстрый переход от тона к тону с характерными шумами, высотой голоса и прочее, что приводит к необходимости распознавания зависимости от времени $t$ амплитуд и фаз первичных компонент и определения на этой основе, какое слово было произнесено. Причем схема распознавания звуков должна быть достаточно жесткой, нечувствительной к специфическим шумам, национальности или диалекту говорящего и т. д.

Таким образом, при квазинепрерывном вейвлет-преобразовании используется временно́й параметр $x$ и дискретный набор точек $a_j = 2^j a_0$, где $a_0$ — исходная координата; $j = 0, 1, 2,...$ На плоскости с координатами $(a, x)$ чистые тона сигнала человеческой речи имеют вид периодического горного хребта конечной ширины по параметру $a$ и бесконечной длины по относительной временно́й координате $x$. Процедура перехода от одной суммы нескольких тонов к другой выглядит как слияние или расщепление этих горных кряжей вдоль координаты $a$. Однако эта схема чувствительна к шуму и к интерференции близких по частотам тонов. Во избежание этого приходится не ограничиваться только информацией о вершинах хребтов, а применять процедуру «синхронного сжатия» хребта в линию путем специального усреднения значений вейвлет-преобразования $W(x, a)$ по $a$ около максимумов. Это дает нечувствительный к шуму механизм определения мгновенных фаз. Для вычисления мгновенных амплитуд применяют еще одну аналогичную оптимизационную процедуру, используя уже известную информацию о $\varphi_k(t)$. При этом для получения превосходных результатов отнюдь не приходится отказываться от методов и техники, разработанных независимо от вейвлет-преобразования.

**Восстановление зашумленных сигналов.** Пусть имеется $n$ наблюдений $y_i$ сигнала $u_i(t)$ на отрезке [0, 1], содержащих «белый» шум

$$y_i = u_i(t) + \sigma_i, \tag{10.28}$$

где $\sigma_i$ — напряжение шума; $i = 1, 2, ..., n$.

Для нелинейного подавления шума при помощи вейвлет-преобразования последовательно применяются следующие процедуры:

- прямое вейвлет-преобразование;
- обнуление незначимых коэффициентов преобразования по уровню порога, пропорционального амплитуде шума;
- обратное вейвлет-преобразование.

На рис. 10.8 приведены результаты подавления шума в некотором ступенчатом сигнале на основе метода Фурье, сплайн-методов (сплайны по природе очень близки вейвлетам и до появления последних удельный вес сплайнов в прикладных задачах был значительно выше; основной их недостаток — присущий и вейвлетам Добеши — наличие корреляции между степенью сплайна и степенью гладкости аппроксимируемой функции) и вейвлет-преобразования.

Данный способ нелинейного (порогового) удаления шума заключается в том, что ортогональное вейвлет-преобразование «сжимает» сигнал до небольшого числа относительно больших коэффициентов. Поскольку «белый» шум при любом ортогональном преобразовании сохраняет свою структуру и ам-

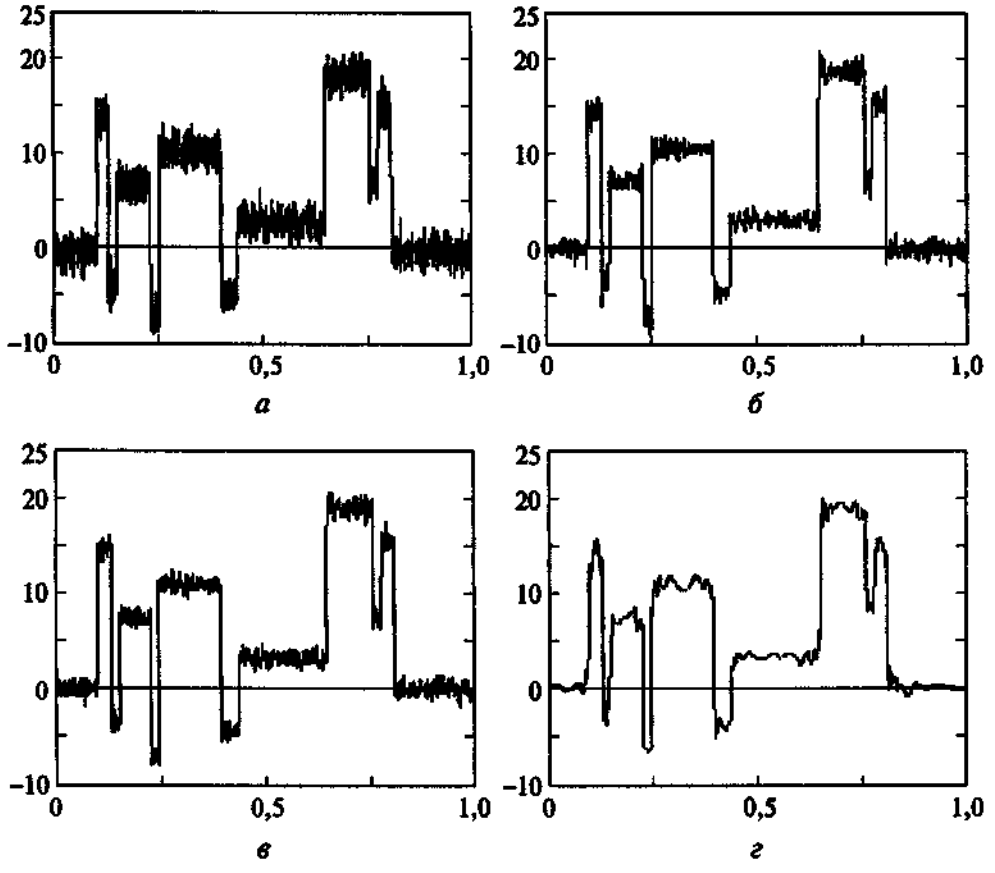



**Рис. 10.8.** Нелинейные методы подавления шума:
*а* — зашумленный ступенчатый сигнал; *б* — на основе метода Фурье;
*в* — сплайн-сглаживание; *г* — фильтрация при помощи вейвлет-преобразования

плитуду и его вейвлет-коэффициенты постоянны, то пороговое обрезание несущественных коэффициентов вейвлет-преобразования сильно понижает шум, не влияя на структуру сигнала.

**Эффективный анализ сильно осциллирующих сигналов.** Выделение этих сигналов на фоне других и т. д.

**Исследование электромагнитных явлений.** Вейвлеты представляются удобным и естественным инструментом для исследований электромагнитных явлений, поскольку уравнения Максвелла (как и вейвлеты) инвариантны относительно перемещений и изменений масштаба. При этом формализм вейвлет-преобразования позволяет сформулировать физические принципы волновых явлений и вопросы обработки сигналов на едином языке.

Именно благодаря выявлению локальных особенностей сигнала, принципиально отсутствующим у преобразования Фурье, вейвлет-преобразование нашло широкое применение для анализа тонкой структуры сигналов и изображений, их сжатия и очистки от шума, что важно и полезно в радиотехнике, электронике, гидроакустике, геофизике, медицине и других областях науки и техники.

При этом стоит отметить, что вейвлет-преобразование ни в коем случае не является заменой традиционного преобразования Фурье и не умаляет его достоинств и значимости при работе со стационарными процессами. Вейвлет-преобразование просто иное и позволяет посмотреть на исследуемый процесс с другой точки зрения.

**Ближайшие перспективы применения вейвлетов.** Что же приносит нового многомасштабный анализ сигнала?

1. Подобный анализ позволяет выявлять локальные (нестационарные) особенности сигнала и классифицировать их по каким-либо специфическим признакам, например по интенсивности. В частности, в обработке изображений широко распространена многомасштабная локализация резких границ (*multiscale edge detection*). Очень резкие перепады яркости заметны и на малых, и на больших масштабах. В некоторых задачах можно считать их наиболее информативной частью изображения и оценивать с большой точностью, пренебрегая всем остальным.

Часто в задачах подобного типа важнее найти не сами разномасштабные версии сигнала, а различия между ними, детали, которые исчезают при переходе от более тонкого масштаба к более грубому.

2. При таком анализе удается проследить динамику изменения сигнала вдоль «оси масштабов». Если резкие скачки во многих случаях можно увидеть невооруженным глазом, то взаимодействие событий на мелких масштабах, перерастающее в крупномасштабные явления, заметить очень сложно. Многомасштабный анализ помогает количественно охарактеризовать эту неоднородность. Скачки динамики по «масштабной переменной» могут нести не менее важную информацию, чем резкие изменения по времени или пространству.

В последнее время база приложений вейвлет-преобразований быстро расширяется. Это происходит как вследствие типизации постановок задач, так и благодаря появлению пакетов программ и интегрированных компьютерных сред для вейвлет-анализа. В частности, широкое распространение получила библиотека программ вейвлет-анализа для пакета *MATLAB*.

Большое количество современных приложений требуют обработки сигналов (функций) в реальном масштабе времени. В связи с этим перспективы приложений вейвлет-анализа связываются с разработкой и производством специализированных (сигнальных) микропроцессоров для *быстрого вейвлет-преобразования* (*Fast Wavelet Transform*). Такие (параллельные на каждом масштабе разрешения сигнала) микропроцессоры используют весьма ограниченный набор операций и достаточно малую локальную память, что обеспечит возможность быстрых стандартизованных вычислений, компактность и дешевизну разрабатываемых устройств. В ближайшее время начнется массовое использование вейвлет-анализа для сжатия и передачи видеоинформации в диалоговых телекоммуникационных системах (видеоконференций, дистанционного образования, компьютерных музеев, систем виртуальной реальности, стереотелевидения и т. п.).

Бурный прогресс в области фундаментальных исследований и теории вейвлетов может быть связан с важными биологическими аналогиями между коди-

рованием информации при помощи вейвлет-преобразования и пространственным устройством чувствительности рецептивных полей нервных окончаний сетчатки глаза. Отдельные клетки «интегрируют» возбуждение от достаточно широкой области сетчатки, при этом сигналы воспринимаются селективно по поверхности сетчатки.

Функции чувствительности глаза (хорошо аппроксимируемые функциями Габора) — типичные представители семейства вейвлетов, по которым сетчатка выполняет «разложение» видимого изображения.

3. Довольно распространено мнение, что вейвлеты позволяют раскрыть внутреннюю природу исследуемого сигнала. Однако это верно лишь отчасти. Аналитическая сторона вейвлет-анализа — вещь довольно тонкая, хотя она очень проста и наглядна. По мнению ряда физиков и математиков, работающих с вейвлет-анализом, простым «прикладыванием» соответствующих формул серьезную задачу не решить — надо быть экспертом в самой проблеме, проработать с ней не один год и уметь отличать внутреннюю проблему от артефактов (т. е. необоснованных показаний методов, не очень хорошо приспособленных к рассматриваемой задаче), порожденных спецификой использованного метода вейвлет-анализа. А так, совершенно не важно, с помощью какой техники решается взятая задача, лишь бы решалась.

Следует помнить, что вейвлет-анализ требует значительного объема вычислений, который быстро растет с увеличением обрабатываемых данных. При обработке одномерного сигнала на выходе получается *трехмерный массив* данных (вейвлет-коэффициенты на плоскости время — частота), а при обработке многомерных массивов данных представление результатов невозможно без компьютеров, имеющих мощные графические возможности.

Выбор вейвлета и глубины разложения, в общем случае, зависит от свойств конкретного обрабатываемого сигнала. Можно дать лишь несколько общих рекомендаций: более гладкие вейвлеты осуществляют более гладкую аппроксимацию сигнала, и наоборот — «короткие» вейвлеты лучше отслеживают пики аппроксимируемой функции (сигнала); глубина разложения влияет на масштаб отсеиваемых деталей. Другими словами, при увеличении глубины разложения модель вычитает шум все большего уровня, пока не наступит «переукрупнение» масштаба деталей и преобразование начнет искажать форму исходного сигнала. Интересно, что при дальнейшем увеличении глубины разложения вейвлет-преобразование начинает формировать сглаженную версию исходного сигнала, т. е. отфильтровывать не только шум, но и некоторые локальные особенности (выбросы) исходного сигнала, что очень важно при сжатии передаваемой по каналам связи информации.

Вейвлет-преобразование представляет собой лишь метод первичной обработки сигнала для повышения эффективности его сжатия. Непосредственно сжатие выполняется после этой предобработки классическими методами. При этом сжатие выполняется, разумеется, для коэффициентов вейвлет-разложения сигнала, а его реконструкция по этим коэффициентам производится на этапе восстановления (декомпрессии).

## 10.2. Общие сведения о фрактальном анализе сигналов

Материалы, изложенные в главе 10, пока не имеют строгого математического описания теории и находятся еще в стадии исследований и поисков. Но их практическое значение велико и бесспорно. На первый взгляд геометрическое изложение теории фракталов имеет мало общего с радиотехническими сигналами. Но это не так. Роль фракталов в компьютерной обработке сигналов и сжатии изображений сегодня достаточно велика. Они незаменимы, например, когда требуется, с помощью нескольких коэффициентов, задать линии и поверхности очень сложной формы. С точки зрения компьютерной графики, фрактальная геометрия незаменима при генерации искусственных облаков, гор, рекламных переливающихся образований, поверхности моря.

Понятие «фрактал» появилось в конце 70-х, а с середины 80-х г. XX в. начало использоваться математиками, физиками, программистами и специалистами по обработке сигналов и сжатию изображений. Слово «фрактал» (от лат. *fractus — дробный* или *состоящий из фрагментов* и *frangere — ломать*) и отражает изломанный, фрагментарный характер объектов и явлений с дробными размерностями, служащими количественной мерой негладкости фракталов. В основе концепции фрактальности лежит некоторая идеализация реального мира — гипотеза самоподобия (скейлинга): *вид фрактальной структуры объектов существенно не меняется при ее масштабных преобразованиях в определенном диапазоне значений масштаба.* Понятие фрактальности было предложено Бенуа Мандельбротом (*Benoit Mandelbrot*; франко-американский математик; р. 1924) в 1975 г. для обозначения нерегулярных, но *самоподобных структур*.

Что же такое *самоподобие*? Система называется самоподобной, если из конечного набора ее копий можно собрать, как в детском конструкторе, такую же систему, которая больше или меньше начальной в определенное число раз. При упаковке шаров наиболее известные плотные конфигурации обладают решетчатой структурой; для упаковки дисков на плоскости — это гексагональная решетка, в трехмерном пространстве задача еще не решена — предполагается, что это будет кубическая решетка. Хотя принцип самоподобия известен уже давно (простейшие примеры — русские матрешки, китайские вложенные шары), самоподобные структуры недавно были, как бы заново, открыты и буквально привели к взрыву теоретической и экспериментальной активности. Отметим, что принцип самоподобия в приближенном виде имеется и в природе: в фигурах снежинок, в кристаллах, в линиях берегов морей и рек, в распределении магнитных силовых линий и полей, в очертаниях облаков и деревьев, в турбулентном потоке жидкости и в иерархической организации живых систем.



Б. Мандельброт

Строгого и полного определения фракталов пока не существует. Определение фрактала, данное Мандельбро-

том, звучит так: «*Фракталом называется структура, состоящая из частей, которые в каком-то смысле подобны целому*». Процессы, порождающие такие структуры, давно изучаются в математике и физике. Это процессы с обратной связью, в которых одна и та же операция выполняется снова и снова, когда результат одной итерации является начальным значением для следующей. Главное требование при этом — нелинейная зависимость между результатом и начальным значением. Формула процесса Мандельброта проста

$$x_{n+1} = f(x_n) = x_n^2 + c. \qquad (10.29)$$

Если начать итерационный процесс указанного вида с произвольного значения $x_0$, то результатом будет некоторая последовательность $x_1$, $x_2$, ..., поведение которой по истечении достаточно большого интервала времени и будет представлять интерес. Будет ли последовательность сходиться к некоторому предельному значению $x$, стремясь к состоянию равновесия? Придет ли она к некоторому циклу значений, которые будут повторяться вновь и вновь? Или эта последовательность будет вести себя все время хаотично, хотя и определена динамическим законом и конкретным начальным моментом, но, тем не менее, непредсказуема? Мандельброт решил вместо действительных чисел $c$ использовать комплексные числа вида $p + jq$ (причем $p$ может быть как положительным, так и отрицательным) и стал наблюдать процесс $x_0 \to x_1 \to x_2 \to \ldots$ не на прямой, а на плоскости. Выбрав произвольное число $x_0$, возведем его в квадрат и прибавим константу $c$ для того, чтобы получить $x_1$; затем повторим вычисления для того, чтобы получить $x_2$, $x_3$ и т. д. Начнем с простейшего из возможных значений константы $c = 0$. Тогда при каждой очередной итерации вычисляется точный квадрат числа: $x_0 \to x_0^2 \to x_0^4 \to x_0^8 \to \ldots$ Для этой последовательности в зависимости от $x_0$ имеются два направления развития:

- числа с каждой операцией уменьшаются с приближением к нулю; можно констатировать, что нуль является *аттрактором* (англ. *attract* — привлекать, притягивать; аттрактор — множество точек в фазовом пространстве динамической системы, к которым стремятся ее траектории) для процесса $x \to x^2$; все точки, находящиеся на расстоянии меньше 1 от этого аттрактора, движутся к нему, числа все время увеличиваются, стремясь к бесконечности; значит, бесконечность также является аттрактором для этого процесса;
- все точки, лежащие на расстоянии больше 1 от нуля, движутся к бесконечности;
- точки находятся и продолжают оставаться на расстоянии 1 от нуля; их последовательности лежат на границе двух областей притяжения, в данном случае на окружности единичного радиуса с центром в нуле.

Очевидно, что плоскость делится на две зоны влияния аттракторов, а граница между ними — окружность. Любая начальная точка $x_0$ в течение процесса либо приходит к тому или другому центру, либо лежит на границе и не может принять определенное решение. С изменением параметра изменяются и области, и границы, принадлежащие аттракторам. Может случиться, что граница превратится в пыль, и такой распад представляет один из наиболее важных сценариев.

Исследования, проведенные Мандельбротом, показали, что сюрпризы начнутся, если выбрать ненулевое значение параметра $c$. Здесь для последовательности значений $x_0 \to x_1 \to x_2 \to \ldots$ имеются три из перечисленных направления развития. В центре внимания оказалась природа границ между различными областями. Подобные процессы обнаруживают в любой точной науке. В радиотехнике в качестве фрактальных структур можно рассматривать реализации случайных процессов, телевизионные изображения, сигналы и изображения на выходе РЛС и радиометров, рассеяние радиоволн подстилающими покровами, фазовые портреты автогенераторов и других динамических систем. Прикладными направлениями исследований в последнее время стали: фрактальная обработка изображений и сжатие информации; обнаружение по фрактальным алгоритмам цели на различных изображениях с предельно малыми отношениями сигнал/шум (называют цифру менее –30 дБ); разработка фрактальных антенн; создание новых моделей рассеяния радиоволн, имеющих конечное число степеней свободы и реализуемые на основе идей детерминированного хаоса.

### Классификация фракталов

**Геометрические фракталы.** Основным свойством фракталов является *нецелое* (*дробное*) *значение их размерности*. Чтобы понять это, обратимся к основам теории размерности. Понятие дробной размерности опирается на анализ понятия целой евклидовой E или топологической размерности $D_0$. Под топологической размерностью $D_0$ множества в линейном пространстве понимают число линейно независимых в нем координат. В частности, точка в евклидовом пространстве E имеет нулевую топологическую размерность $D_0 = 0$. Гладкие кривые — прямые, окружности, различные линии и т. п. — имеют одно измерение, а значит, их размерность $D_0 = 1$. Поверхности имеют размерность $D_0 = 2$ (двумерны), объемные тела — размерность $D_0 = 3$ (трехмерны), гипертела — более высокие значения $D_0$. А какова размерность плоскости домашнего сита для просеивания муки, неплотно заштрихованного листа бумаги или клубка (шара) степного травянистого растения перекати-поле? Твердое тело, жидкость и газ — три привычных для нас состояния однородного вещества, существующего в трехмерном мире. Но какова размерность облака, клуба дыма, точнее их границ, размываемые турбулентным движением воздуха? Оказалось, что она больше 2-х, но меньше 3-х. Аналогичным образом можно посчитать размерности других реальных объектов вроде береговой линии или кроны дерева.

Пытаясь ответить на эти вопросы, можно прийти к системам — фракталам, чья размерность будет измеряться нецелыми (дробными) числами. Фракталы же — это не линии и не поверхности, а, если можно так сказать, что-то среднее. Размерность объекта показывает, по какому закону растет его внутренняя область. Аналогичным образом с ростом размеров возрастает объем фрактала, но его размерность — величина не целая, а дробная, поэтому граница фрактальной фигуры не линия: при большом увеличении становится видно, что она размыта и состоит из спиралей и завитков, повторяющих в малом масштабе саму фигуру. Все объекты с нечеткой, хаотичной, неупорядоченной структурой оказались состоящими из фракталов.

Понятия фрактальных меры и размерности множеств, характеризующих их в целом, базируются на определениях, введенных Феликсом Хаусдорфом (*Felix Hausdorf*; немецкий математик, основатель современной топологии; 1868–1942). *Размерность по Хаусдорфу* определим следующим образом. Пусть на плоскости имеется некая сильно изломанная линия, имеющая «причудливую», фрактальную природу. Будем измерять длину этой линии отрезками с характерным размером $\delta$ (пробными отрезками, их длину $\delta$ называют *коэффициентами подобия*). Тогда ее суммарная длина $L_\delta$ стремится не к конечному пределу, как это должно быть для «обычной» линии, а к бесконечности по степенному закону

$$L_\delta = A\delta^{1-D}, \tag{10.30}$$

где $A$ — некая константа размерности $L^D$ ($L$ — размерность длины), а D — постоянная, называемая *фрактальной размерностью* или *размерностью Хаусдорфа*. Формулу (10.30) можно записать асимптотически при $\delta \to 0$:

$$L_\delta \sim \delta N(\delta), \tag{10.31}$$

где $N$ — число отрезков, необходимых для покрытия фрактальной линии.

Если $l$ — длина каждого такого отрезка (т. е. расстояние между двумя точками по прямой) и $N(\delta)$ определяется законом подобия при $\delta \to 0$, то

$$N(\delta) \sim (l/\delta)^D. \tag{10.32}$$

Из (10.32) вытекает, что фрактальная размерность D находится как

$$D = \lim_{\delta \to 0} \frac{\ln N(\delta)}{\ln \delta}. \tag{10.33}$$

Для связи фрактальной и топологической размерностей используют показатель Херста H ($0 < H < 1$)

$$H = D - D_0. \tag{10.34}$$

Геометрические фракталы наиболее наглядны. Рассмотрим один из первых простейших фракталов (рис. 10.9) — кривую Хельге фон Коха (*Helge von Koch*; шведский математик; 1870 – 1924). Построение кривой начинают с отрезка единичной длины $\delta = 1$ (рис. 10.9, *а*), являющийся *инициатором* (затравкой), и называется *предфракталом* 0-го порядка ($n = 0$). Далее инициатор заменяют на образующий элемент — ломаную линию (кривую) из четырех прямолинейных звеньев, каждое из них имеет длину $\delta = 1/3$. Так образуется предфрактал 1-го порядка ($n = 1$), его длина $\delta = 4/3$. Для построения предфрактала следующего порядка каждое звено заменяют на уменьшенный образующий элемент. В результате получают линию, состоящую из $4 \cdot 4 = 16$ звеньев, каждое из которых имеет длину $(1/3)/3 = 1/9$; общая длина $\delta = 16/9$. Длина предфрактала $n$-го порядка $\delta = (4/3)^n$, а длина звена $\delta = 3^{-n}$ при их числе $N(\delta) = 4^n = \delta^{-D}$.

Очевидно, что предел длины кривой при $n \to \infty$ равен бесконечности. В итоге получили кривую бесконечной длины, заполняющую ограниченное множество на плоскости.

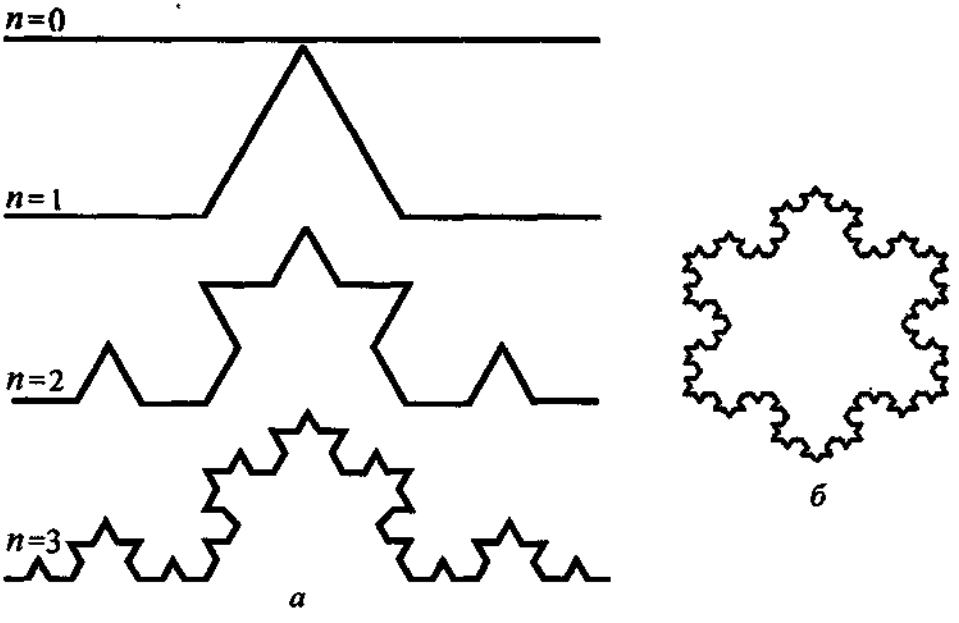



**Рис. 10. 9.** Кривые Коха:
*а* — построение триадной кривой; *б* — снежинка

Если построение кривой начать не с отрезка, а с треугольника, и применить перечисленные действия к каждой его стороне, получим замкнутую фигуру — фрактал *снежинку Коха* (рис. 10.9, *б*). Сопоставим топологическую и *хаусдорфову* размерности триадной кривой Коха. Топологическая размерность кривой $D_0 = 1$, а *хаусдорфова*, согласно соотношению (10.33), $D = \ln 4/(\ln 3) = 1{,}26$.

Построение двух интереснейших фрактальных фигур — *треугольной салфетки* и *ковра* Вацлава Серпинского (*Wacław Sierpinski*; польский математик; 1882–1969) — показано на рис. 10.10, *а, б* соответственно. Это четвертые поколения предфракталов. В первом случае инициатор — треугольник со всеми внутренними точками, а во втором — квадрат.

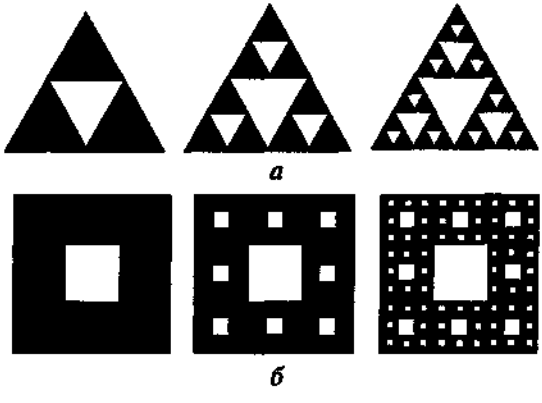
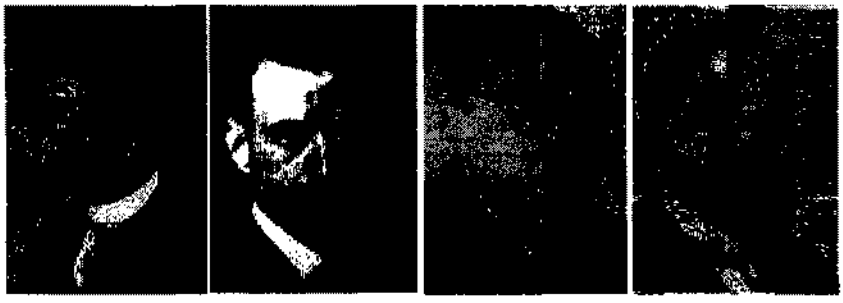



**Рис. 10.10.** Предфракталы Серпинского:
*а* — салфетка; *б* — ковер



Ф. Хаусдорф Х. Кох В. Серпинский Г. Кантор

На каждом следующем этапе треугольник в салфетке заменяют тремя треугольниками, уменьшенными с коэффициентами 1/3. Для ковра образующий элемент состоит из $N = 8$ квадратов с коэффициентом сжатия 1/3. Их бесконечное множество поколений порождает фрактальную кривую. В этом случае черные участки исчезают, а полный периметр отверстий в ковре Серпинского становится бесконечным. Нетрудно заметить, что топологическая размерность и салфетки и ковра Серпинского одинакова $D_0 = 1$, а *хаусдорфова* — соответственно $D = \ln 3/\ln 2 = 1{,}58$ и $D = \ln 8/(\ln 3) = 1{,}89$.

Еще один фрактал — *канторовское множество*, созданное в 1883 г. Георгом Кантором (*Georg Cantor*; немецкий математик;1845–1918). Его построение позволяет сформировать фрактальное множество с топологической размерностью $0 < D_0 < 1$. На рис. 10.11 показано образование триадного канторовского множества (первых пяти поколений отрезков) — *канторовской пыли*. Основой служит единичный отрезок [0, 1], разбиваемый на три одинаковые части. Образующий удаляет среднюю треть. Затем образующий элемент применяется к каждому из двух оставшихся отрезков и т. д. Сумма длин $L$ удаленных отрезков

$$L = 1/3 + 2/9 + 4/27 + \ldots = (1/3) = 1. \qquad (10.35)$$

Топологическая размерность канторовского множества $D_0 = 0$, а *хаусдорфова* — $D = \ln 2/(\ln 3) = 0{,}63$. Отметим, что упомянутый ковер Серпинского является аналогом канторовского множества на квадрате.

**Алгебраические фракталы.** Получают их с помощью нелинейных процессов в $n$-мерных пространствах. Наиболее изучены двухмерные процессы. Интерпретируя нелинейный итерационный процесс как дискретную динамическую систему, можно пользоваться терминологией теории этих систем: *фазовый портрет*, *установившийся процесс*, *аттрактор* и т. д.



**Рис. 10.11.** Канторовское множество

Известно, что нелинейные динамические системы обладают несколькими устойчивыми состояниями. То состояние, в котором оказалась динамическая система после ряда итераций, зависит от ее начального состояния. Поэтому каждое устойчивое состояние обладает некоторой областью начальных состояний, из которых система обязательно попадет в конечные состояния. Таким образом, фазовое пространство системы разбивается на *области притяжения* аттракторов. Если фазовым является двухмерное пространство, то, окрашивая области притяжения различными цветами, можно получить *цветовой фазовый портрет* этой системы. Меняя алгоритм выбора цвета, можно получить сложные фрактальные картины с причудливыми многоцветными узорами.

**Стохастические фракталы.** Они получаются тогда, когда в итерационном процессе случайным образом меняют какие-либо его параметры. При этом возникают объекты, похожие на природные — несимметричные деревья, изрезанные береговые линии и т. д.

### Фрактальный алгоритм сжатия информации

В радиотехнике фракталы позволили разработать высокоэффективные методы сжатия информации. В 1988 г. известный американский специалист в теории динамических систем и эргодической теории Майкл Барнсли (*Michael Barnsley*; р. 1947) предложил некоторые идеи для сжатия и хранения графической информации. Он назвал свой метод *фрактальным сжатием информации.* Происхождение названия связано с тем, что геометрические образы, возникающие в этом методе, обычно имеют фрактальную природу.

Фрактальный алгоритм сжатия информации основан на представлении изображения в более компактной форме — с помощью коэффициентов *системы итерирующих функций* (*Iterated Function System — IFS*). Система итерирующих функций — совокупность сжимающих *аффинных преобразований* (от лат. *affinis* — родственный). Аффинные преобразования — геометрические преобразования плоскости или пространства, включающие в себя масштабирование, поворот, и параллельный перенос, при котором фигуры сохраняют свои свойства. В изображениях преобразованию подвергают точки в трехмерном пространстве (координата *X*, координата *Y*, яркость). По существу *IFS* представляет собой систему из фиксированного класса функций, отображающих одно многомерное множество на другое. Аффинное преобразование считается сжимающим, если коэффициент масштабирования меньше единицы.



М. Барнсли

Прежде чем рассматривать процесс архивации, разберем, как *IFS* строит изображение. Его продемонстрировал Барнсли в книге «Фрактальное сжатие изображения», где введено понятие «фотокопировальной машины», состоящей из экрана с изображенной исходной картинкой и системы линз, проецирующих картину на другой экран (рис. 10.12). Каждая линза проецирует часть исходного изображения. Расставляя линзы и меняя их характеристики, можно управлять получаемым изображением.

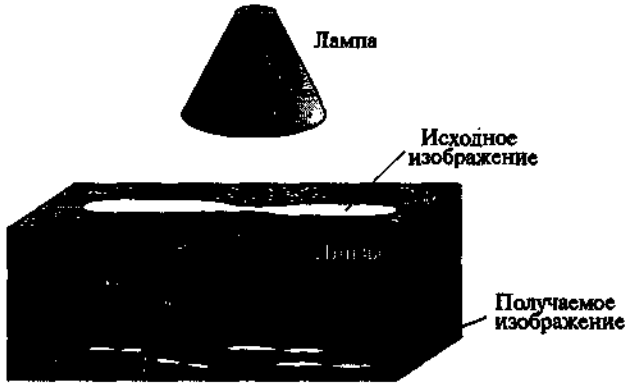



**Рис. 10.12.** Фотокопировальная машина Барнсли

На линзы накладывается требование: линзы должны уменьшать (масштабировать) свой фрагмент изображения. Кроме того, они могут менять яркость и уменьшать контрастность фрагмента, проецируя не круги, а области с произвольной границей. Каждая линза проецирует часть исходного изображения. Расставляя линзы и меняя их характеристики, можно управлять получаемым изображением. Одна итерация работы машины заключается в том, что по исходному изображению с помощью проектирования строится новое, после чего новое берется в качестве исходного. В процессе итераций получается изображение, которое перестанет изменяться. Оно будет зависеть только от расположения и характеристик линз и не будет зависеть от исходной картинки. Это изображение называется *неподвижной точкой*, или *аттрактором* данной *IFS*. Соответствующая теория гарантирует наличие ровно одного аттрактора для каждой *IFS*. Поскольку отображение линз является сжимающим, каждая линза в явном виде задает *самоподобные области* в изображении. Благодаря самоподобию получают сложную структуру изображения при любом увеличении. Итак, интуитивно понятно, что система итерируемых функций задает *фрактал*.

По существу алгоритм Барнсли выделяет в изображении пару областей, меньшая из которых подобна большей, и сохраняет нескольких коэффициентов, кодирующих преобразование. Требуется, чтобы множество «меньших» областей покрывало все изображение. При этом в файл, кодирующий изображение, записывают не только коэффициенты, характеризующие найденные преобразования, но и местоположение и линейные размеры «больших» областей, которые вместе с коэффициентами будут описывать локальное самоподобие кодируемого изображения. Восстанавливающий алгоритм, в этом случае, должен применять каждое преобразование не ко всему множеству точек, получившихся на предыдущем шаге алгоритма, а к некоторому их подмножеству, принадлежащему области, соответствующей применяемому преобразованию.

Одной из основных проблем, с которой пришлось столкнуться при построении алгоритма фрактальной компрессии, является поиск самоподобных участков в изображении. Это основная идея, благодаря которой осуществляется сжатие.

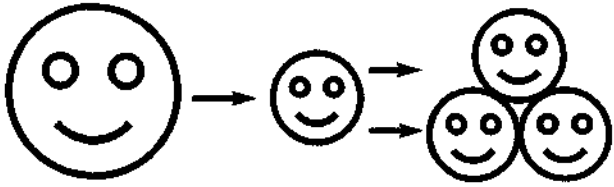


**Рис. 10.13.** Способ фрактального преобразования изображения

Подобный метод можно применить и при архивации видео. Как правило, соседние кадры отличаются несильно, и изменения между ними в основном состоят в сдвиге, повороте или растяжении какой-либо части изображения. Таким образом, изменения между двумя кадрами можно задать небольшим числом аффинных преобразований. Для фрактального алгоритма компрессии, как и для других алгоритмов сжатия с потерями, очень важны механизмы, с помощью которых можно будет регулировать степень сжатия и степень потерь.

К настоящему времени разработан достаточно большой набор таких методов. Во-первых, можно ограничить число преобразований, заведомо обеспечив степень сжатия не ниже фиксированной величины. Во-вторых, можно потребовать, чтобы в ситуации, когда разница между обрабатываемым фрагментом и наилучшим его приближением будет выше определенного порогового значения, этот фрагмент дробился обязательно. В-третьих, можно запретить дробить фрагменты размером меньше, допустим, четырех точек. Изменяя пороговые значения и приоритет этих условий, можно очень гибко управлять коэффициентом компрессии изображения: от побитного соответствия до любой степени сжатия.

Рассмотрим фрактальный алгоритм, который преобразует изображение некоторым заданным способом. Например, этот способ предполагает уменьшение линейных размеров исходного изображения (например, рисунок «улыбающееся лицо») в два раза и тройное его копирование (рис. 10.13).

Пусть имеется три изображения: улыбающееся лицо, буква А и салфетка Серпинского, показанных слева на рис. 10.14. Если процесс преобразования повторять к изображениям итеративно несколько раз, то возникающие из различных исходных изображений рисунки станут похожими. При достаточно большом числе итераций рисунки будут неразличимы. Конечное изображение обладает рядом интересных свойств, присущим лишь аттракторам фракталов:

- не зависит от начального изображения, поскольку при достаточно большом числе итераций оно уменьшится до аттрактора (точки);
- определяется исключительно процедурой преобразования;
- последующие преобразования преобразуют его в самое себя;
- может иметь сколь угодно мелкие детали.

Процедуры преобразования могут быть и другие. Единственное ограничение — требование сходимости изображения в указанном смысле. В противном случае, если две разные точки исходного изображения при последовательном преобразовании не сойдутся в одну, конечное изображение будет зависеть от исходного и не будет аттрактором.

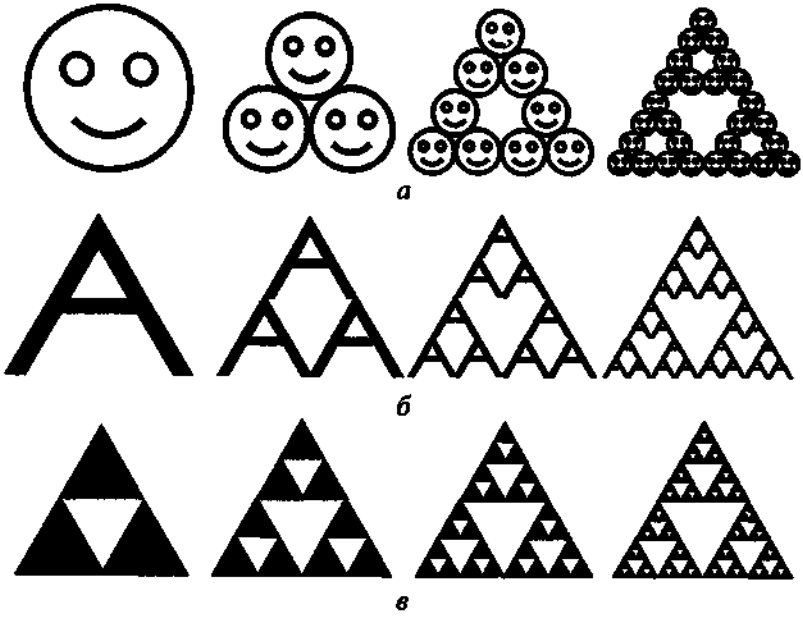


**Рис. 10.14.** Фрактальное преобразование различных изображений:
*а* — смеющегося лица; *б* — буквы А; *в* — салфетки Серпинского

На практике достаточно большое количество преобразований можно описать с помощью матричного уравнения, определяющего линейное преобразование координат $x$ и $y$:

$$\mathbf{W}_i\begin{bmatrix}x\\y\end{bmatrix}=\begin{bmatrix}x\\y\end{bmatrix}\cdot\begin{bmatrix}a_i & b_i\\c_i & d_i\end{bmatrix}+\begin{bmatrix}g_i\\h_i\end{bmatrix}, \qquad (10.36)$$

где $a_i$, $b_i$, $c_i$, $d_i$, $g_i$ и $h_i$ — параметры аффинного преобразования.

На рис. 10.15 показаны ряд процедур преобразования изображений и их аттракторов. Последний случай (рис. 10.15, *в* — правый крайний рисунок) относится к *папоротнику Барнсли*. Это внешне сложное изображение получено в результате четырех аффинных преобразований, каждое из которых имеет шесть параметров (см. (10.36)). Тот же папоротник Барнсли создается с помощью *IFS* четырьмя аффинными преобразованиями (в нашей терминологии «линзами»). Каждое преобразование кодируется буквально считанными байтами, в то время как изображение, построенное с их помощью, может занимать и несколько мегабайт. Если для штрихового изображения папоротника Барнсли перемножить число преобразований 4, число параметров 6 и число бит под хранение каждого из параметров например, 32, то получим 4·6·32 = 768 бит — столько бит необходимо для хранения способа получения этого изображения. В то же время изображение папоротника (1 бит/пиксел) имеет разрешение 256 · 256 пиксел. Для прямого хранения такого изображения необходимо 65 536 бит, т. е. рассматриваемая схема позволяет «сжать» изображение примерно в 85 раз. Такое определение коэффициента сжатия в некоторой мере условно.

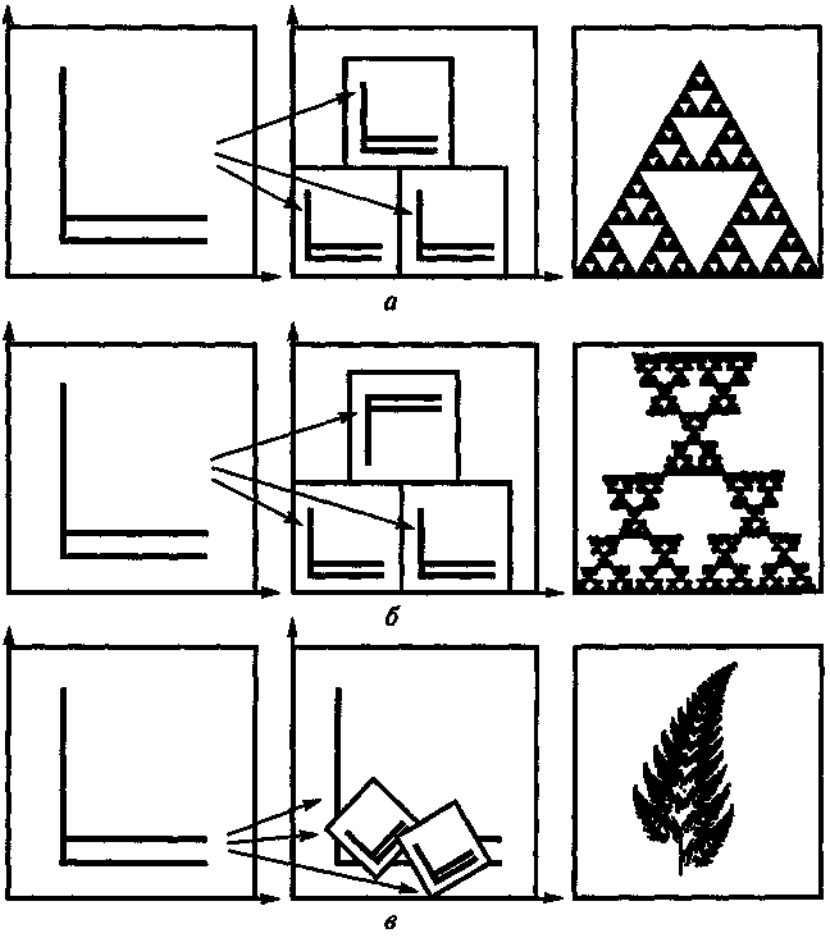



**Рис. 10.15.** Различные фрактальные преобразования изображений:
*а* — известное; *б* — преобразование при дополнительном повороте верхней детали;
*в* — сложное преобразование с поворотом деталей на разные углы и с их масштабированием

Дело в том, что для хранения алгоритма преобразования требуется определенное, заранее известное, число бит. Но этот алгоритм позволяет создать изображение любого размера с достаточно мелкими деталями, для хранения которого требуется другое (возможно, большее) число бит. Соответственно размеру изображения будет меняться и коэффициент сжатия. Возникает вопрос: можно ли для любой исходной детали изображения подобрать деталь, которая после некоторых преобразований станет достаточно похожей на исходную? Строгое математическое доказательство отсутствует, однако практика показывает, что это возможно практически во всех случаях. Эти преобразования для полутоновых монохромных изображений можно формально описать следующим образом. Пусть яркость пикселов изображения $z$ задана функцией

$$z = f(x; y), \tag{10.37}$$

где $x$, $y$ — координаты пикселов.

Пусть $z$ имеет 256 фиксированных уровней. Само аффинное преобразование $i$-го блока полутонового изображения имеет вид:

$$\mathbf{W}_i\begin{bmatrix}x\\y\\z\end{bmatrix}=\begin{bmatrix}x\\y\\z\end{bmatrix}\cdot\begin{bmatrix}a_i & b_i & 0\\c_i & d_i & 0\\0 & 0 & s_i\end{bmatrix}+\begin{bmatrix}g_i\\h_i\\r_i\end{bmatrix}, \qquad (10.38)$$

где $a_i$, $b_i$, $c_i$, $d_i$, $g_i$ и $h_i$ — параметры аффинного преобразования; $s_i$, $r_i$ — коэффициенты преобразования контраста и яркости блока.

Теперь исходное изображение необходимо разбить на блоки (*домены*), для которых будут подбираться подобные блоки (*ранговые области*). Вычисление оптимальных коэффициентов преобразования контраста и яркости квадратов путем минимизации среднеквадратичной разности яркостей пикселов домена и кандидата в ранговую область может быть произведено по разным формулам, приведенным в специальной литературе. Итак, понятие фрактала привело к развитию новой математической модели, дающей единое описание форм, присущих многим природным явлениям, сигналам, изображениям и т. д. Оказывается, природа устроена так, что именно фрактальные модели хорошо описывают многие реальные объекты, структура которых не отражается традиционными моделями. Этим объясняется современная популярность фрактального подхода к анализу различных объектов и процессов в физике, радиотехнике, системах передачи информации и др. В частности, сейчас большое внимание уделяется разработке широкополосных фрактальных антенн (рис. 10.16).

Очевидно, что новые геометрические и топологические представления фрактального анализа в скором времени станут такой же непременной частью анализа сигналов и процессов в радиоэлектронике и теории передачи информации, какой стал Фурье-анализ. Отметим, что между вейвлетным и фрактальным анализами много общего, поскольку в них используется принцип самоподобия.

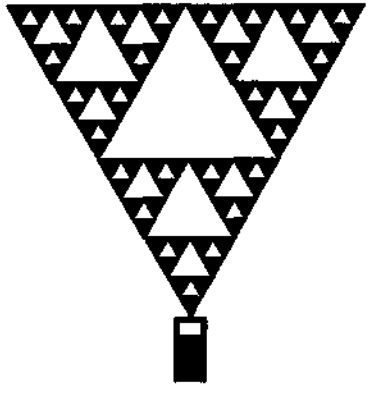


**Рис. 10.16.** Фрактальная антенна

## Контрольные вопросы

1. Какие сигналы можно использовать в качестве вейвлетов?
2. В чем отличие непрерывного и дискретного вейвлет-анализов?
3. Чем характерен вейвлетный алгоритм сжатия информации?
4. Дайте определение и приведите классификацию фракталов.
5. Какими параметрами можно описать фрактальные функции?
6. На каком принципе основано действие фотокопировальной машины Барнсли?
7. Чем характерен фрактальный алгоритм сжатия информации?

# ЗАКЛЮЧЕНИЕ

Перевернув последнюю страницу, подведем некоторые итоги, чтобы оценить полученные знания и наметить пути дальнейшей работы.

Современные курсы радиотехники и теории связи используют понятия и методы из различных областей, прежде всего математики, физики, информатики, вычислительной техники, теории электрических цепей, технической электродинамики, оптоэлектроники, микроэлектроники и т. д. Все эти понятия и методы образуют в курсе «Основы радиоэлектроники и связи» определенное единство и должны рассматриваться как нечто целое в рамках системного подхода, отвечающего требованиям современной науки и техники. Фундаментальное понятие этого курса — математическая модель радиотехнического сигнала.

В данном учебном издании рассмотрен целый ряд современных математических моделей сигналов — непрерывных, дискретных и цифровых, а также методов и устройств, генерирующих, преобразующих и обрабатывающих эти сигналы. Некоторые из моделей таких сигналов, методов и устройств еще не только не нашли должного отражения в учебной литературе по радиоэлектронике и технике связи, но пока не имеют четкого математического и физического толкования. В частности, практически впервые в книге такого широкого плана рассмотрены методы вейвлетного и фрактального анализа сигналов и обработки различных видов информации. Вместе с тем, все представленные модели позволяют с достаточной степенью полноты и точности выполнять две взаимосвязанные операции — анализ и синтез.

До сих пор радиотехника и теория связи развивались так, что методы анализа всегда опережали методы синтеза. Однако в последние годы тенденция существенно меняется, главным образом под воздействием внедрения компьютерных технологий в практику научных исследований. Переход к цифровым методам передачи информации при использовании микропроцессоров приводит к интеграции радиотехнических устройств и средств вычислительной техники.

Тенденции развития современной радиоэлектроники и техники связи:

1. Исследование и практическое освоение новых все более коротковолновых диапазонов электромагнитных волн. В последние годы интенсивно осваиваются миллиметровый, субмиллиметровый, инфракрасный и оптический диапазоны. Укорочение длины волны связано, прежде всего, с необходимостью резкого увеличения скорости передачи информации (емкости канала связи), пропорциональной полосе частот $\Delta f$ радиосигнала. Это также позволяет применять более помехоустойчивые виды модуляции, например, ИКМ, при которой из смеси радиосигнала $u(t)$ и помехи $r(t)$ легче выделить полезный сигнал.

2. Повышение уровней мощности передающих и чувствительности радиоприемных устройств. Следует, однако, иметь в виду, что [illegible] мощности радиоустройств могут нанести вред природе и генофонду [illegible].

3. Переход к интегральной и оптической технологии производства радиоэлектронных устройств, разработка интегральных микросхем [illegible] более высокочастотных диапазонов.

4. Всеобъемлющий переход к системам цифровой обработки сигналов. Сжатие передаваемой информации с полным ее восстановлением при приеме, что позволит существенно повысить пропускную способность каналов связи. Ожидается, что к первой четверти XXI в. большинство радиоэлектронных систем будет функционировать на цифровой основе, причем цифровая обработка будет вестись на все более высоких частотах вплоть до СВЧ-диапазона.

5. Создание мировой глобальной системы радиосвязи и объединение под «руководством» центральных компьютеров различных систем связи в единую автоматизированную систему. Многофункциональный характер глобальной системы радиосвязи, позволяющей передавать не только речь, но и разнообразные данные, в том числе и информацию, считываемую с компьютера. Объединение системы радиосвязи с радионавигационной системой, позволяющей с высокой точностью определять местоположение корреспондентов.

6. Установление фундаментальных пределов радиоэлектроники с точки зрения физики, экологии, экономики, социологии: возможностей микроминиатюризации, размеров антенных систем, генерируемых мощностей, мощностей минимально обнаружимых сигналов, полос пропускания устройств и т. п.

7. Проникновение идей и методов радиоэлектроники в самые разнообразные отрасли человеческой жизни и деятельности. В наше время трудно представить себе какую-либо область науки и техники, в которой не использовалась бы радиоэлектроника. С ее применением связаны наиболее заметные открытия последних десятилетий во многих отраслях.

Написать книгу было непросто. Над авторами все время довлели математические аспекты радиотехники и теории связи. Дать меньше математики можно, но существенно сократить ее нельзя. Ведь и теория сигналов, и теория связи — прежде всего математические теории. В очень сложных местах авторы просто указывали основные свойства используемого математического выражения и старались меньше вдаваться в его сущность. И все же учебное издание содержит разделы, трудные для понимания студента. Советуем изучать эти разделы, пытаясь уловить из математических трактовок все, что удастся. После того как вся книга будет прочитана, можно вернуться к этим разделам снова, чтобы уже по-настоящему попытаться их понять. И последнее, авторы надеются, что результаты этой работы не окажутся бесполезными, и книга послужит основательным фундаментом для изучения дополнительной литературы по интересующим читателя вопросам и самостоятельного освоения специальных и новых дисциплин, формирующих грамотного и вдумчивого исследователя в области радиоэлектроники и теории и техники связи.

# СПИСОК ЛИТЕРАТУРЫ


1. *Баскаков С.И.* Радиотехнические цепи и сигналы. – М.: Высшая школа, 2005.
2. *Гоноровский И.С., Демин М.П.* Радиотехнические цепи и сигналы. – М.: Радио и связь, 1994.
3. *Сергиенко А.Б.* Цифровая обработка сигналов. – СПб.: Питер, 2005.
4. *Нефедов В.И.* Основы радиоэлектроники и связи. – М.: Высшая школа, 2005.
5. Теория электрической связи /*А.Г. Зюко, Д.Д Кловский, В.И. Коржик и др.*; Под ред. *Д.Д. Кловского.* – М.: Радио и связь, 1999.
6. Теоретические основы радиотехники /*М.Т. Иванов, А.Б. Сергиенко, Ушаков В.Н.*; Под ред. *В.Н. Ушакова.* – М.: Высшая школа, 2002.
7. *Прокинс Дж.* Цифровая связь. – М.: Радио и связь, 2000.
8. *Гоноровский И.С.* Радиотехнические цепи и сигналы. – М.: Радио и связь, 1986.
9. *Каганов В.И.* Радиотехнические цепи и сигналы. – М.: ACADEMA, 2003.
10. *Хотунцев Ю.Л., Лобарев А.С.* Основы радиоэлектроники. – М.: Агар, 1998.
11. *Манаев Е.И.* Основы радиоэлектроники. – М.: Радио и связь, 1990.
12. *Шеннон К.* Работы по теории информации и кибернетике /пер. с англ.; Под ред. *Р.Л. Добрушина, О.Б. Лупанова.* – М.: ИЛ, 1963.
13. *Громаков Ю.А.* Стандарты и системы подвижной радиосвязи. – М.: Эко-Трендз, 1997.
14. *Першин В.Т.* Основы современной радиоэлектроники. – Ростов н/Д.: Феникс, 2009.
15. *Феер К.* Беспроводная цифровая связь. – М.: Радио и связь, 2000.
16. *Ван Трис Г.* Теория обнаружения, оценок и модуляции. – М.: Радио и связь, 1972.
17. *Новожилов О.П.* Основы микропроцессорной техники: В 2 т. – М.: РадиоСофт, 2007.
18. *Сигов А.С., Нефедов В.И.* Метрология, стандартизация и технические измерения. – М.: Высшая школа, 2008.
19. *Быстров Ю.А., Мироненко И.Г.* Электронные цепи и микросхемотехника. – М.: Высшая школа, 2002.
20. *Сиберт У.М.* Цепи, сигналы, системы: В 2 ч. – М.: Мир, 1988.
21. Передача дискретных сообщений /*В.П. Шувалов, В.О. Захарченко, В.О. Шварцман и др.*; Под ред. *В.П. Шувалова.* – М.: Радио и связь, 1990.
22. *Варакин Л.Е.* Системы связи с шумоподобными сигналами. – М.: Радио и связь, 1985.
23. *Френкс Л.* Теория сигналов /пер. с англ.; Под ред. Д.Е. Вакмана. – М.: Советское радио, 1974.
24. Радиотехнические цепи и сигналы. *Д.В. Васильев, М.Р. Витоль, Ю.Н. Горшенков и др.*; Под ред. *К.А. Самойло.* – М.: Радио и связь, 1982.
25. *Андрианов В. И., Соколов А.В.* Средства мобильной связи. – СПб.: Питер, 1998.
26. Радиотехнические и радиооптические системы /*Э.А. Засовин, А.Б. Борзов, Р.П. Быстров*; Под ред. *Э.А. Засовина.* – М.: Круглый год, 2001.
27. *Румянцев К.Е., Землянухин П.А., Окорочков А.И.* Радиотехнические цепи и сигналы. – М.: Академия, 2005.
28. *Зубарев Ю.Б., Дворкович В.П.* Цифровая обработка телевизионных и компьютерных изображений. – М.: МЦНТИ, 1997.
29. *Брейсуэлл Р.* Преобразование Хартли /пер. с англ. *А.И. Папкова.* – М.: Мир, 1990.
30. *Скляр Б.* Цифровая связь. Теоретические основы и практическое применение. – М.: Издательский дом «Вильямс», 2004.
31. *Блейхут Р.* Быстрые алгоритмы цифровой обработки сигналов. – М.: Мир, 1989.
32. *Гольденберг Л.М.* Цифровая обработка сигналов. – М.: Радио и связь, 1990.

33. *Шелухин О.И., Румянцев К.Е.* Радиоэлектронные средства бытового назначения. – М.: Академия, 2008.
34. *Гутников В.С.* Фильтрация измерительных сигналов. – Л.: Энергоатомиздат, 1990.
35. *Нефедов В.И., Сигов А.С., Битюков В.К., Самохина Е.В.* Электрорадиоизмерения измерения. – М.: Форум, 2009.
36. *Даджион Д., Мерсеро Р.* Цифровая обработка многомерных сигналов. – М.: Мир, 1988.
37. *Дмитриев В.И.* Прикладная теория информации. – М.: Высшая школа, 1989.
38. *Марпл-мл. С.Л.* Цифровой спектральный анализ и его приложения. – М.: Мир, 1990.
39. *Рабинер Л., Гоулд Б.* Теория и применение цифровой обработки сигналов. – М.: Мир, 1978.
40. *Хемминг Р.В.* Цифровые фильтры. – М.: Недра, 1987.
41. *Дьяконов ВП., Абраменкова И.А.* MATLAB. Обработка сигналов и изображений: Специальный справочник. – СПб.: Питер, 2002.
42. *Петухов А.П.* Введение в теорию базисов всплесков. – СПб.: СПбГТУ, 1999.
43. Адаптивные фильтры /Под ред. *К.Ф. Коуэна и П.М. Гранта.* – М.: Мир, 1988
44. *Дьяконов В.П.* Вейвлеты. От теории к практике. – М.: СОЛОН-Р, 2002.
45. *Шинаков Ю.С.* Теория передачи сигналов в электросвязи. – М.: Радио и связь. 1989.
46. Телекоммуникационные системы и сети. Т. 2 /*Г.П. Катунин, Г.В. Мамчев, В.Н., Попантонопуло, В.П. Шувалов* – Новосибирск: ЦЭРИС, 2000.
47. Многоканальные системы передачи /*Н.Н. Баева, В.Н. Гордиенко, С.А. Курицин и др.* – М.: РиС, 1996.
48. *Миддлтон Д.* Введение в статистическую теорию связи: В 2 т. – М.: Советское радио, (1961, 1962).
49. *Вудворд Ф.М.* Теория вероятностей и теория информации с применениями в радиолокации. – М.: Советское радио, 1965.
50. Цифровые и аналоговые системы передачи /*В.И.Иванов, В.Н.Гордиенко, Г.Н. Попов и др.* – М.: РиС, 1995.
51. Спутниковые системы персональной и подвижной связи для обслуживания абонентов на территории России. – М.: ИПРЖР, 2001.
52. Системы мобильной связи /*В.П. Ипатов, В.К. Орлов, И.М. Самойлов, и др.*; Под ред. *В.П. Ипатова.* – М.: Горячая линия – Телеком, 2003.
53. *Лайонс Р.* Цифровая обработка сигналов: 2-е издание. Пер. с англ. – М.: ООО «Бином-Пресс», 2006 г.
54. *Фомин Н.Н.* Сжатие изображений. – М.: Горячая линия – Телеком, 2007.
55. *Cooley J.W., Tukey J.W.* An algorithm for the machine calculation of complex Fourier series //Math. Comput. – 1965. – Vol. 19. – № 90.
56. Антенно-фидерные устройства систем сухопутной подвижной радиосвязи /*А.Л. Бузов, Л.С. Казанский, В.А. Романов и др.*; Под ред. *А.Л. Бузова.* – М.: Радио и связь, 1997.
57. *Hartley R.V.L.* Transmission of Information // Bell System Technical Journal. 1928. – Vol. 7. № 3. – P. 535–563.
58. *Добеши И.* Десять лекций по вейвлетам. – Ижевск: РХД, 2001.
59. *Козлов П.В., Чен Б.Б.* Вейвлет-преобразование и анализ временны́х рядов //Вестник КРСУ. – № 2. –2002.
60. *Галкин В.А.* Цифровая мобильная радиосвязь. – М.: Горячая линия–Телеком, 2007.
61. *Улахович Д.А.* Основы теории линейных электрических цепей. – СПб.: БХВ-Петербург, 2009.
62. *Кловский Д.Д., Шилкин В.А.* Теория электрической связи. Сб. задач и упражнений. – М.: Радио и связь, 1990.
63. *Тихонов В.И.* Статистическая радиотехника. – М.: Радио и связь, 1982.

64. *Райс С.О.* Математический анализ случайного шума // Теория передачи электрических сигналов при наличии помех. – М.: Иностранная литература, 1953.
65. *Давенпорт В Б., Рут В.Л.* Введение в теорию случайных сигналов и шумов /Пер. с англ.; Под ред. *Р.М. Добрушина.* – М.: ИЛ, 1960.
66. *Nyquist H.* Certain topics in telegraph transmission theory //Trans. AIEE. – Vol. 47. Pp. 617 – 644. 1928.
67. *Петров Б.М.* Электродинамика и распространение радиоволн. – М.: Радио и связь, 2000.
68. Электродинами́ка и распространение радиоволн. */В.А. Неганов, О.В. Осипов, С.Б. Раевский и др.* / Под ред. *В.А. Неганова и С.Б. Раевского.* – М.: Радио и связь, 2005.
69. *Котельников В.А.* Теория потенциальной помехоустойчивости. – М.: Советское радио, 1956.
70. Справочник по радиолокации: В 4 т. / Под ред. *К.Н. Трофимова.* – М.: Радио и связь, 1977.
71. *Хармут Х.* Теория секвентного анализа: основы и применения. – М.: Мир, 1980.
72. Радиотехника: Энциклопедия /Под ред. *Ю.Л. Мазора., Е.А. Мачусского, Правды В.И.* – М.: Издательский дом «Додэка-XXI». 2002.
73. *Лосев А.К.* Теория линейных электрических цепей. – М.: Высшая школа, 1987.
74. Антенно-фидерные устройства и распространение радиоволн / *Г.А. Ерохин, О.В. Чернышев и др.*; Под ред. *Г.А Ерохина.* – М.: Горячая линия – Телеком, 2004.
75. *Гусев В.Г., Гусев Ю.М.* Электроника и микропроцессорная техника. – М.: Высшая школа, 2008.
76. *Кук Ч., Бернфельд М.* Радиолокационные сигналы /Пер. с англ.; Под ред. *В.С. Кельзона.* – М.: Советское радио, 1971.
77. *Баскаков С.И.* Радиотехнические цепи и сигналы. Руководство к решению задач. – М.: Высшая школа, 2002.
78. *Анго А.* Математика для электро- и радиоинженеров. / Пер. с франц.; Под ред. *К.С. Шифрина.* – М.: Наука, 1965.
79. *Фомин Н.Н.* Радиоприемные устройства. – М.: Радио и связь, 2007.
80. Метрология и радиоизмерения */В.И. Нефедов, В.И. Хахин, В.К. Битюков и др.*; Под ред. *В.И. Нефедова.* – М.: Высшая школа, 2006.
81. Устройства СВЧ и антенны. Проектирование фазированных антенных решеток /Под ред. *Д.И.Воскресенского.* – М.: Радиотехника, 2003.
82. *Догадин Н.Б.* Основы радиотехники. – М.: Лань, 2008.
83. *Каганов В.И.* Радиоэлектронные системы автоматического управления. Компьютеризированный курс. – М.: Горячая линия –Телеком, 2009.
84. *Левин Б.Р.* Теоретические основы статистической радиотехники. – М.: Советское радио, 1989.
85. *Mandelbrot B.B.* The fractal geometry of nature. – San Francisco.: Freeman,1982.
86. *Ансон Л., Барнсли М.* Фрактальное сжатие изображений // Мир ПК. –.№4. – 1992.
87. *Воробьев В.И., Грибунин В.Г.* Теория и практика вейвлет-преобразования. – СПб.: Изво ВУС, 1999.
88. *Крухмалев В.В., Гордиенко В.Н., Моченов А.Д.* Основы построения телекоммуникационных систем и сетей. – М.: Горячая линия–Телеком, 2009.

# ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ


Автогенератор 476
- баланс амплитуд 479
- баланс фаз 479
- жесткий режим 485
- кварцевый 484
- мягкий режим 484
- с мостом Вина 489
- стационарная амплитуда 477
- типа *LC* 480
- типа *RC* 487
- трехточечный 486
- условия самовозбуждения 477

Антенна 46
- адаптивная (интеллектуальная) 114
- параболическая 48
- плоскостная (низкопрофильная) 147
- спиральная 145
- типа «волновой канал» 47
- четвертьволновая 48
- фрактальная 721
- шлейфовая 145

Аппроксимация характеристик 440
- кусочно-линейная 440
- степенна́я 440
- показательная 440

Аффинное преобразование 617

Базис 140
- ортогональный 140
- ортонормированный 140

Биения 223, 439
Бит 27
Байт 28
Билинейное преобразование 586
Буняковского-Шварца неравенство 632

Варикап 453
Варактор 470
Вейвлет 690
- анализ (преобразование) 690
- - дискретный (ортогональный
- - непрерывный 692
- Добеши 692
- Морле 692
- сомбреро 692
- Хаара 701

Верность приема 78, 614
Вибратор 47
- заземленный 48
- полуволновый 48
- четвертьволновый 48

Видео 89
- детектор 89
- сигнал 86
- импульс 87
- усилитель 86

Виртуальный нуль 415

Гармоника 184
- высшая 184
- основная (первая) 184

Генератор
- задающий 87
- ждущий 476
- импульсов 476
- накачки 470
- на туннельном диоде 497
- несущей частоты 66
- опорный 607
- развертки 87
- - кадровой 87
- - строчной 87

- сверхвысоких частот 496
- синхроимпульсов 88
- - кадровых 88
- *M*-последовательности 495
- - строчных 88
- цифровой 493
- типа *LC* 480
- типа *RC* 487
Гауса (нормальный) закон 305
Гиббса явление 192
Гильберта преобразование 284
Годограф 412

Двухполюсник 355
Девиация частоты 235
Девиация фазы 235
Декада 169
Декодер 31
- звука 95
- изображения 95
Дельта-функция (Дирака) 203
- спектральная плотность 204
- фильтрующее свойство 204
Демодулятор 69
Детектор 69
- амплитудный 69
- диодный 457
- входное сопротивление 4571
- квадратичный 457
- линейный 457
- на ОУ 460
- на аналоговом перемножителе 463
- фазовый 464
- на *JK*триггере 526
- - на логических элементах 525
- - цифровой 525
- частотный (дискриминатор) 463
- - цифровой 524
Диаграмма
- векторная 221
- - АМ-сигнала 221
- - ЧМ-сигнала 236
- направленности антенны 47
- - вибратора 47
- - ФАР 49
- спектральная 182

Диапазон 32
- волн 32
- частот 32
- динамический 29
- - сигнала 29
- - усиления 29
Динамическая система 359
Дискретизация сигнала 531
Дискретная гармоника 578
Дискретная свертка 558
Дисперсия 307

Добротность 378
Добротность контура 378
Дрейф нуля 412
Дюамеля интеграл 366

Емкость канала 53
- пропускная способность 53
- конденсатора 355

Зона молчания 43
*z*-преобразование 559
Избирательность 612
Импульс 87
- кадровый 87
- строчный 87
- синхронизации 87
- прямоугольный 87
Индекс модуляции
- амплитудной 220
- частотной 234
- фазовой 236
Интеграл свертки 366
Интервал
- дискретизации 533
- корреляции 214
Интерфейс 58
Информация 21
Искажения сигнала
- линейные (частотные) 383
- нелинейные 384

Канал
- зеркальный 613
- интермодуляционный 621
- комбинационный 613
- обработки 58
- связи 39
- соседний 621
- спутниковый 39

Кинескоп 85
Кластер 113
Кодер 31
- звука 97
- изображения 97
- канала 32
Кодек (кодер-декодер) 31
Кодирование 657
- компонентное 657
- примитивное 657
- помехоустойчивое 657
- экономное (сжатие данных) 657
Код
- Баркера 344
- Бодо 27
- двоичный 32
- Морзе 28
- неравномерный 657
- помехоустойчивый 657
- - с исправлением ошибок 657
- - с обнаружением ошибок 657
- примитивный 657
- равномерный 657
Колебание
- несущее 61
- поднесущее 64
- с одной боковой полосой 224
Контур 377
- добротность 378
- параллельный 379
- последовательный 377
Котельникова
- ряд 533
- теорема 532
- функция 534
Коэффициент
- амплитудной модуляции 220
- гармоник 393
- направленного действия антенны 8
- передачи цепи (петли) ОС 405
- усиления антенны 47
- усиления усилителя 386
Критерий устойчивости
- Найквиста 411
- Михайлова 412
- Рауса-Гурвица 411

Лапласа
- изображение 363
- преобразования 363
- функция 362
Линеаризация 438
Линия связи 52
- волоконно-оптическая 142

Манипуляция 250
Математическое ожидание 307
Метод графоаналитический 389
Многополюсник 356
Многостанционный доступ с разделением 108
- временным (МДВР) 108
- временным дуплексным 108
- кодовым (МДКР) 108
- частотным (МДЧР) 108
Модем (модулятор-демодулятор) 73
- проводной 73
- цифровой 73
Модулятор 449
- амплитудный 449
- Армстронга 455
- на аналоговых перемножителях 451
- частотный 453
- - цифровой 453
- фазовый 455
Модуляция 218
- амплитудная 218
- - балансная 223
- - однополосная 224
- - цифровая 248
- амплитудно-импульсная 249
- импульсная 249
- импульсно-кодовая 249
- однотональная 220
- полярная 230
- угловая 234
- фазовая 238
- широтно-импульсная 2491
Момент 2-го порядка 307
Напряженность электрического поля 34

Обработка сигнала
- дискретная 531
- цифровая 531
Обратная связь 402
- глубокая 247
- комбинированная 403
- отрицательная 402
- паразитная 402

- параллельная 403
- по напряжению 403
- положительная 404
- последовательная 403
- по току 403
Огибающая 197
- АМ-сигнала 225
- физическая 278
- комплексная 278
Отношение сигнал/шум 78

Парсеваля равенство 201
Пейджер 128
Передатчик (радиопередатчик) 598
Перемножитель аналоговый 451
Перемодуляция 221
Пиксел 154
Плотность вероятности
- одномерная 229
- двумерная 229
- многомерная 229
Полоса
- заграждения 423
- задержания 423
- пропускания 423
Поляризация 36
Помеха 167
- внешняя 167
- внутренняя 167
- импульсная 1671
- индустриальная 167
- соканальная 112
- флуктуационная 167
Преобразователь 58
- аналого-цифровой (АЦП) 58
- частоты 69
- цифроаналоговый (ЦАП) 59
Приемник 597
- детекторный 616
- прямого усиления 617
- супергетеродинный 618
Принцип
- дуальности 171
- суперпозиции (наложения) 538
Процесс (сигнал) случайный 295
- ансамбль 298
- сечение 299
- стационарный 315
- реализация 297
- центрированный 299
- эргодический 315

Радиоимпульс 30, 115
Радиоканал 52
Радиоприемник 597
Радиосигнал 54
Развертка 87
- кадровая 87
- построчная 89
- строчная 87
- чересстрочная 89
Режим
- активный 386
- ключевой 386
- насыщения 386
- отсечки 386
- покоя 289
- термостабилизации 388
- холостого хода 386
Ряд Фурье 181
- комплексный 181
- обобщенный 179
- тригонометрический 182

Свертка 366
- сигналов 366
- круговая 558
Селективность 612
Сигнал 9
- амплитудно-модулированный 218
- аналоговый (непрерывный) 160
- детерминированный 160
- дискретный 160
- импульсный 160
- комплексный 160
- логический 500
- многомерный (векторный) 160
- непериодический 160
- одномерный 160
- периодический 160
- сопряженный по Гильберту 286
- телеграфный 27
- цифровой 160
- шумоподобный (широкополосный) 340
Синтез 178,
- гармонический 192
- фильтров 430
- цифровых по инвариантности

- - импульсных характеристик 584
- - частотных характеристик 585
Система
- линейная динамическая 359
- обнаружения и измерения 51
- передачи информации 51, 53
- радиолокационная 51
- радионавигационная 51
- радиотелеуправления 51
- радиотехническая 51
- радиоэлектронная 51
- связи 61
- - космическая 51
- - мобильная (подвижная) 1
- - пейджинговая (радиовызова) 61
- - дуплексная 62
- - профессиональная 61
- - симплексная 62
- - сотовая 61
- - спутниковая 61
- - с уплотнением временны́м 62
- - с уплотнением частотным 62
- - телевизионная 85
- - транкинговая 61
- счисления 28
Скалярное произведение 175
Скейлинг-функция 690
Смеситель 69
Сопротивление параметрическое 467
Соты 61
Спектр
- АМ-сигнала 219
- АИМ-сигнала 253
- дискретного сигнала 545
- ЛЧМ-импульса 245
- периодического сигнала 195
- ЧМ-сигнала 236
- энергетический 215
Спектральная плотность 195
- мощности случайного процесса 322
- радиоимпульса 230
Сплайн 542
Способ представления сигнала
- аналитический 283
- временно́й 173
- спектральный 173
- корреляционный 158
Среднее квадратическое отклонение 302
Субгармоника 438
Схема эквивалентная
- каскада с ОЭ 391
- каскада с ОИ 398

Таблица истинности (состояний) 503
Телевидение 24
Теорема
- Винера–Хинчина 323
- запаздывания 200
- Коши 561
- линейности 200
- смещения 200
- центральная предельная Ляпунова 304
Точка
- покоя 398
- рабочая 398
Транзистор
- биполярный 386
- комплементарный 242
- полевой (униполярный) 395
Трафик 56
Триггер 507
- интегральный 508
- симметричный 507
- Шмитта 523

Угол отсечки 445
Умножитель частоты 447
Усилитель 385
- бестрансформаторный 242
- высокой частоты 385
- дифференциальный 412
- избирательный 385
- импульсный 385
- логарифмический 472
- инвертирующий 414
- малошумящий 122
- неинвертирующий 416
- операционный 413
- промежуточной частоты 69, 385, 619
- резонансный 263, 290
- сверхвысоких частот 385
- с общей (им)
- - базой 419
- - затвором 395
- - истоком 3958
- - коллектором (эмиттерный повторитель) 419
- - стоком (истоковый повторитель) 395
- - эмиттером 419

- устойчивый 406

Фаза
- мгновенная 214
- начальная 214
- полная 214
Фазированная антенная решетка 48
Физическая реализуемость 370
Фильтр
- активный 435
- Баттерворта 425
- верхних частот 423
- нерекурсивный (трансверсальный) 567
- нижних частот 423
- оптимальный (согласованный) 628
- полосовой 423
- режекторный (заграждающий) 423
- рекурсивный 403
- трансверсальный 567
- цифровой 567
- Чебышева 425
- электрический 421
Фрактал 185
Функция
- АКФ (корреляционная) 213
- - ЛЧМ-сигнала 244
- Адамара 209
- Берга (коэффициент гармоник) 446
- Бесселя 237
- взаимокорреляционная 215
- передаточная 452
- распределения 305
- Уолша 209
- Хевисайда (включения, единичная) 207
Фурье
- преобразование 193
- - быстрое (БПФ) 552
- - дискретное (ДПФ) 551
- - обратное 194
- - обратное дискретное 551
- - прямое 194

Характеристика
- амплитудная 392
- амплитудно-фазовая 74
- амплитудно-частотная 74
- вольтамперная 387
- - входная (базовая) 387
- - выходная (коллекторная) 387
- вольт-фарадная 453
- детекторная 465
- импульсная 367
- - линейной цепи 367
- - цифрового фильтра 578
- модуляционная 450
- переходная 367
- стоковая 397
- стоко-затворная 397
- фазочастотная 370
- амплитудно-частотная 370
- - цифрового фильтра 580

Цепь
- дифференцирующая 372
- интегрирующая 374
- линейная 353
- нелинейная 354
- параметрическая 357
- с распределенными параметрами 357
- с сосредоточенными параметрами 357

Частота
- квазирезонансная 421
- комбинационная 469
- круговая (угловая) 162
- несущая 30
- поднесущая 239
- промежуточная 69, 467
- резонансная 432
- среза 424
- циклическая 31
Частотный коэффициент передачи 361
Четырехполюсник 356
Чувствительность 612

Ширина спектра 197
Шум
- белый 170
- квантования 591

Элемент нелинейный 438
Электромагнитное излучение 34
Электромагнитная совместимость 45
Энтропия 642
Эталонная модель ВОС 57

# ОГЛАВЛЕНИЕ

**ПРЕДИСЛОВИЕ** ..... 3

**ВВЕДЕНИЕ** ..... 5

**ЧАСТЬ I. РАДИОТЕХНИЧЕСКИЕ СИСТЕМЫ И СИГНАЛЫ** ..... 19

**Глава 1. Основные принципы передачи и приема информации** ..... 19
1.1. Общие сведения об информации, сообщениях и сигналах ..... 19
1.2. Элементы теории распространение радиоволн ..... 33
1.3. Антенны ..... 46
1.4. Основные сведения о радиотехнических системах и сетях связи ..... 50
*Контрольные вопросы* ..... 156

**Глава 2. Основы теории детерминированных сигналов** ..... 157
2.1. Общие сведения о радиотехнических сигналах ..... 157
2.2. Спектральное представление детерминированных сигналов ..... 177
2.3. Основы корреляционного анализа детерминированных сигналов..... 212
2.4. Аналоговая модуляция несущих колебаний ..... 217
2.5. Сигналы с импульсной и цифровой модуляцией ..... 249
2.6. Узкополосные сигналы ..... 277
*Контрольные вопросы* ..... 293

**Глава 3. Случайные и шумоподобные сигналы** ..... 294
3.1. Общие сведения о случайных сигналах ..... 294
3.2. Законы распределения случайных процессов ..... 303
3.3. Стационарные случайные процессы ..... 314
3.4. Спектральный и корреляционный анализ случайных процессов ..... 320
3.5. Узкополосные случайные процессы ..... 329
3.6. Шумоподобные сигналы ..... 340
*Контрольные вопросы* ..... 352

**ЧАСТЬ II. РАДИОЭЛЕКТРОННЫЕ ЦЕПИ И УСТРОЙСТВА** ..... 353

**Глава 4. Линейные цепи** ..... 353
4.1. Общие сведения о методах анализа линейных цепей ..... 353
4.2. Дифференцирование и интегрирование сигналов ..... 372
4.3. Резонансные цепи ..... 377
4.4. Схемотехника усилительных устройств ..... 384
4.5. Элементы теории обратной связи. Устойчивость систем с обратной связью 402
4.6. Интегральные и специальные усилители ..... 412
4.7. Электрические фильтры ..... 421
*Контрольные вопросы* ..... 436

**Глава 5. Нелинейные и параметрические цепи** ..... 437
5.1. Общие сведения об аппроксимации характеристик нелинейных элементов 437
5.2. Отклик нелинейной цепи на гармонический входной сигнал ..... 443
5.3. Нелинейные усилители мощности и умножители частоты ..... 447
5.4. Модуляция колебаний ..... 449
5.5. Детектирование колебаний ..... 456
5.6. Преобразования сигналов в параметрических цепях ..... 466
5.7. Логарифмирование сигналов ..... 472
*Контрольные вопросы* ..... 473

**Глава 6. Генерирование электромагнитных колебаний** ..... 474
6.1. Общие сведения. Генераторы гармонических колебаний ..... 474
6.2. *LC*-генераторы ..... 480
6.3. *RC*-генераторы ..... 487
6.4. Стабилизация частоты в автогенераторах ..... 490
6.5. Цифровые генераторы ..... 493
6.6. Генераторы сверхвысоких частот ..... 496
*Контрольные вопросы* ..... 499

**Глава 7. Импульсные и цифровые устройства** ..... 500
7.1. Общие сведения. Логические сигналы и элементы ..... 500
7.2. Триггеры ..... 507
7.3. Счетчики, шифраторы и дешифраторы ..... 512
7.4. Мультивибраторы ..... 516
7.5. Схемотехника импульсных и цифровых детекторов ..... 525
*Контрольные вопросы* ..... 529

**ЧАСТЬ III. ЦИФРОВАЯ ОБРАБОТКА, ПЕРЕДАЧА И ПОМЕХОУСТОЙЧИВЫЙ ПРИЕМ СИГНАЛОВ** ..... 530

**Глава 8. Цифровая обработка сигналов** ..... 530
8.1. Общие сведения ..... 530
8.2. Цифровое представление сигналов. Теорема Котельникова ..... 531
8.3. Дискретные сигналы. Алгоритмы дискретного и быстрого преобразований Фурье ..... 544
8.4. Дискретная свертка сигналов. Теория $z$-преобразования ..... 556
8.5. Основы теории цифровой фильтрации. Синтез цифровых фильтров ..... 561
8.6. Аналого-цифровые и цифроаналоговые преобразователи ..... 589
*Контрольные вопросы* ..... 596

**Глава 9. Устройства передачи и помехоустойчивого приема сообщений** ..... 597
9.1. Радиопередающие устройства ..... 597
9.2. Радиоприемные устройства ..... 610
9.3. Оптимальная линейная фильтрация сигналов ..... 628
9.4. Элементы теории помехоустойчивого приема ..... 641
9.5. Кодирование сообщений в системах связи ..... 653
9.6. Модемные устройства систем связи ..... 675
*Контрольные вопросы* ..... 687

**Глава 10. Элементы вейвлетного и фрактального анализа сигналов** ..... 688
10.1. Основы вейвлет-анализа сигналов ..... 688
10.2. Общие сведения о фрактальном анализе сигналов ..... 710
*Контрольные вопросы* ..... 721

**ЗАКЛЮЧЕНИЕ** ..... 722

**СПИСОК ЛИТЕРАТУРЫ** ..... 724

**ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ** ..... 727

*Учебное издание*

**Нефедов** Виктор Иванович
**Сигов** Александр Сергеевич

# ОСНОВЫ РАДИОЭЛЕКТРОНИКИ И СВЯЗИ

Редактор *Е.Н. Гарденина*
Внешнее оформление *К.И. Мандель*
Технический редактор *Л.А. Маркова*
Корректор *Г.Н. Петрова*
Компьютерный набор и верстка *Е.Г. Елисеева*

Изд. № PHH-562. Подп. в печать 09.04.09. Формат 70×100 1/16.
Бум. офсетная. Гарнитура «Ньютон». Печать офсетная.
Объем 59.8 усл. печ. л., 60.78 усл. кр.-отт.
Тираж 3000 экз. Заказ № 805.

ОАО «Издательство «Высшая школа». 127994, Москва,
Неглинная ул., 29/14, стр. 1.

Тел.: (495) 694-04-56.
http://www.vshkola.ru. E-mail: info_vshkola@mail.ru

*Отдел реализации*: (495) 694-07-69, 694-31-47, факс: (495) 694-34-86.
E-mail: sales_vshkola@mail.ru

Отпечатано в ООО «Великолукская городская типография».
182100, Псковская область, г. Великие Луки, ул. Полиграфистов, 78/12
Тел./факс: (811-53) 3-62-95
E-mail: zakaz@veltip.ru